Choice Intransitivity and Revealed Salience

Nicola Gennaioli, Matthias Rodemeier, Gregory Sun

March 20, 2025

PRELIMINARY AND INCOMPLETE

Abstract

In a field experiment with a major European e-commerce outlet, we randomize Linear promotions applying a $\delta\%$ discount on all baskets and Nonlinear promotions applying the same discount only to baskets larger than M EUR. Nonlinear promotions exhibit dominated choices and no bunching compared to Linear counterfactuals. These facts point to context dependence: consumers are less price sensitive in Nonlinear than in Linear promotions. The data reject frictions and challenges leading biases. A standard consumer model enriched with salience-driven attention to hedonic features offers a good qualitative and quantitative account of the data.

JEL Codes: D01, D12, D40, D60, D91

Keywords: Revealed Preference, Salience, WARP, Welfare, Bunching, Field Experiments

---

Acknowledgments: Affiliations: Gennaioli and Rodemeier: Bocconi University, BAFFI, and IGIER. Sun: Olin Business School, Washington University. We thank Pedro Bordalo and seminar participants at Bocconi, Harvard, and NYU Abu Dhabi for comments, and Silvia Amalia Meneghesso for excellent research assistance. Gennaioli thanks the European Research Council for financial support under an Advanced Grant (GA 101097578).

1 Introduction

Revealed preference (Samuelson, 1938) is the bedrock of economics. It allows empiricists to infer utility functions and study welfare using choice-based statistics such as price elasticities. Its validity relies on stable utility, embedded in revealed preference or transitivity axioms (Kreps, 1988). Violation of these axioms has been detected (Tversky and Thaler, 1990), but only in the lab and without offering a general mechanism. In a large field experiment on nonlinear pricing we show systematic violation of revealed preference, pointing to context-dependent price elasticity. The data is inconsistent with "frictions", the common account of anomalies in the field, and with behavioral context effects. A standard consumer model enriched with salience-driven attention to hedonic features offers a good qualitative and quantitative account of the data.

Nonlinear pricing is widely used by governments, firms and generally regarded as optimal by economists (Mirrlees, 1971). Many observational studies of nonlinear taxes or prices detect an anomalous insensitivity to incentives, typically attributed to frictions such as adjustment costs. Methods have been developed to measure frictions, allowing to infer structural elasticities that inform welfare analysis (Moffitt, 1990; Kleven, 2016). Compared to observational studies, our experiment allows to compare responses across Nonlinear incentives and Linear counterfactuals. This is key to understand how Nonlinear incentives affect behavior and the drivers of the observed insensitivity.

We randomize Linear and Nonlinear promotions to 720,401 website visitors of a major European e-commerce outlet, resulting in 21,247 buying decisions. Nonlinear promotions give a $\delta\% \in \{10, 15\}$ discount if a consumer's basket is above $M \in \{99, 199\}$ EUR, Linear promotions give the same discount on all baskets, Control gives no discount. The rational model makes a sharp prediction: Nonlinear promotions should create i) an empty region of baskets just below M and ii) excess mass at $b \geq M$ compared to Linear

promotions with the same discount δ. Prediction ii) is conventionally called "bunching".

Spending densities seem to support bunching: in Nonlinear promotions, not in other conditions, there is a visible spike near M and missing mass to its left. Treatment comparisons however show a striking violation. First, the mass below M drops too little compared to Control: 5% of buyers pick dominated baskets (they would spend less by raising b to M) and an estimated 18% of buyers is in excess at lower baskets.

Second, and key, bunching fails. The mass at $b \geq M$ in Nonlinear promotions is directionally $smaller$ than, and statistically indistinguishable from, the mass at $b \geq M$ in Linear ones. But then, where do the spikes come from? There data reveals a third surprising fact: mass to the right of $b >> M$ is smaller in Nonlinear compared to Linear promotions. The rational model predicts no difference in this region. That is, nonlinear promotions stunt high spending. The anomalies are sizable. Across Nonlinear promotions about 2,000 consumers (12.5% of buyers) make mistakes, causing a 163,000 EUR revenue loss for the firm ($\approx 7\%$ of revenues).

It is known that there is "too little" bunching in observational data (e.g. Saez, 2010). We show a deeper problem: there is no bunching compared to the Linear counterfactual. This fact rejects standard stories for insufficient bunching based on search, information, and other costs. We show that these "frictions" stunt consumers' response to Nonlinear and Linear discounts, but they continue to predict that Nonlinear discounts should cause excess mass at $b \geq M$ compared to Linear ones. This is not what we see in the data.

Standard context effects also have problems. Two obvious candidates are anchoring and reference dependence (see Bartels et al., 2023): M may act as a "sticky" starting point for spending or as a target for consumers. These forces go in the wrong direction, though: they strengthen rational incentives to spend $b \geq M$ compared to Linear promotions, again at odds with the data. Context effects can also arise from signaling

(Kamenica, 2008): M signals that the firm is "unfair" or that it has low-quality goods. This is not fully plausible: consumers engage with nonlinear prices in many normal transactions. More important, if Nonlinear promotions carry a bad signal, they should discourage any spending compared to Control, which is not true in the data.

The overarching anomaly in our data, that standard forces fail to explain, is that Nonlinear promotions stunt consumers' price sensitivity compared to Linear ones, shifting observed baskets "to the left" and causing failures of revealed preference. Consumers who shift to the left of M invalidate bunching and create excess mass right below M. Consumers who shift to the left from $b >> M$ contribute to the spike near M. Different puzzles emerge as a byproduct of this force.

Why does the mere presence of M reduce sensitivity to discounts? We offer a theory based on salience-driven attention: The "M" feature in the choice context increases the salience of total spending. When seeing the promotion "10% off above 99 EUR minimum order value" some consumers focus on spending and think "That's a lot! I don't want to spend 99 EUR!". By focusing on the pain of paying, these consumers neglect the 10% discount and act thrifty. Upon seeing only "10% off" many consumers instead think "This is a good deal! I can buy more stuff I need!". By focusing on the lower prices, they increase spending. A third group of consumers attends to both features and behaves optimally, exhibiting also bunching. Overall, promotion context affects attention to different hedonic features of choice. By boosting focus on spending, M interferes with the discount, reducing sensitivity to it.

We formalize this idea building on Bordalo et al. (2025). The model rationalizes the observed choice anomalies. It also endogenizes "M-error messages", received by consumers upon redeeming a Nonlinear promotion at $b < M$. These messages reveal the incidence of representations that focus on δ and neglect M. New comparative stat-

ics follow from bottom up-salience. A more striking discount should boost focus on δ fostering neglect of M. Consistently, higher δ increases error messages, bunching, and reduces missing mass at $b >> M$. A more salient M should foster neglect of δ. Consistently, displaying M more prominently on the website, which we randomize, causes fewer "M-error messages" but not higher sensitivity to the discount. We estimate salience parameters using the frequency of "M-error messages" which are linked to the attention mechanism. The model reproduces choice anomalies and their instability fairly well. The welfare impact of partial representations is sizable.

Consumers choose sub-optimally in several domains, including health (Baicker et al., 2015; Handel and Kolstad, 2015; Abaluck and Adams-Prassl, 2021), investments (Hortaçsu and Syverson, 2004), energy savings (Allcott and Taubinsky, 2015; Rodemeier and Löschel, 2025), rebates (Rodemeier, 2024). Suboptimal response to nonlinear incentives is shown, among others, by (Grubb and Osborne, 2015) for phone plans, (Ito, 2010) for energy prices, Gars et al. (2025) for lotteries, (Saez, 2010) for taxes. According to "schmedul-ing" (Liebman and Zeckhauser, 2004) people react to average rather than marginal prices (see Feldman and Katuščák, 2006, Rees-Jones and Taubinsky, 2020). A key debate is whether these anomalies reflect "frictions" due to real costs or "mental gaps" (Handel and Schwartzstein, 2018). Our findings rule our frictions and point to mental gaps.

We stress salience-driven attention, a general cognitive force that may help reconcile anomalies. When choosing a health plan people may underweight their expected out of pocket costs, not due to a stable bias or friction, but because - in the choice set they face - a product feature such as premiums is very salient (Abaluck and Gruber, 2011). Crucially, then, the same people may also overweight a salient "donut hole" protection or, as shown by Bhargava et al. (2017) salient deductibles. People may exhibit a sluggish reaction to nonlinear incentives not due to stable frictions, but rather due to changing salience of

the features of nonlinear schedules, consistent with Chetty et al. (2013). This implies that methods for measuring frictions and infer true elasticities (e.g. Chetty et al., 2011, Kleven and Waseem, 2013) may rely too much on choice data within a context. Non choice data may be useful, as we elaborate in the Conclusions.

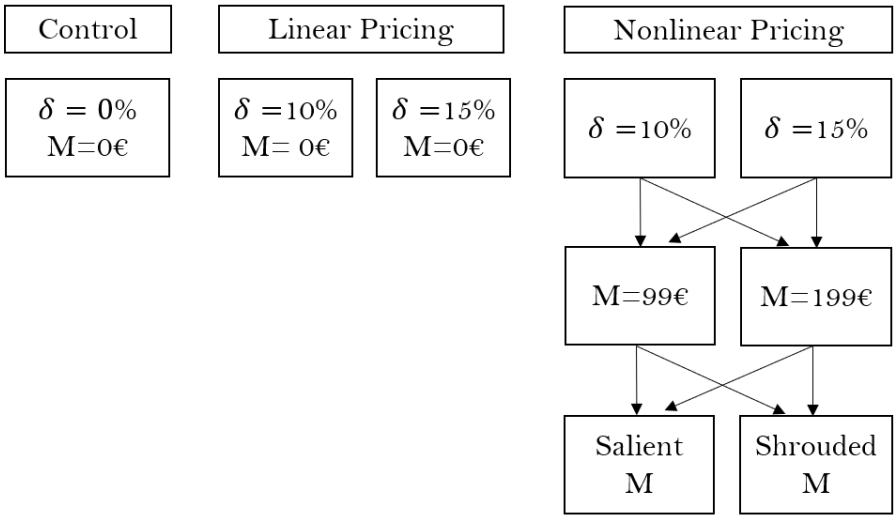
We follow growing work on the cognitive drivers of choice and its anomalies. This work studies costly attention/noise (Woodford, 2012; Caplin and Dean, 2015; Gabaix, 2014), salience (Bordalo et al., 2012; Li and Camerer, 2022) selective memory (Bordalo et al., 2020), and other forces. Building on memory and salience, Bordalo et al. (2025) offer a general theory accounting for puzzles in statistical problems, consumer choice, and risky choice. We develop the first field application of this theory, showing how changing choice context can cause violations of rationality axioms. Laboratory or observational tests of these axioms often find small violations (see Varian, 2006). Our model helps reconcile this evidence with ours. People may act consistently within a choice context, when their representation (even if incorrect) is fixed, but act inconsistently across contexts, when their representation changes. Evidence in de Clippel et al. (2025) and Ariely et al. (2003) seems in line with this notion.

Section 2 describes the experiment. Section 3 derives rational predictions and show their violation, discussing possible mechanisms. Section 4 presents our model. Section 5 tests its key comparative statics. Section 6 structurally estimates it. Section 7 concludes.

2 The Experiment

The experiment took place in the German version of a large online shop for household appliances in the EU. It entails 11 treatments. Linear promotions apply a discount $\delta \in \{10, 15\}\%$ on basket value. Nonlinear promotions apply the same discounts only if basket value is above a minimum order value $M \in \{99, 199\}$ EUR, which we cross

Figure 1: Experimental Design



Notes: This figure shows the experimental groups. The discount size is denoted by δ, the minimum order value by M. The arrows between boxes indicates cross-randomization.

randomize. We also randomize whether M is prominently described or not. The control group gets no promotion, $\delta = M = 0$. See Figure 1.

Upon entering the website, subjects are randomized into one treatment based on their HTTP-Cookie and stay in the same group on follow-up visits. Subjects in a promotion immediately learn about it from a banner placed at the top of the webpage, and keep seeing the latter when browsing. Figure 2 shows in Panel (a) the banner for a Linear promotion with $\delta = 10\%$, and in Panel (b) for a Nonlinear promotion with $\delta = 10\%$ and minimum order value $M = 99$. If M is shrouded, the banner reports only the discount, as in Panel (a). By clicking on the banner, subjects access the promotion page which reports (δ,M) and further details, see Figure 15. To redeem *any* promotion, subjects must access this page and obtain the coupon code. M can also be learned when a subject checks out with a basket value $b < M$, which prompts an error message saying that the basket is too small to qualify for the promotion.[1]

[1]One concern with e-commerce experiments is that subjects could see more than one treatment if they switch devices. Prior experiments with the same shop find this event to be rare (Rodemeier, 2024). Also in

Figure 2: Promotion Banners



(a) Promotion banner under Linear (or Nonlinear with Shrouded M



(b) Promotion Banner under Nonlinear with Salient M

Notes: This figure shows examples of the promotions. Panel (a) shows the Linear promotion banner with a discount of 10%, Panel (b) the Nonlinear promotion with he same discount and a salient $M = 99$

The banners were programmed to be a prominent integral part of the website (i.e. they could not be avoided using pop-up blockers). At checkout, consumers were also reminded to redeem their promotion. Subjects are thus in all likelihood aware of promotions (but for us it is enough that they are equally aware of different promotions).

The experiment was held between October 12 and 25, 2020. We observe 1,318,655 online sessions by 720,401 unique subjects (so the average subject visits the website roughly twice). Each treatment consists of around 65,500 subjects. In Appendix Table 5 observables are balanced across groups, showing successful randomization. 3% of subjects make a purchase, amounting to 21,247 buyers ($\approx$ 1,930 per treatment).

3 Basic Facts and Rational Choice Violations

We present our key findings, organized around a neoclassical consumer model.

3.1 The Rational Model

The set of consumer types is $J \subseteq \mathbb{R}_+$. Type $j \in J$ maximizes a continuous, increasing and convex utility function $u_j(\mathbf{x})$ in the budget set $\left\{\mathbf{x} \in \mathbb{R}^{\mathbf{n}}_+ : \mathbf{p}'\mathbf{x} \leq w_j\right\}$. Across our

our experiment, among customers who purchase more than once (who are identifiable to the firm because buyers must report info such as name, address, etc.) we see multiple cookies only in 2.8% of cases. Internal validity is confirmed in Table 5: observables are balanced across treatments.

treatments *relative* prices in the store are constant, so by Hicks separability (Varian 1992, p. 148) the consumer chooses, without loss of generality: i) in store consumption $b/p_{j,r}$ where b is the basket value at regular prices and $p_{j,r}$ is a type-specific regular price index, and ii) an outside good x, whose price is set to one.

Let t_i be the generic treatment, where $i = C$ is Control, $i = \delta$ is a Linear promotion with discount δ, and $i = M, \delta$ is a Nonlinear promotion with discount δ and minimum order value M. Note that Linear treatments set $M = 0$ and Control ones set $M = \delta = 0$, in the generic treatment, type j solves:

$$\max_{b,x} \; u_j(b, x) \tag{1}$$

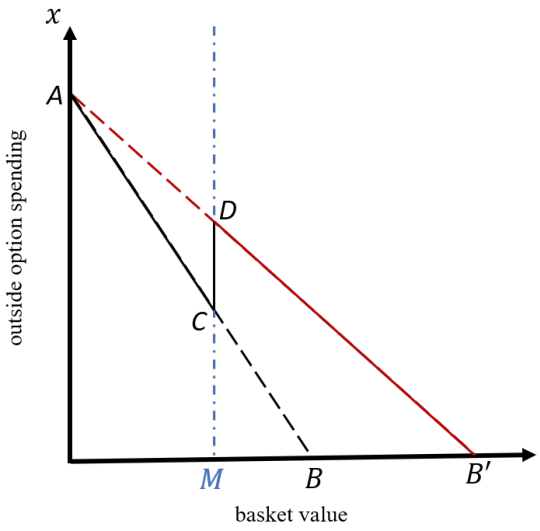
$$s.t. \quad x + (1 - \delta \cdot \mathbb{I}_{b \geq M}) \cdot b \leq w_j \,, \tag{2}$$

where $\mathbb{I}_{b \geq M} = 1$ if $b \geq M$ and $\mathbb{I}_{b \geq M} = 0$ otherwise.[2]

Treatment i affects the store price index p_{ji} and hence the consumer's budget set as in Figure 3. The abscissa reports the outside good x, the ordinate the regular price basket b. AB is the budget line under Control, which has slope -1 because store prices are regular, $p_{j,C} = p_{j,r}$. AB' is the budget line under Linear, which has slope $-(1 - \delta)$ because all regular prices are discounted by $\delta\%$, $p_{j,\delta} = (1 - \delta) \cdot p_{j,r}$. $ACDB'$ is the budget under Nonlinear: store prices are regular for $b < M$ and discounted by $\delta\%$ otherwise. The jump from C to D is the infra-marginal discount at $b = M$. Compared to Control, Linear and Nonlinear promotions expand the budget set, the latter only for $b \geq M$.

[2]The budget set in t_i is $x + b \cdot (p_{ji}/p_{jr}) \leq w$, where (p_{ji}/p_{jr}) is the price of store goods in t_i relative to regular. Utility $u_j(b, x)$ subsumes the index $p_{j,r}$ as a taste parameter.

Figure 3: Budget Lines



Notes: This Figure shows the budget line in Control (AB), Linear (AB'), and Nonlinear (ACDB').

Optimization by different types j produces a CDF of baskets in t_i denoted by:

$$\Pr(\widetilde{b} \leq b|t_i) = B\left(b|t_i\right),$$

As proven in the Appendix, under general conditions three predictions obtain. To aid intuition, we discuss them under quasi-linear utility, $u_j\left(b\right)+w_j-b$, and single crossing.[3]

Proposition 1. *under Nonlinear treatment $i = M, \delta$ there is a threshold $\underline{b}_i \leq (1-\delta)\cdot M$ such that the CDF of baskets obeys:*

$$B\left(b|t_i\right) = \begin{cases} B\left(b|t_C\right) & if \quad b \leq \underline{b}_i \\ B\left(\underline{b}_i|t_C\right) & if \quad \underline{b}_i < b < M \\ B\left(b|t_\delta\right) & if \quad b \geq M \end{cases}. \tag{3}$$

[3]Single crossing says that higher types have higher marginal utility, e.g. $u_j = j \cdot u(b)$. We use the definiton: if $b' > b$ and $j' > j$ or $b' < b$ and $j' < j$ then $u_j(b') > u_j(b)$ implies $u_{j'}(b') > u_{j'}(b)$.

Low consumers types (low j) choose $b \leq \underline{b}_i$ and do not do not earn a discount. In this range, the CDF under Nonlinear $t_{M,\delta}$ is equal to that under Control, $B(b|t_C)$. Nonlinear promotions matter at higher basket values, in particular around M. This gives rise to the first two predictions, which are the focus of bunching studies.

Rational Prediction 1 (Empty Region) *The "bunching region" $\mathcal{B} \equiv (\underline{b}_{M,\delta}, M)$ is empty under $t_{M.\delta}$,. The dominated region $\mathcal{D} \equiv (M(1-\delta), M) \subseteq \mathcal{B}$ is thus empty too.*

Types who buy in $\mathcal{B}$ under a Linear promotion t_δ, under a Nonlinear promotion $t_{M,\delta}$ bunch at $b = M$ so as to earn the infra-marginal discount. Thus, $\mathcal{B}$ is empty. A fortiori, subregion $\mathcal{D}$ is also empty: if a consumer originally setting $b \in \mathcal{D}$ buys an extra item to reach $b = M$, her actual spending drops to $M(1-\delta)$. Baskets in $\mathcal{D}$ are dominated by $b = M$, so they should not be not-chosen given increasing utility. In Figure 3, bundle (b, x) to the immediate left of C is dominated by D. To measure avoidance of dominated choice, we compare the mass in $\mathcal{D}$ under Nonlinear promotions to the mass in the same region under Control (and under Linear).

Rational Prediction 2 (Bunching) *Suppose that* $\Pr(b \in \mathcal{B}|t_\delta) > 0$. *Then, under Nonlinear promotion $t_{M,\delta}$ the mass of baskets at or above M is strictly larger than under Linear promotion t_δ,* $\Pr(b \geq M|t_{M,\delta}) > \Pr(b \geq M|t_\delta)$.

Consumers shopping in $\mathcal{B}$ under Linear t_δ should bunch at $b \geq M$ under the Nonlinear promotion, creating excess mass at $b \geq M$ compared to t_δ. Observational studies (Kleven and Waseem, 2013) detect bunching as "missing mass" below M and a spike at $b = M$. We can measure bunching directly, by comparing the mass in $b \geq M$ under

Nonlinear $t_{M,\delta}$ to the mass in the same region under Linear t_δ. This test is more precise that looking at $b = M$: it allows consumers to "bunch" at $b > M$ due to lumpy prices (absent lumpiness, bunchers would set $b = M$). Our treatments also allow us to compare excess bunching mass at b to the "bunching potential": the mass in $\mathcal{B}$ under t_δ.

The third rational prediction is "Independence of Irrelevant Alternatives" (IIA).

Rational Prediction 3 (IIA) *The mass of baskets in* $b > M$ *is the same under Nonlinear and Linear promotions,* $\Pr(b \geq M + \theta | t_{M,\delta}) = \Pr(b \geq M + \theta | t_\delta)$ *for any* $\theta > 0$.

Consumer types spending $b \geq M$ under Linear are not affected by the irrelevant M. The Linear budget set contains the Nonlinear one, see Figure 3. Thus, a basket $b > M$ chosen under Linear should also be chosen under Nonlinear because it is affordable in both promotions. Aggregating across consumers, Prediction 3 follows. Due to lumpy prices also some "bunchers" may set $b > M$, so the condition can be weakened to $\Pr(b \geq M + \theta | t_{M,\delta}) \geq \Pr(b \geq M + \theta | t_\delta)$, but equality should be expected for $\theta >> 0$.

These predictions hold under our general utility function. Quasi-linearity further implies that discounts increase spending, $B(b|t_\delta) \leq B(b|t_{M,\delta}) \leq B(b|t_C)$. Single crossing implies that the CDF under Nonlinear $t_{M,\delta}$ coincides with Control for $b < \underline{b}_{M,\delta}$.

Section 3.2 documents our key puzzles. Section 3.3 discusses their implications.

3.2 Failures of Rationality

Appendix Table 6 shows summary statistics of our treatments. The share of subjects buying something is 2.88% in Control, it is on average 2.91% in Nonlinear and 3.13% in Linear promotions. The average basket per subject is respectively equal to 2.17 EUR, 2.34 EUR, and 2.53 EUR, which becomes 155.2 EUR, 157.6 EUR, and 158.2 EUR con-

ditional on purchasing. Intuitively, promotions increase spending.

Figure 4 plots the pdf of baskets in Nonlinear promotions, reporting on the abscissa the share of subjects choosing b in a particular range. We pool $\delta = 10\%, 15\%$, omit the very large bar at $b = 0$ and focus on $b < 300$ EUR. Consistent with observational studies, Nonlinear promotions induce a visible spike at the minimum order value and "missing mass" to its left, $M = 99$ in Panel (a) and $M = 199$ in Panel (b).[4]

We now test for Predictions 1 and 2. By Prediction 1, the dominated region $\mathcal{D}$ should be empty. Figure 5 Panel (a) reports on the left the average dominated mass in Nonlinear treatments, $\Pr(b \in \mathcal{D}|t_{M,\delta})$, on the right the mass in $\mathcal{D}$ under Control (right). We divide these masses by 3%, the average probability of $b > 0$, so that numbers are indicative of the share of buyers making dominated choices.

About 5% of buyers make dominated choices $b \in \mathcal{D}$, which range from 3.01% of buyers for $M = 199$ and $\delta = 10$, to 6.26% for $M = 99$ and $\delta = 15$. In our Nonlinear promotions there are about 15,000 buyers, corresponding to about 700 people making dominated choices in total. This is a striking violation of rationality: these consumers could spend less by reaching M.

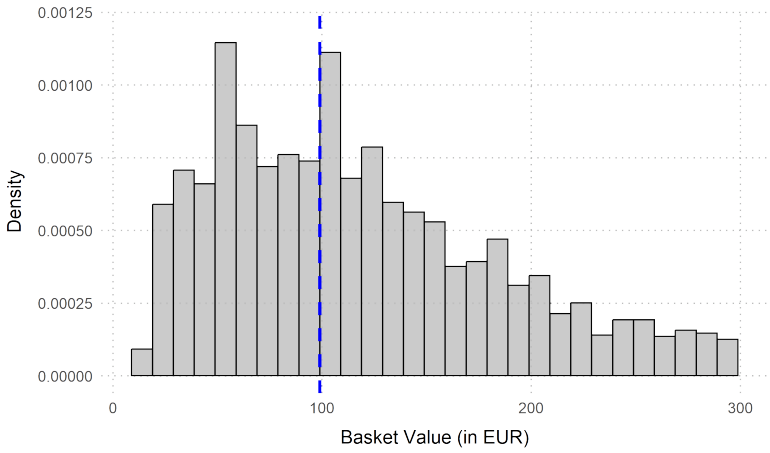
Under Control, the mass in $\mathcal{D}$ is 6%. The difference with 5% is significant (p-value of 0.03), but says that only 1 out of 6 buyers in $\mathcal{D}$ move out of it based on a Nonlinear promotion. Using standard methods (see Appendix J), we can identify the bunching region $\mathcal{B}$ by estimating threshold $\underline{b}_{M,\delta}$ in different treatments.[5] On average, this region extends to about 30 EUR below M. That is, a consumer is roughly indifferent between

[4]The firm chose $M = 99, 199$ EUR to match the 50% and 75% spending percentiles. In Control, these percentiles are 109 and 190 EUR, so targeting was fairly accurate. The pdf under Control, and that under Linear, are in Appendix Figure 12

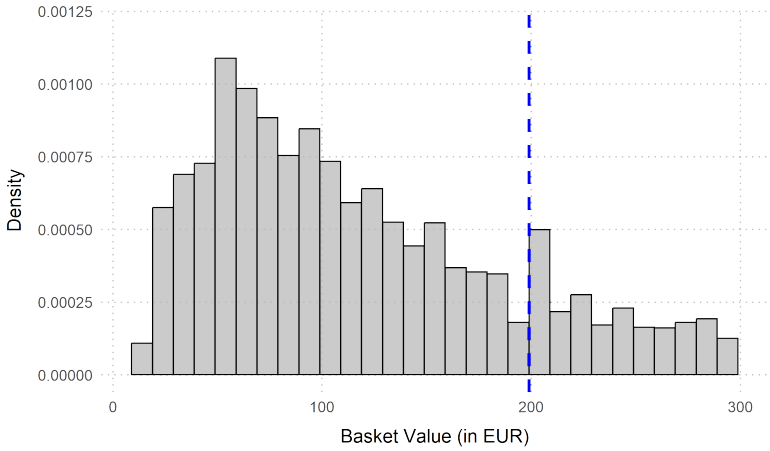
[5]The threshold is inferred by imposing an indifference condition between the regular price basket $\underline{b}_{M,\delta}$ and the discounted M for a second order approximated utility. The condition depends on price elasticity ε, which we estimate from Linear treatments. If such elasticity varies across types, the threshold would be type specific. However, because the latter is a nearly linear function of ε, we can approximate the size of $\mathcal{B}$ using a unique threshold based on average elasticity.

Figure 4: Spending Distributions

(a) Nonlinear, $M = 99$



(b) Nonlinear, $M = 199$



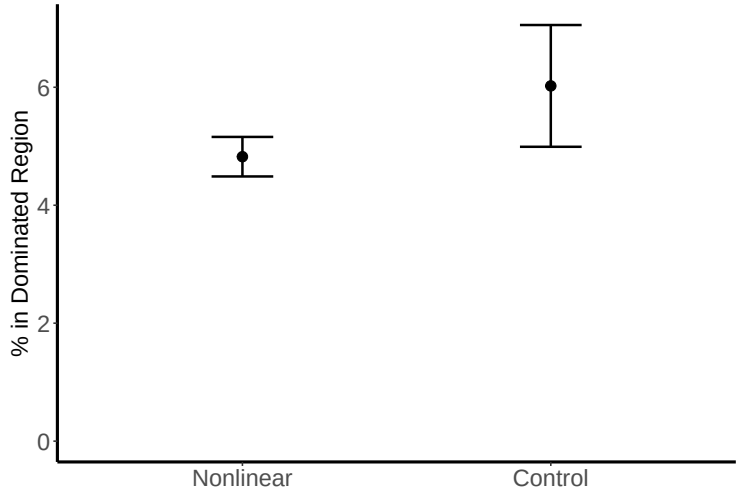
Notes: This Figure plots the density of baskets in Nonlinear with $M = 99$ (Panel a) and with $M = 199$ (Panel b). In both panels we pool $\delta = 10, 15$. We exclude the mass $b = 0$ and $b > 300$. The first bin starts 10 and ends at 19, and bins continue in 10 EUR increments.

spending $b = 70$ and bunching at $M = 99$ EUR. In Nonlinear promotions, $\mathcal{B}$ contains on average 18.0% buyers. Under Control it contains 28.2% of buyers, under Linear 25.9% of buyers. Again, there is a significant reduction, 10 p.p. compared to control and 8 p.p. compared to Linear, but Prediction 1 of empty $\mathcal{B}$ is strongly rejected.[6]

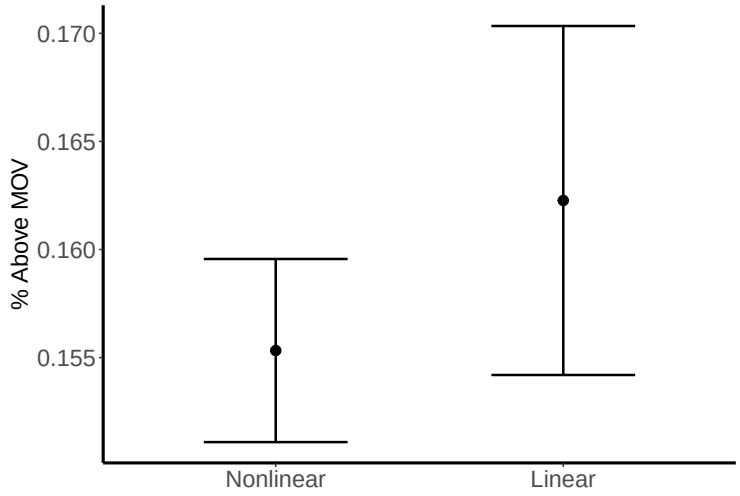
[6]To identify $\mathcal{B}$ we relied on the average elasticity estimated from Linear. Results change quantitatively but not qualitatively if we estimate price elasticity using bunching methods, which rely on the Nonlinear

Figure 5: Rational Predictions 1 & 2

(a) Incidence in Dominated Region



(b) Incidence Above M



Notes: Panel (a) plots the percent of baskets in $\mathcal{D} \equiv (M \cdot (1-\delta), M)$ arithmetically averaged across all (M, δ) and divided by the 3% overall purchase probability, on the left for Nonlinear, on the right Control. Panel (b) report the same computation for event $b \geq M$, on the left for Nonlinear, on the right for Linear.

Excess mass in $\mathcal{B}$ is large. It entails 2,831 subjects in total, and it has significant revenue consequences for the firm. If all buyers in $\mathcal{B}$ set $b = M$, the revenues for the

CDF only. In this case, the mass in the bunching region would be 9.2% in Nonlinear compared to 14.8% in Control and 13.4% in Linear.

firm increase by 40,879 Euros, which accounts for 1.7% of revenues.[7]

Dominated choices and excess mass in $\mathcal{B}$ reveal a stunted sensitivity to Nonlinear discounts. This phenomenon was also noted in observational studies, e.g. on taxes (Saez, 2010) and attributed to "frictions" created by choice set constraints (e.g. 8 daily work hours) or to costs of search, information, etc (e.g. Chetty, 2012). In our setup these costs are arguably small, so insensitivity is more puzzling. Consumers see in almost every webpage cheap items such as lightbulbs, allowing them to easily avoid dominated choices if they wish to. Frictions may however arise from the cost of redeeming a discount, or of deciding whether to buy an additional item, etc.[8]. In Section 3.3 we show that, even if present, such frictions cannot explain the data.

Consider next the key Prediction 2 on Bunching: the mass at $b \geq M$ should be strictly larger in Nonlinear promotions $t_{M,\delta}$ than in Linear t_δ ones. Observing the Linear counterfactual is key for this test. For each promotion type, we compute the share of subjects in the treatment spending $b \geq M$, namely $\Pr(b \geq M|t_i)$. We again scale this mass by 3% to express it as a share of buyers. Figure 5 reports in Panel (b) the average scaled mass in $b \geq M$ under Nonlinear promotions (left) and Linear promotion (right). Across promotions, roughly 30% of buyers spend $b \geq M$. Critically, the mass in $b \geq M$ is if anything $smaller$ in Nonlinear than Linear (the difference is not significant). There is no evidence of bunching, even directional, violating Prediction 2. A one sided t test rejects that the mass in $b \geq M$ is larger under Nonlinear compared to Linear against the alternative $X \leq 0$ with a p-value of 0.1. Remember that, as we estimated before, the "bunching potential" in $\mathcal{B}$ is about 26% of buyers. In light of this, it is truly surprising that no extra mass is present under Nonlinear.

[7] If we make a finer computation, accounting for the fact that some consumers in $\mathcal{B}$ would optimally spend $b > M$, we find that the revenue loss for the firm is 44,039 Euros.

[8] Costs of redeeming do not seem promising also. About 40% of consumers (so well above 1 out of 6) redeem Linear coupons at basket values close to M, see the regression in Figure 13.

This fact, which we call "no bunching", is striking for two reasons. First, it rejects a strong and general rational prediction: Nonlinear discounts should cause more people to move to $b \geq M$ compared to Linear, but they do not. Second, in Figure 4 we saw clear spikes right above M which are not merely visual: in our experiment the share of subjects in right neighborhoods $[M, M + \upsilon]$ of M significantly larger in Nonlinear than in Linear treatments (see Appendix Figure 16). Thus, Nonlinear promotions cause the spikes, but where do the latter come from?

A clue to unlock the answer is the IIA Prediction 3 on behavior at $b > M$. IIA cannot also be tested in observational studies, where the counterfactual Linear CDF is unobserved. We measure, in analogy with our bunching test, excess mass under Nonlinear compared to Linear for several right tails $b > M + \theta$, where $\theta > 0$. We compute percent excess mass in Nonlinear relative to Linear:

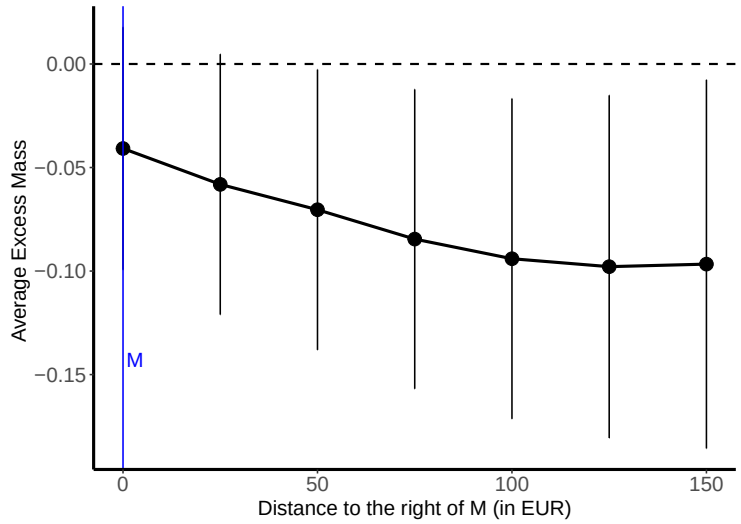
$$X(\theta)_{M,\delta} = \frac{\overbrace{\Pr\left(b \geq M + \theta | t_{M,\delta}\right)}^{\text{Mass at or above } M+\theta \text{ in Nonlinear}} - \overbrace{\Pr\left(b \geq M + \theta | t_{\delta}\right)}^{\text{Mass at or above } M+\theta \text{ in Linear}}}{\underbrace{\Pr\left(b \geq M + \theta | t_{\delta}\right)}_{\text{Mass at or above } M+\theta \text{ in Linear}}} \tag{4}$$

for 6 values $\theta \in \{25, \ldots, 150\}$ in 25 EUR increments. Figure 6 reports average excess mass $X(\theta)$ across treatments $t_{M,\delta}$ for the 6 values of θ on the ordinate. For comparison we report excess mass at $b \geq M$, i.e. $X(0)$. The IIA Prediction 3 says that for $\theta > 0$ excess mass should be zero.

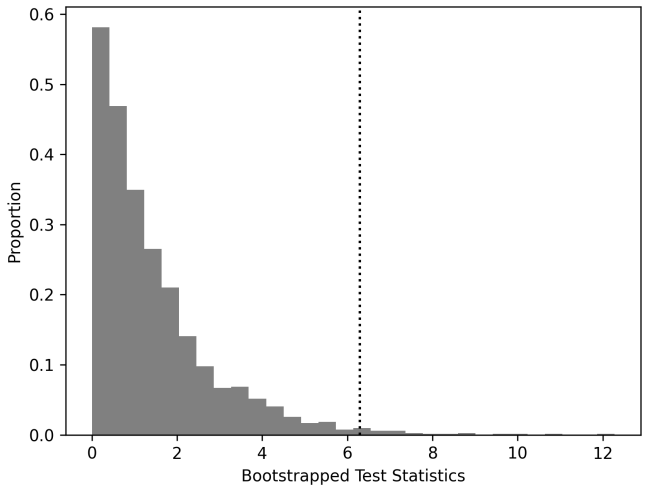
Contrary to Prediction 3, excess mass is $negative$ and significant for $b \geq M+50$. The Nonlinear promotion causes "missing right tail mass" compared to the Linear promotion. This is a striking violation of IIA: for $b \geq M + 50$ or above the presence of M should be irrelevant. This violation can help explain why there are visible spikes near M and yet we see "no bunching" : some consumers move toward M from above. The mere

Figure 6: Rational Prediction 2: No Excess Mass in the Right Tail

(a) Missing Mass



(b) Bootstrap Test



Notes: Panel (a) plots excess mass in Equation 4 averaged across (M, δ), as a function of θ. $\theta = 0$ shows excess mass for $b \geq M$. Any $\theta > 0$ shows excess mass farther above M. The bars represent 95%-CIs computed using the delta method. Panel (b) plots the distribution of bootstrap draws for the test in Equation (24), and the test statistic in the true sample, given by the dotted line.

presence of M seems to stunt spending in Nonlinear versus Linear promotions.

We further probe IIA using a bootstrap test based on the rational null: "all upper tails

$b > M$ have higher or equal mass in $t_{M,\delta}$ than in t_δ".[9] We construct a test statistic by "counting" instances in which - for a range of values $b' > M$ - the null is violated. The exact procedure is described in Appendix F. Figure 6 plots the the test statistic in our sample (dashed line) and its bootstrap pdf under the null (where Nonlinear and Linear right tails and drawn from the same CDF). The test statistic in the data is too extreme to be generated by sampling error, rejecting the IIA Prediction 3 (p-value $= 0.014$).

Violations of IIA range from 5% of the right tail under Linear treatments when $\theta = 25$ to 10% when $\theta = 150$. At $\theta = 25$, the 5% missing mass translates into 316 people missing from the upper tail, which is roughly 1.4% of buyers. Critically, because these buyers are "big spenders", the revenue implications are significant. Average consumer spending conditional on being above M is 6% lower in Nonlinear compared to Linear treatments. Because in Linear treatments the average basket in $b \geq M$ costs about 300 EUR, aggregated across the roughly 7,000 consumers choosing $b \geq M$ in Nonlinear treatments, the drop in spending caused by the violation of IIA is about $123,000$EUR = $7000 \times 300 \times 0.06$ EUR. This is a large amount. It corresponds to about 5% of the firm's revenue from Nonlinear treatments during the experimental period.

3.3 Understanding the Puzzles

We reject three general rational predictions of Nonlinear promotions:

1. Non-Empty Region. About 5% of buyers choose dominated baskets, $b \in (M \cdot (1-\delta), M)$. About 18% of buyers choose baskets in the "bunching region". Both regions should be empty.

2. No Bunching. The share of subjects choosing $b \geq M$ is not larger in Nonlinear

[9] Formally, the null is $\forall b > M$, and (M, δ), we have $B(b|t_{M,\delta}) - B(b|t_\delta) \leq 0$. The alternative is "$\exists b > M$ and (M, δ) such that $B(b|t_{M\delta}) - B(b|t_\delta) > 0$.

promotion $t_{M,\delta}$ compared to Linear promotion t_δ.

3. Failure of IIA. The share of subjects choosing $b >> M$ is smaller in Nonlinear promotion $t_{M,\delta}$ compared to Linear promotion t_δ.

These violations are significant: 1) and 3) entail about 3,100 buyers (about 20% of the total) and amount to a revenue loss of about $163,000$ EUR ($\approx 7\%$ of revenues).

In observational studies, insufficient bunching and dominated choice are explained with "frictions" (Dubé et al., 2009; Chetty et al., 2011). We now show that this force cannot explain our data. We then argue, less formally, that rational inattention or standard context effects also have trouble explaining the data. This opens the room for our approach, based on salience-driven mental representations.

Frictions. Suppose that type j bears additional cost $A_j(b, b_C^j) + \epsilon \cdot \mathbf{1}_{b>b_C^j}$ of choosing b, where $\epsilon \geq 0$ is iid across consumers with full support. Function $A(.)$ embeds the search, information or deliberation costs of increasing the basket above the Control optimum b_C^j: it is zero for $b \leq b_C^j$ and increases in b otherwise.[10] ϵ captures random costs such as hassle, distraction (including failure to see the promotion banner, or forgetting). The key restriction is that, consistent with existing work, frictions affect any promotion.

Given Nonlinear parameters δ and M, we partition the choice set in three regions: the lower tail $\mathcal{L} \equiv [0, M \cdot (1-\delta)]$, the dominated region $\mathcal{D} \equiv (M \cdot (1-\delta), M)$, and the right tail $\mathcal{U} \equiv [M, \infty)$. A consumer's Control basket b_C^j lies in one region. We define her reaction to a Nonlinear $t_{M,\delta}$ or Linear t_δ promotion based on her "transition" to a different region of the partition. Aggregate reaction to t_i is thus captured by the mass of consumers choosing $b \in \mathcal{X}$ under Control and $b' \in \mathcal{Y}$ under t_i itself:

$$\pi(\mathcal{X}, \mathcal{Y} | t_i) \equiv \Pr(j \in J : b_C^j \in \mathcal{X} \cap b_{i,A}^j \in \mathcal{Y}),$$

[10] $A(.)$ can be specified relative to any basket lower than the full rational promotion optimum.

with $\mathcal{X}, \mathcal{Y} \in \{\mathcal{L}, \mathcal{D}, \mathcal{U}\}$. $b^j_{i,A}$ is the consumer's optimal basket under t_i with frictions. The rational frictions model makes the following prediction.

Proposition 2. *If* $\Pr(b \in \mathcal{D}|t_{M,\delta}) < \Pr(b \in \mathcal{D}|t_C)$*, more consumers should "transition" from* $\mathcal{D}$ *to* $\mathcal{U}$ *under a Nonlinear than a Linear promotion,* $\pi(\mathcal{D},\mathcal{U}|t_{M,\delta}) > \pi(\mathcal{D},\mathcal{U}|t_\delta)$*. The bunching Prediction 2 must then hold:* $\Pr(b \geq M|t_{M,\delta}) > \Pr(b \geq M|t_\delta)$.

Frictions allow for some dominated choices, but they also imply bunching by consumers whose adjustment cost is small enough. The reduction of the mass in $\mathcal{D}$ (and in $\mathcal{B}$) in Nonlinear promotions ensures that some such consumers exist. Thus, the Nonlinear promotion should exhibit bunching compared to the Linear one. This is not what we see in the data. By Proposition 2, then, frictions are rejected.[11]

Rational Inattention. Each promotion has two features. Linear has $M = 0$ and $\delta > 0$, Nonlinear has $M > 0$ and $\delta > 0$, Control has $M = \delta = 0$. With costly attention, consumers may optimally confuse different treatments (Caplin and Dean, 2015). This force can give positive mass in $\mathcal{D}$ and no bunching if for some people the optimal information structure gives an imperfect yet informative signal about the presence of a promotion, distinguishing $\delta = 0$ versus $\delta > 0$, but does not distinguish between $M = 0$ and $M > 0$. Attending to M may be too costly, even for consumers who are likely to spend near it.

This explanation is however unable to explain why behavior differs in Nonlinear versus Linear promotions (e.g., the probability of buying is larger in the latter), but at the same time we see no bunching. If some people attend to M, they must be consumers near M, for which the incentive to bunch is strongest. This then implies positive excess mass at $b \geq M$ under Nonlinear compared to Linear, which is counterfactual.

Rational inattention also fails to explain violations of IIA. For such violations to occur, high types j must be less likely to receive a signal about the presence of $\delta > 0$

[11] The proof shows that friction are also inconsistent with behavior in $b >> M$, violation 3). It also shows that Proposition 2 extends to random errors in implementing the optimal basket, $b^j_{i,A} = b^j_i + \epsilon$.

when the true promotion is Nonlinear than when it is Linear. Given that when $b \geq M$ consumers face the same benefit of attending to δ in Nonlinear and Linear promotions, this signal structure can only be optimal if attending $only$ to δ is costlier when the current promotion has $M > 0$ rather than $M = 0$. That is, neglecting $M > 0$ should be costly, an assumption that is hard to swallow.

In sum, neither frictions nor rational inattention seem capable of explaining our key puzzle: lower sensitivity to a Nonlinear compared to an equivalent Linear discount. This context dependent price elasticity challenges core axioms of economic theory such as WARP and transitivity.[12] Relaxing these axioms implies that preferences are unstable, including a key welfare metric such as price elasticity.

We can directly check for this implication by estimating price elasticity across promotions. In Nonlinear treatments we apply standard bunching methods (Kleven, 2016). In Linear treatments we rely on exogenous experimental variation, using both a linear demand model and a quantile regression (which allows for highly nonlinear demand). Both methods yield estimates around M, which we compare with the estimates from Nonlinear treatments. Appendix G describes these methods. Table 1 reports the average (across δ) elasticities estimated for $M = 99$ and $M = 199$ in Nonlinear and Linear treatments using quintile regressions and linear demand. We bootstrap standard errors for the Nonlinear and quantile estimates, we use the delta method for linear demand estimates.

Price elasticity is uniformly lower in Nonlinear promotions. The difference with the linear demand model is large and significant (p-value of 0.02). the difference with the quantile estimates is directionally the same but marginally significant due to limited power (p-value 0.16).[13] These estimates imply different welfare computations for the

[12] For a consumer j WARP says: if a bundle y is affordable when x is chosen, then if y is chosen x must be not affordable. Consider failures of IIA: for them to occur, at least one consumer type j must choose $b > M$ under Linear and $b' < b$ under Nonlinear. In Figure 3, though, b' and b are affordable in both promotions. Thus, the change in behavior violates WARP, creating also room for intransitivities.

[13] The linear model exploits variation across the entire basket value, the quantile effects approach only

Table 1: Elasticities From Various Estimation Methodologies

Nonlinear/Bunching	Linear/Quantile Effects	Linear/OLS
-0.264	-0.752	-0.957
(0.063)	(0.338)	(0.292)

Notes: The first column reports the average elasticity estimated in Nonlinear promotions. The second column computes elasticities using quantile treatment effects across Linear versus Control. The third column uses OLS estimates comparing Linear and Control. Standard errors are in parentheses.

same consumer with respect to the very same choices. Context dependence matters.

Why does the mere presence of M reduces consumers' sensitivity to the same discount δ, producing context dependent price elasticity? Standard context effect theories suggest that M may act as a signal (Kamenica, 2008) or as a reference point/anchor (Tversky and Kahneman, 1974; Kahneman and Tversky, 1979).

Signalling. The presence of M may signal that the firm has low quality goods (e.g. it wants to get rid of many of them), or that it is "unfair". A Nonlinear promotion may thus reduce consumers' willingness to spend by conveying "bad news". Signalling however implies that the Nonlinear promotion would also reduce spending in the left tail, in particular the probability of buying anything, compared to Control t_C. This prediction is rejected in the data. As shown in Appendix Figure 15, below the bunching region ($b < \underline{b}_{M,\delta}$) the CDF of baskets is indistinguishable across Nonlinear and Control treatments, inconsistent with M sending a bad signal to consumers.

Anchors and Reference Points. Anchors and reference points imply that choice can be influenced by "salient" yet irrelevant numbers. By anchoring (Tversky and Kahneman, 1974), the consumer in a Nonlinear promotion starts from M and adjust too little away from it (perhaps thinking that M is a reasonable amount). By reference dependence, the consumer views "reaching M" as a goal, analogously to a marathon runner striving to fulfill a salient 3 hours finishing time target Allen et al. (2017).

uses CDF values near M. Thus, the linear model requires stronger assumptions to be valid, but produces lower standard errors.

Bartels et al. (2023) discusses evidence for anchors and reference points. Anchors imply that consumers should move toward M symmetrically, from above and below. Reference points imply that consumers should strive toward reaching M from below, not from above. Critically, for transitions across $\mathcal{L}, \mathcal{D}, \mathcal{U}$, allowing anchor or reference M to distort rational behavior should increase transitions from $\mathcal{D}$ to $\mathcal{U}$ to $\pi^*(\mathcal{D}, \mathcal{U}|t_{M,\delta}) > \pi(\mathcal{D}, \mathcal{U}|t_{M,\delta})$. Of course, this effect would not influence behavior in t_δ, where M is absent, nor would it influence other relevant transitions. Thus, the inequalities in Proposition 2 must a fortiori hold when M is as an anchor or a reference point. "No bunching" therefore rejects also these mechanisms.[14]

We are back to the question: how can the mere presence of M reduce sensitivity to the same δ compared to a Linear promotion? Building on Bordalo et al. (2025), we offer a model in which the features of a Nonlinear promotion, δ and M, do not just affect budget sets. They also cue different reasons for choice. Seeing the minimum order value M cues a consumer to think about her total spending. Attention to M focuses her on the pain of paying and hence on "controlling costs". Seeing the discount δ cues the consumer about temporarily lower prices. Attention to δ focuses her on "making a good deal".

Selective attention to features creates partial representations that distort choice. A consumer focusing on M and neglecting δ acts thrifty. She may reason "I do not need extra goods", even if $b > M$. A consumer focusing on δ and neglecting M may instead reason "it is a good time to buy items I need". She increases her basket and tries to redeem the discount, even if $b < M$. Different representations and errors coexist due to a basic cognitive force: salience driven attention. In this theory, the "M context" of Nonlinear promotions renders total spending salient compared to Linear ones. Thus, it interferes with attention to δ, triggering the choice instability that violates rational axioms.

[14] Problems arise also with loss aversion to spending $b \geq M$. Now consumers would avoid dominated choices and also gain from cutting spending in the left tail compared to Control, which is counterfactual.

4 Selective Attention and Choice

Section 4.1 formalizes how attention shapes choice. Section 4.2 models salience-driven attention and derives the equilibrium CDF of baskets. We derive and test new comparative statics in Section 5. In Section 6 we estimate the model.

4.1 How Attention Causes Choice

Consumer j has utility $u_j(b) + x$, where u_j is increasing, concave, and obeys single crossing (see footnote 4). Perceived utility is shaped by attention to the features of b.

Features reflect choice-relevant attributes of b. We focus on those we experimentally control: the discount, D, and the minimum order value, MOV. Basket b has feature vector $(D(b), MOV(b))$. Discount reports how much the consumer would save if δ was applied, $D(b) = \delta \cdot b$. MOV reports whether the discount condition is met: it is $MOV(b) = 1$ for $b \geq M$ and zero otherwise. In Linear and Discount all baskets meet the condition ($M = 0$), so MOV does not draw attention in either.[15,16]

Attention and Mental Representations. The perceived utility from basket b is:

$$u_j(b) + w - [b - \Delta(b; \boldsymbol{\alpha})], \tag{5}$$

where $\Delta(b; \boldsymbol{\alpha})$ is the perceived discount, $\boldsymbol{\alpha}$ is the attention vector to promotion features, $\boldsymbol{\alpha} = (\alpha_D, \alpha_{MOV})$, where each component is binary. Attention affects the perceived cost of b and thus outside consumption x. The perceived discount is:

[15] Following Bordalo et al. (2025), the vector of b can include the baseline utility $u_j(b)$ and its regular price b. We omit these features because relative attention to them is constant across treatments.

[16] Other features derive from goods in the basket (e.g. design, energy efficiency). Many such features are constant across treatments so we abstract from them. One feature varying across treatments is whether the basket contains a good whose price is above M. Appendix H shows that including this feature relaxes Hicks separability but does not change our key results. In fact it predicts that, compared to Linear, Nonlinear should foster purchases of items whose price is above M. Our data supports this prediction.

$$\Delta(b;\boldsymbol{\alpha}) = (\mathbb{1}_{b\geq M})^{\alpha_{MOV}} \cdot b \cdot \delta \cdot \alpha_D \tag{6}$$

with the convention $0^0 = 1$. Attention to discount, $\alpha_D = 1$, prompts the consumer to focus on the benefit of a lower price. Attention to MOV, $\alpha_{MOV} = 1$, prompts the consumer to focus on regular price spending and on whether $b \geq M$ or not. Under a Linear promotion t_δ, the discount is the only feature that varies across baskets, so attention to it is full $\alpha_D = 1$. Under a Nonlinear promotion $t_{M,\delta}$, two features draw attention, creating three possible representations, $\boldsymbol{\alpha} = (\alpha_D, \alpha_{MOV})$:

$$\Delta(b,\boldsymbol{\alpha}) = \begin{cases} \mathbb{1}_{b\geq M} \cdot \delta \cdot b & if \quad \boldsymbol{\alpha}_F \equiv (1,1) \\ \delta \cdot b & if \quad \boldsymbol{\alpha}_D \equiv (1,0) \\ 0 & if \quad \boldsymbol{\alpha}_{MOV} \equiv (0,1) \end{cases} \tag{7}$$

If the representation is "Full", $\boldsymbol{\alpha}_F = (1,1)$, the consumer attends to both features and correctly perceives the promotion. If the representation is "Discount", $\boldsymbol{\alpha}_D = (1,0)$, the consumer focuses only on the lower than usual price. she perceives $\Delta(b;\boldsymbol{\alpha}_D) = \delta \cdot b$, boosting the value of any basket, including $b < M$. If the representation is "MOV", $\boldsymbol{\alpha}_{MOV} = (0,1)$, she focuses only on the spending cost b. She perceives $\Delta(b;\boldsymbol{\alpha}_{MOV}) = 0$, which reduces her valuation, including of baskets $b \geq M$. Full neglect of δ aids tractability but our results hold as long as $\alpha_D < 1$.[17]

[17] We could allow for more promotion features and attention profiles. The consumer may attend to the jump from C to D in Figure 3, producing anchoring to $b = M$. We exclude this feature as we see little anchoring to M in the data (this is consistent with our model because anchoring requires attention to two features rather than one). We could allow the consumer to attend to nothing, $\boldsymbol{\alpha} = (0,0)$, causing neglect of discount under Linear. This profile would not change the key point: having more features, $t_{M,\delta}$ is more difficult to represent, which produces a stronger neglect of D than Linear.

Choice Behavior. The consumer chooses b subject to her perceived budget:

$$\max u_j(b) + x \tag{8}$$

$$s.t.\ b - \Delta(b; \boldsymbol{\alpha}) + x \leq w_j \tag{9}$$

Under t_δ and t_C she correctly represents the budget set. Under $t_{M,\delta}$, different representations span the budget lines in Figure 3: Linear AB' under $\boldsymbol{\alpha}_D$, Control AB under $\boldsymbol{\alpha}_{MOV}$, and Nonlinear $ACDB'$ under $\boldsymbol{\alpha}_F$.

Representations shape choice. In Linear and Control treatments representations are correct and choice is rational, giving the CDF $B(b|t_i)$, $i = C, \delta$, which we observe. In Nonlinear treatments, the CDF of baskets $B(b|t_{M,\delta}, \boldsymbol{\alpha})$ is rational as in Equation (3) if all consumers use $\boldsymbol{\alpha}_F$ but generally not otherwise. If all consumers use $\boldsymbol{\alpha}_{MOV}$ they behave "as if" they are in Control, so $B(b|t_{M,\delta}, \boldsymbol{\alpha}_{MOV}) = B(b|t_C)$. The representation is wrong but chosen baskets fall within the true budget set. If all consumers use $\boldsymbol{\alpha}_D$ they behave as if they are in Linear, so $B(b|t_{M,\delta}, \boldsymbol{\alpha}_D) = B(b|t_\delta)$. In this case, consumers choosing $b < M$ violate their budget constraint. When redeeming the discount they receive an error message, which we observe in our data.

4.2 Attention Determination

In Bordalo et al. (2025) the salience of a representation is shaped by: 1) bottom up contrast and prominence of its features, and 2) the DM's past reliance on it. We seek to capture basic intuitions, not to have a full model of attention, so we simplify the Bordalo et al. (2025) model. This suffices also for estimation, given our data.

Contrast. Following Bordalo et al.; Bordalo et al. (2012; 2013) and Kőszegi and Szeidl (2013), attention increases in a feature's maximal variation across considered bas-

kets $b \in [\underline{b}, \overline{b}]$. We abstract from individual differences in the consideration set, assuming that $\underline{b} = 0$ and $\overline{b}$ is a large basket (e.g. top 10%). For discount, then, variation is $|\delta \cdot \overline{b} - 0|$ and contrast is an increasing function of it $\sigma_D(\delta \cdot \overline{b}) \equiv \sigma(\delta)$. A larger, more striking δ draws attention. For MOV, variation is maximal $|1 - 0|$ because $\overline{b}$ always qualifies, $\overline{b} > 199$.[18] MOV contrast is constant at $\sigma_{MOV} > 0$.

Prominence. A feature draws attention if it is visually conspicuous, e.g. price is in boldface as in Li and Camerer (2022). Prominence is a scalar ν. The discount δ is on the banner, so its prominence is maximal, $\nu_D = 1$. M is also prominent, $\nu_{MOV} = 1$, when it is on the banner, not otherwise ($\nu_{MOV} = 0$). Variation in the prominence of M may affect attention. Recall, though, that M is also seen when the consumer goes to the promotion page or receives an error message. Placing it on the banner matters for the initial reaction to the promotion, but the representation can subsequently change.

Bottom-up salience is the product between a feature's contrast and prominence, $\sigma \cdot \nu$. Prominently shown prices are especially salient if they vary sharply across options. We denote the bottom-up conditions of a Nonlinear treatment by $\beta = (\delta, \nu_{MOV})$.

Top-down salience. In Bordalo et al. (2025) a frequently used representation is more likely to be retrieved from memory. We often consider hotel prices, so we think about them even if currently shrouded. Our dataset does not report past experiences, so we capture them with a latent scalar κ that we estimate, as in Bordalo et al. (2023). The overall salience of a representation $\boldsymbol{\alpha}_R$ is additive in its top down salience and in the bottom up salience of its features:

$$S_\beta(\boldsymbol{\alpha}_R) = \begin{cases} \kappa + \sigma(\delta) + \sigma_{MOV} \cdot \nu_{MOV} & if \quad \boldsymbol{\alpha}_F \\ \sigma(\delta) & if \quad \boldsymbol{\alpha}_D \\ \sigma_{MOV} \cdot \nu_{MOV} & if \quad \boldsymbol{\alpha}_{MOV} \end{cases} \tag{10}$$

[18] A minuscule M would have zero contrast with $\underline{b} > 0$ and a huge M would have zero contrast, too.

κ captures the familiarity of a Full representation $\boldsymbol{\alpha}_F$ compared to Discount and MOV ones. The relative familiarity of the latter two is indistinguishable from $\sigma(\delta)$, hence we do not parameterize it. Equation (10) can be derived, following Bordalo et al. (2025), from the similarity between a mental representation and its best-matching category of experiences: making good deals, paying regular prices, and nonlinear price experiences.[19]

The cognitive system selects the most salient representation subject to an extreme value attention shock ϵ that is iid across consumers with mean zero and precision 1. Thus, all consumers draw representation R with the same probability:

$$\mu_\beta(\boldsymbol{\alpha}_R) = \frac{e^{S_\beta(\boldsymbol{\alpha}_R)}}{e^{S_\beta(\boldsymbol{\alpha}_D)} + e^{S_\beta(\boldsymbol{\alpha}_{MOV})} + e^{S_\beta(\boldsymbol{\alpha}_F)}} \quad \text{for } R = D, MOV, F. \tag{11}$$

This simple attention model captures two key forces. First, consumers represent the same promotion differently, even if utility is the same. Second, as feature contrast and prominence β change, representations also change for many people, causing instability. As we now show, both forces help make sense of our data.

4.3 Endogenous Observables: Choice and Coupon Errors

Endogenous observables are generated as follows. The banner of a Nonlinear promotion sets the initial representation by Equation (11). Based on it, the consumer picks b_{co} to maximize utility in Equations (8) and (9). If $b_{co} > M$ the consumer buys this basket, so $b = b_{co}$. If $b_{co} < M$ the consumer receives an "M-message" saying "M has not been reached" causing MOV to become prominent. With probability η_{MOV} the surprising MOV interferes with δ, the consumer adjusts attention to $\boldsymbol{\alpha}_{MOV}$, and buys her "Control"

[19] By the similarity function in Bordalo et al. (2025), a description with weights $\boldsymbol{\alpha}_d = (\alpha_{d,D}, \alpha_{d,MOV})$ has similarities $S(\boldsymbol{\alpha}_F, \boldsymbol{\alpha}_d) = \sigma_{MOV} \cdot \alpha_{d,MOV} + \sigma_\delta$, $S(\boldsymbol{\alpha}_D, \boldsymbol{\alpha}_d) = \sigma_{MOV} \cdot (1 - \alpha_{d,MOV}) + \sigma_\delta$ and $S(\boldsymbol{\alpha}_{MOV}, \boldsymbol{\alpha}_d) = \sigma_{MOV} \cdot \alpha_{d,MOV}$ (using distance $d(.) = |.|$). If each representation has a fully fitting category (with similarity normalized to one) and the relative frequency of $\boldsymbol{\alpha}_F$ is κ this gives the salience ranking of Equation (10) by defining $\nu_{MOV} = (2 \cdot \alpha_{MOV} - 1)$.

basket. Else, the consumer integrates M with δ, attending to both, $\boldsymbol{\alpha}_F$, and buys her "rational" basket. We later estimate η_{MOV}.[20]

The probability of receiving an M-message is the probability that the initial representation neglects MOV, $\boldsymbol{\alpha}_D$,times the probability that a consumer with this representation chooses $b \in (0, M)$:

$$\Pr(MOVerror) = \mu_\beta(\boldsymbol{\alpha}_D) \cdot \Pr(b \in (0, M)|t_\delta). \tag{12}$$

This is a key equation because it ties two observables, the frequency of M-messages and choice behavior in t_δ, to mental representations. Consider the other observable, the CDF of b. Let $\boldsymbol{\mu}_{i,\beta}$ be the distribution of initial representations in Equation (11). The probability of shopping under a "MOV" representation at $b < M$ adds the probability of choosing it initially with the probability of adjusting to it after receiving an M-message, so $\widetilde{\mu}_\beta \equiv \mu_\beta(\boldsymbol{\alpha}_{MOV}) + \mu_\beta(\boldsymbol{\alpha}_D) \cdot \eta_{MOV}$. Choice then works as follows.

Proposition 3. *In Nonlinear promotion t_i with $\beta = (\delta, \nu_{MOV})$, the CDF of b obeys:*

$$B(b|\boldsymbol{\mu}_{i,\beta}) = \begin{cases} B\left(b|t_C\right) & \text{if } b \leq \underline{b}_i, \\ B(b|t_C) \cdot \widetilde{\mu}_\beta + B(\underline{b}_i|t_C) \cdot (1 - \widetilde{\mu}_\beta) & \text{if } \underline{b}_i < b < M \\ B(b|t_\delta) \cdot [1 - \mu_\delta(\boldsymbol{\alpha}_{MOV})] + B(b|t_C) \cdot \mu_\beta(\boldsymbol{\alpha}_{MOV}) & \text{if } b \geq M. \end{cases} \tag{13}$$

As in the rational model, spending in $t_{M,\delta}$ is a function of the CDFs under Control t_C and Linear t_δ. Now, however, these CDFs are mixed using weights reflecting different representations. Shifts in attention affect behavior at constant utility, causing choice heterogeneity and instability that is at the heart of of our puzzles.

In the left tail, $b \leq \underline{b}_i$, the CDF is the same as under Control, consistent with rationality. People initially focusing on MOV, $\boldsymbol{\alpha}_{MOV}$, or attending to both features, $\boldsymbol{\alpha}_F$, buy

[20] Not much changes if with probability η_{MOV} the consumer only partially adjusts her basket.

the optimal basket. People initially focusing on discount, $\boldsymbol{\alpha}_D$, initially choose too large a basket b_{co}. As they try to redeem a discount, the M-message prompts them to buy the optimal Control basket. Anomalies arise at larger baskets.

The first departure is positive mass in the bunching region, $b \in [\underline{b}_{M,\delta}, M)$. It is not caused by stable "frictions", but by people who neglect the discount due to interference from M. Some people focus on spending and fail to bunch at M. Other people would rationally set $b > M$ under Linear but "slide back" to $b < M$. These forces account for positive mass right below M (anomaly 1) and "no bunching" (anomaly 2).

Some people bunch at $b = M$, using the correct representation $\boldsymbol{\alpha}_F$. These consumers account for part of the spike at $b = M$ and for part of the shrinkage of the bunching region compared to Control, which we see in the data. Representation are heterogeneous, so some people are sensitive to Nonlinear discounts.

To the near right of M, $b \in [M, M + v]$ some consumers would have rationally set $b >> M$. Due to interference from M, they focus on controlling spending and "move to M from above". They contribute to the spike near M but critically they cause missing right mass at at $b >> M$ (anomaly 3).

Selective attention to the Nonlinear context creates heterogeneous representations explaining seemingly contradictory facts. By cuing some consumers to think about spending, M causes an average stunted reaction to the $same$ discount δ under a Nonlinear versus a Linear promotion, explaining dominated choice, insufficient bunching, missing right tail mass. Other consumers are *too sensitive* to δ: they try to redeem even at $b < M$, producing error messages we observe. Finally, some consumers focus on both features, creating some bunching and avoidance of dominated choices. Critically, these heterogeneous behaviors do not reflect different "types". They are unstable upon changes in feature salience, as we show next.

5 Comparative Statics

Changes in bottom-up salience $\beta = (\delta, \nu_{MOV})$ alter, via representations, observable errors and spending. We focus on basic implications, leaving more complex comparative statics for our quantitative exercise.

Proposition 4. *Higher δ reduces "MOV" representations, $\partial \mu_\beta(\boldsymbol{\alpha}_{MOV})/\partial \delta < 0$, and fosters other representations $\partial \mu_\beta(\boldsymbol{\alpha}_R)/\partial \delta > 0$ for $R = D, F$.*

Higher prominence ν_{MOV} reduces "Discount" representations, $\partial \mu_\beta(\boldsymbol{\alpha}_D)/\partial \delta < 0$, and fosters other representations $\partial \mu_\beta(\boldsymbol{\alpha}_R)/\partial \delta > 0$ for $R = MOV, F$.

Increasing the discount δ draws attention to lower prices, reducing the share of people focused $only$ on M. Placing M on the banner, higher ν_{MOV}, draws attention to it, reducing the share of people initially focusing $only$ on δ. Intuitively, rendering a feature more salient reduces its neglect. In part, this effect promotes accuracy: the incidence of $\boldsymbol{\alpha}_F$ increases in both cases. In part, though, the salient feature interferes with the other, favoring a partial representation. A striking discount, higher δ, interferes with M, fostering $\boldsymbol{\alpha}_D$ and hence overreaction to the promotion. Placing M on the banner, $\nu_{MOV} = 1$, interferes with δ, fostering $\boldsymbol{\alpha}_{MOV}$ and hence neglect of lower prices. Interference is key: making one piece of information salient does not necessarily "debias" people.

Our strongest predictions concern changes in discount contrast, which we fully control. Changes in the prominence of M are harder to trace to final choices, given that in our experiment M can be learned in many ways.

5.1 The Left Tail: Error Messages

In the left tail $b < M$ our comparative statics concern M-messages received by consumers redeeming a discount at $b < M$. Using the CDF of Linear promotion t_δ and the frequency

of these messages we construct an observable proxy for Discount representations:

$$\frac{\Pr(\textit{M-messages})}{\Pr(b \in (0, M)|t_\delta)} = \mu_\beta(\boldsymbol{\alpha}_D) \tag{14}$$

Higher frequency of M-messages relative to the probability of $b \in (0, M)$ under Linear promotion t_δ signals stronger neglect of M. Proposition 4 gives the following result.

Model Prediction 1. (*M-messages*). *The proxy for Discount representations (left hand side of Equation (14)) should increase δ and decrease in prominence ν_{MOV}.*

A striking discount δ and a shrouded M boost $\boldsymbol{\alpha}_D$, increasing our errors based proxy for it. Before testing this prediction, we describe our data. As in many online stores consumers receive error messages for typing the wrong coupon code, trying to redeem an expired promotion, etc. These errors are not part of our theory and we do not observe them. We observe error messages tied to correctly inserted promotion codes that refer to the failure to meet M. Every such message starts with "The minimum order value of M has not been reached."[21] These errors can be tied to the M of our experimental treatments, giving a key input for our proxy in (14).

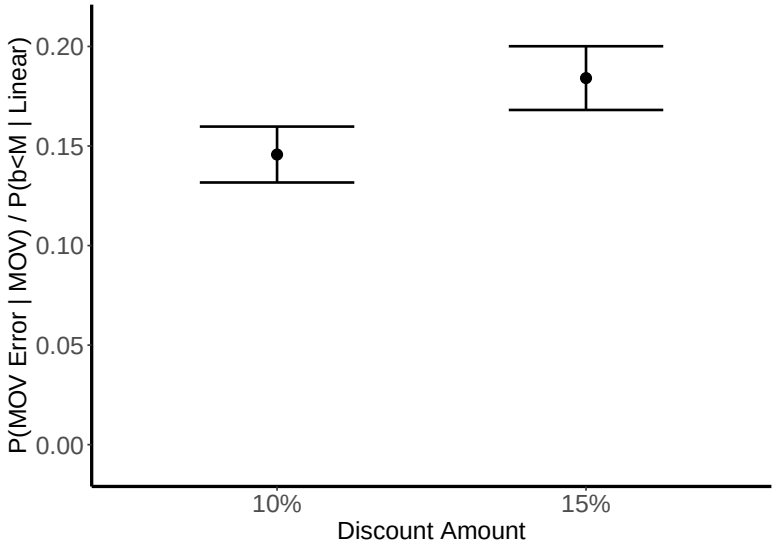
We observe 1583 messages due to failure to meet the M of our Nonlinear promotion treatment. These treatments have ≈ 15000 buyers, so the incidence of M-messages is $\approx 10\%$. For each discount δ we can then compute $\Pr(b \in (0, M)|t_\delta)$ from our Linear treatment and construct the proxy in Equation (14). Figure 7 reports in Panel (a) a split based on discount size, Panel (b) a split based on the prominence of M.

Putting M on the banner reduces by our proxy the probability that consumers initially choose a Discount representation $\boldsymbol{\alpha}_D$. Increasing discount size instead increases the

[21] In rare cases the error also entails an explanatory sentence (e.g. a specific product is not part of the promotion or a specific product is out of stock).

Figure 7: Prevalence of M.messages

(a) M-messages by Discount



(b) M-messages by MOV Salience



Notes: This figure reports our proxy for the incidence of $\boldsymbol{\alpha}_D$. Panel (a) compares $\delta = 10$ versus $\delta = 15$, Panel (b) salient versus shrouded M. The bars represent 95%-CIs.

incidence of MOV errors, consistent with higher initial prevalence of this representation. This latter effect showcases interference: a more striking discount δ causes more people to "over-react" to the promotion, trying to redeem a discount they are not entitled to.[22]

[22] Based on the same mechanism, Bordalo et al. (2023) show in an inference problem that a more striking

The model's comparative statics about positive mass in $\mathcal{D}$ are ambiguous. Higher δ or ν_{MOV} reduce such mass by fostering the correct representation, but increase such mass via interference. Higher δ does it through people who focus on M upon receiving coupon errors, higher ν_{MOV} through outright neglect of δ. In the data, the mass in $\mathcal{D}$ increases from 4.9% when M is shrouded to 5.1% ($p = 0.58$) when M is on the banner. Higher M does not changes $\mathcal{D}$ much, possibly even due to the fact that M is eventually learned anyway. The same mass on the other hand increases from 4.0% when $\delta = 10$ to 6.0% ($p = 0.00$) when $\delta = 15$. The effect is larger. It is consistent with interference and with η_{MOV} large. We later estimate this parameter.

5.2 The Right Tail

In the right tail $b \geq M$, changing β has ambiguous effects on bunching at $b = M$, also due to interference effects. The clear prediction now is that higher δ should reduce missing right tail mass, the violation of IIA.

Model Prediction 2 (*Right tail mass*). *The missing right tail mass in $t_{M\delta}$ compared to t_δ for baskets $b > M$ is smaller when the discount δ is larger, $\partial X_{M,\delta}(\theta)/\partial\delta > 0$.*

Higher discount promotes neglect of M, boosting sensitivity to δ and rational behavior at $b \geq M$. Crucially, they key effect is not that a larger discount boosts spending by lowering its marginal cost, which occurs in both the Nonlinear and Linear promotions. The key effect is that a striking discount draws attention toward itself and away from M in the Nonlinear promotion, increasing consumers' *price elasticity* in $t_{M,\delta}$ compared to t_δ. This effect reduces violations of IIA and fosters bunching at $b \geq M$.[23]

signal causes more people to focus on it and neglect the prior, fostering belief overreaction.

[23]Higher prominence ν_{MOV} of M boosts sensitivity to δ by fostering the correct representation $\boldsymbol{\alpha}_F$,

Figure 8 reports in Panel (a) the excess mass above $M + \theta$ across the 10% and 15% discount treatments. Consistent with the model, the missing right tail mass is present when $\delta = 10\%$ but it becomes smaller and insignificant when the discount is 15%, also in the bootstrap test in Panels (b) and (c) (at 10% $p = 0.006$, at 15% $p = 0.588$).

The same holds for no bunching. With a 10% discount Prediction 2 ($X(0) > 0$) is rejected with a p-value of 0.023. With a 15% discount the p-value is 0.525, so the rational null is not rejected. See Table 7 in the Appendix. Similarly, we find a significant right tail revenue of $113,000$ EUR ($p = 0.01$) in the 10% Nonlinear promotions, and an insignificant $13,000$ EUR ($p = 0.63$) in the 15% Nonlinear promotions.

Overall, in line with our core mechanism promotion context affects the salience of different features, shaping attention and choice instability. We now estimate the model.

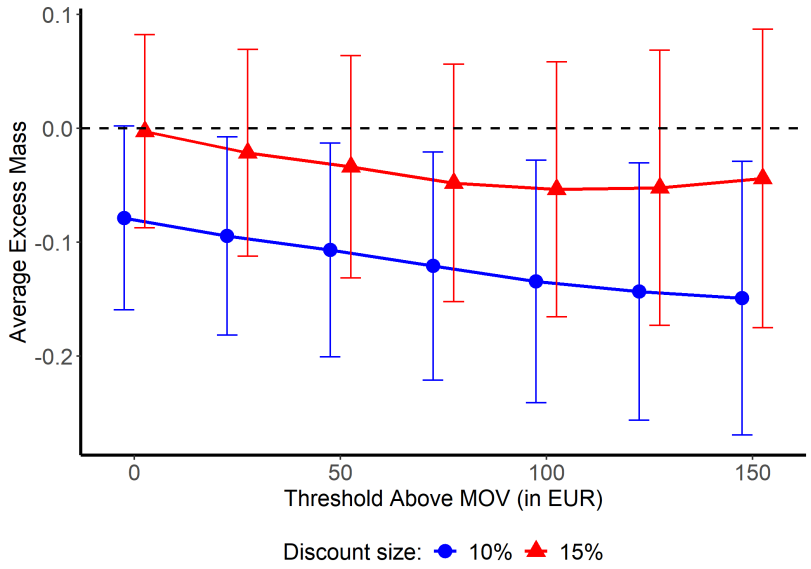
6 Model Estimation and Welfare

We estimate salience by relying on M-messages and evaluate its choice and welfare effects. We do not seek to fit a "best practice" attention model, given that we only observe one simple proxy for attention. Future work may rely on more extensive measurement, including mouse tracking, visual salience (Li and Camerer, 2022), reasoning (Haaland et al., 2024), etc. Here we however show that even our simple proxy can shed significant light on choice puzzles. Once salience and hence the representations $\boldsymbol{\mu}_{i,\beta}$ is pinned down, there are no degrees of freedom in our model: attention weights mix behavior in "pure" Linear and Control conditions, which we observe in our experiment. behavior is stable across contexts sharing the same representation.

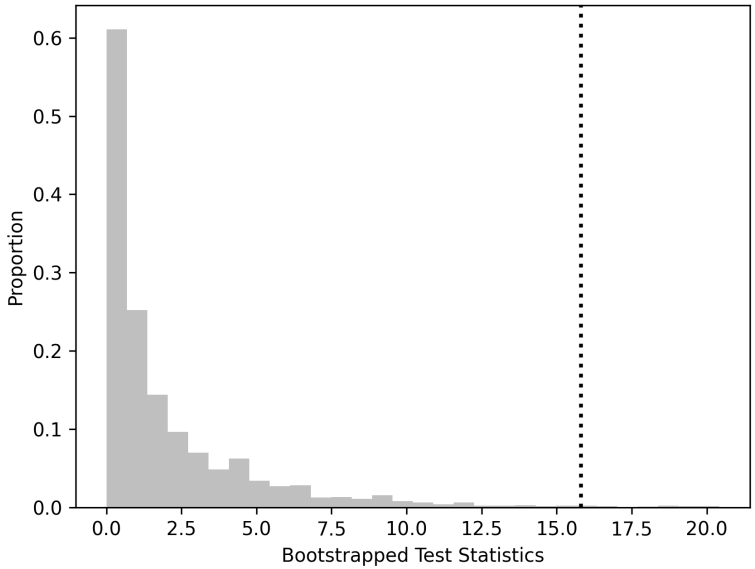
reduces it by fostering $\boldsymbol{\alpha}_{MOV}$. In our model the latter force wins, but the effect is generally ambiguous.

Figure 8: Right Tail Test, by Discount Size

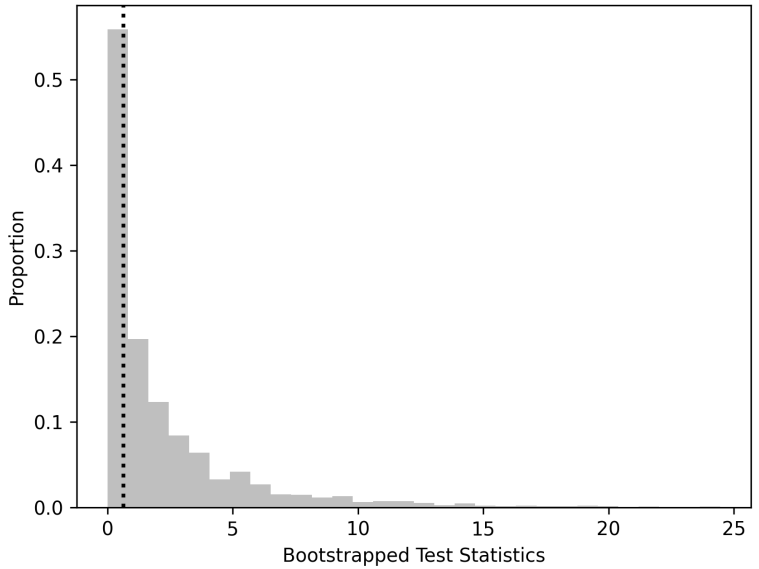
(a) Missing Mass



(b) 10% Off Bootstrap



(c) 15% Off Bootstrap



Notes: This figure compares missing upper tail mass for the 10% discount versus the 15% discount. Panel (a) plots the excess mass at different thresholds, with vertical bars reporting 95% CIs. Panels (b) and (b) contain the bootstrap tests for the 10% and 15% discounts respectively.

6.1 Estimation of Salience Parameters

In Nonlinear promotion $i = (M, \delta)$ the distribution of initial representations $\boldsymbol{\mu}_{i,\beta}$ in Equation (11) is pinned down by four parameters: contrast and prominence of MOV, σ_M if MOV is prominent and zero otherwise; Contrast of Discount, $\sigma(0.1)$ when $\delta = 10\%$ and $\sigma(0.15)$ when $\delta = 15\%$; Top-down salience of the correct representation (κ). A fifth parameter sets the representation after receiving an M-message: the probability η_{MOV} with which this consumer refocuses only on MOV, $\boldsymbol{\alpha}_{MOV}$, or on both features, $\boldsymbol{\alpha}_F$.

Our M-message moments are given by:

$$\frac{\Pr(\textit{MOV error})}{\Pr(b \in (0, M)|t_\delta)(1-h)} - \mu_\beta(\boldsymbol{\alpha}_D) = 0. \tag{15}$$

The only difference with Equation (14) is that we adjust the ratio on the left (which we observe), by the probability with which consumers may fail to redeem discounts due to distraction, forgetting, loss of the promotion code, etc. Due to such events, the frequency of errors reflects only in part the frequency of the representation causing them.[24] In our experiment we observe these redemption failures from Linear treatments t_δ, allowing us to compute $h = \Pr(\text{not redeem}|t_\delta, b \in [0, M))$. Moment condition (15) then reflects the fact that MOV errors correspond to Discount representations by the $(1-h)$ redeemers. This adjustment improves accuracy, but our main results go through with $h = 0$. There are eight combination of (15) across $\delta = 10, 15\%$, $M = 99, 199$ EUR, and $\nu_{MOV} = 0, 1$.

η_{MOV} is the probability that in treatment $t_{M,\delta}$ a consumer stays in the bunching region $\mathcal{B} = [\underline{b}_{M,\delta}, M)$ upon receipt of a MOV error. Our second moment condition thus uses the mass in this region, normalized by its mass in Control t_C (which homogenizes the mass across promotions). This ratio is the leftmost term in the moment condition:

[24] Failure to redeem due to hassle has been studied, in the same online shop, by Rodemeier (2024). He finds that hassle reduces the redemption probability without affecting the probability to buy in the store.

$$\frac{\Pr(b \in [\underline{b}_{M,\delta}, M)|t_{M\delta})}{\Pr(b \in [\underline{b}_{M,\delta}, M)|t_C)} - \mu_\beta(\boldsymbol{\alpha}_{MOV}) - \frac{\Pr(b \in [\underline{b}_{M,\delta}, b^*_{M,\delta}]|t_{M\delta})}{\Pr(b \in [\underline{b}_{M,\delta}, b^*_{M,\delta}]|t_C)} \cdot (\eta_{MOV}(1-h)+h)\mu_\beta(\boldsymbol{\alpha}_D) = 0 \quad (16)$$

By Equation (3), the relative mass in $\mathcal{B}$ is the share of consumers focusing on MOV, term $\mu_\beta(\boldsymbol{\alpha}_{MOV})$, and the share of consumers focusing on δ and spending in $\mathcal{B}$, the third term. A share $(1-h)$ of them redeems, gets an error, and stays in $\mathcal{B}$ with probability η_{MOV}. A share h of them does not even try to redeem.

To compute the ratios in (16) we run an OLS regression of baskets on discount in Linear promotions t_δ, $b_i = \gamma + \psi \cdot \delta_i + \epsilon_i$, which allows us to compute thresholds $\underline{b}_{M,\delta}$, the lower limit of the bunching region, and $b^*_{M,\delta}$ the basket chosen under Control by a consumer choosing $b = M$ under Linear t_δ. See Appenidx J for details [25]

Using these estimates we compute the ratios in (15) and obtain eight additional moments. Primitives are $\zeta \equiv (\sigma_\delta(0.1), \sigma_\delta(0.15), \sigma_M, \kappa, \eta_M)$. We estimate them by solving:

$$\begin{aligned} \hat{\zeta} = &\underset{\zeta}{\text{argmin}} \sum_{i,\beta} \sum_{r \in \{err, bunch\}} \omega^r_{i,\beta} Z^r_{i,\beta}(\zeta)^2 \\ &\text{s.t. } \eta_{MOV} \in [0,1], \end{aligned} \quad (17)$$

where $Z^{err}_{i,\beta}$ is the left hand side of Equation (15) in treatment $i = (M, \delta)$ under contrast and prominence conditions β, while $Z^{bunch}_{i,\beta}$ is the left hand side of Equation (16). The weights ω are equal to the inverse variance of each of the moments. We compute GMM standard errors according to the sandwich formula.

Point estimates and standard errors are in Table 2. Estimates are highly significant.

[25] The OLS recovers a structural price elasticity because consumers properly represent Linear promotions. To increase precision, for each promotion we consider only baskets below the 0.995^{th} percentile.

The largest parameter, $\sigma_{MOV} = 2.10$, suggests that a visually prominent MOV attracts more attention than any of the discounts, also consistent with greater familiarity of paying regular compared to discounted ones. As for discount salience, $\sigma(0.1) = 0.26$ is smaller than $\sigma(0.15) = 0.49$, consistent with contrast: a more striking discount draws attention bottom up. $\kappa = -2.41$ suggests that consumers are relatively unfamiliar with Nonlinear pricing, which reduces the likelihood of them correctly representing the promotion. Finally, and interesting, η_{MOV} is close (here rounded) to one: upon receiving an error message, many subjects focus on total spending and stay below M. This result, due to the stability of the dominated region across Nonlinear and Control, also suggests that information (here through error) may not lead to correct representations.

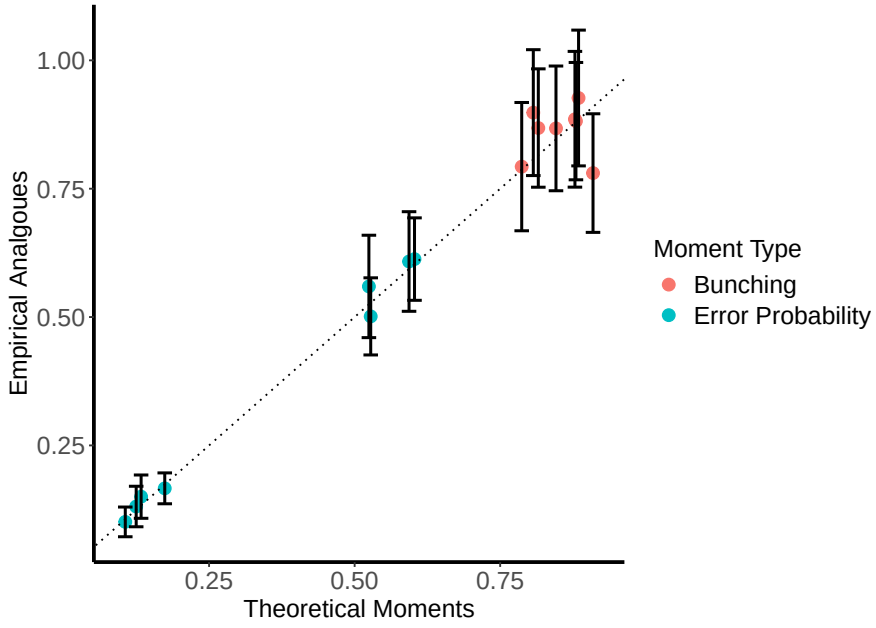
Salience point estimates are driven by coupon errors. They are almost unchanged, but noisier, if we use *only* the error moment condition (15). Data directly linked to faulty representations - and only indirectly linked to choice - is key to to pin down salience parameters. This of course also implies that our test of the model is demanding.

Table 2: Salience Parameter Estimates

Parameter	Estimate
σ_{MOV}	2.10
	(0.10)
$\sigma(0.1)$	0.28
	(0.14)
$\sigma(0.15)$	0.50
	(0.14)
η_{MOV}	1.00
	(0.17)
κ	-2.39
	(0.36)

Notes: This table reports estimates of the salience model from the GMM estimation in equation 17.

Figure 9: Model Fit



Notes: This figures compares the theoretical moments targeted by the structural estimation (the abscissa) with the empirically observed value (the ordinate). The dashed line shows the 45 degree line.

6.2 In-Sample Model Fit

To assess explanatory power, consider moment fit first. Figure 9, plots empirical targets and their standard errors (normalized errors and mass in $\mathcal{B}$ (Equations (15) and (16), respectively) against their model-implied analogue. The model does a good job at fitting 16 empirical moments with 5 parameters only. Thus, the model reproduces well the behavior of MOV errors in Figure 7

6.3 Predicting Consumption with Non-Choice Data

We turn to a more ambitious "out-of-sample" validation. We show that salience can reproduce key features in different part of the CDF and qualitatively captures the instability of the right tail upon changes in δ documented in (Section 5.2).

As a first step, we use salience parameters to compute the distribution of representations, reported in Table 3. The first column shows that, across treatments, only 8% of consumers correctly represent Nonlinear promotions. Most consumers, 57%, focus only on M and neglect δ while the remaining 35% focus solely on δ, neglecting M.

In columns 2 and 3, the model reproduces some instability across the 10% versus 15% discounts linked to missing right tail mass. A more striking discount reduces the share of consumers focusing only on M by 4.5 percentage points, moving most of those consumers toward Discount $\mu(\boldsymbol{\alpha}_D)$, while the share of correct representations remains fairly small. Again, interference is an important force in the model.

Table 3: Distribution of Attention Profiles

	Aggregate	**10% Off**	**15% Off**
$\mu(\boldsymbol{\alpha}_D)$	0.35	0.33	0.37
$\mu(\boldsymbol{\alpha}_{MOV})$	0.57	0.59	0.55
$\mu(\boldsymbol{\alpha}_F)$	0.08	0.08	0.08

Notes: This table presents the implied distribution of attention profiles across different discount conditions.

Given the estimated representations, choice behavior in $t_{M,\delta}$ is determined, based on Equation (3), from the observed behavior in Linear t_δ and Control t_C treatments. We assess our key empirical findings: i) dominated choices (fact 1), ii) bunching behavior (facts 2 and 4), and iii) missing right tail mass (facts 3 and 5). We measure bunching by excess bunching mass $BU(v) \equiv \Pr(b \leq M + v) - \Pr(b \leq M)$ compared to Linear as defined in Equation (23). We measure dominated choices by the mass in $\mathcal{D}$ compared to Control. We then compute a proxy for the missing right tail that ties to representations: the share of $\boldsymbol{\alpha}_{MOV}$ that reproduces the observed missing mass:

$$\mu(\boldsymbol{\alpha}_{MOV}) = -\frac{X(\theta)}{1 - \frac{\Pr(b \geq M+\theta|t_C)}{\Pr(b \geq M+\theta|t_\delta)}}, \tag{18}$$

which follows from (3), where $X(\theta)$ is excess right tail mass in Nonlinear $t_{M,\delta}$ compared to Linear t_δ as defined in Equation (4). In the model the presence of missing mass $X(\theta) < 0$ corresponds to a positive mass of people neglecting δ, $\mu(\boldsymbol{\alpha}_{MOV}) > 0$. This proxy for missing mass can be computed from observed CDFs in our treatments.

Table 4 reports empirical and model implied statistics, averaged across the eight MOV treatments. The model produces realistic bunching: an excess 0.16% of baskets near M in Nonlinear compared to Linear relative to 0.14% in the data. In our model this is produced by two attentional forces influencing consumer types locating around M: 8% of consumers just below M who have correct representations and set $b = M$, and 57% of consumers above M who neglect δ and slide back towards M.

The model does a fair job at predicting incidence in the dominated region as found in Figure 5. The theoretical estimate is 0.64% of control group incidence, and the empirical estimate equals 0.83%. In the model there are no frictions, so representations alone can go a long way toward explaining dominated choices, even in our simple model. Finally, the model matches fairly well also the missing right tail mass. Consumption data "want" 44% of consumers to focus only on M, and the model implied estimate for this profile from error data is 57%, higher but in the right ballpark.

The model can qualitatively capture also the instability of upper tail behavior when moving from a 10% to a 15% discount documented in Section 5.2. In the data, bunching increases from 0.08 to 0.19 when the discount changes from 10 to 15%, in the model it increases from 0.13 to 0.19. The upper tail mass data require a large reduction in MOV representations, from 0.8 with a 10% discount to 0.06 with 15%. The movement in the estimated model is directionally correct but much smaller, corresponding to a 4.5 percentage point reduction. The improvement in behavior for $b \geq M$ reflects the fact that a more striking discount draws attention, increasing consumers' sensitivity to it.[26]

[26]The reduction in missing mass predicted by the model is arguably larger if higher consumer types find

The model does not instead capture instability below M. The increase in the discount is associated with an increase in dominated choices from 0.73 to 0.9 in the data, while the model implies a very minor decrease from 0.62 to 0.599 (see Table X in the appendix). As we saw in Section 5.1 the predictions of the model for $\mathcal{D}$ can go either way.

Salience estimated from M-messages and the bunching region below M do a good job at matching dominated choices, bunching at M, and upper tail missing mass.

Table 4: Model Validation for Moments not Targeted by Estimation

Quantity	Empirical Value	Predicted Value
Bunching in MOV (% of Linear)	0.144	0.165
	(0.045)	
Incidence in Dominated Region (% Of Control)	0.834	0.641
	(0.053)	
Upper Tail Missing Mass Ratio	0.440	0.572
	(0.259)	
Upper Tail Missing Mass Ratio, $\theta = 0$	0.217	0.572
	(0.150)	

Notes: This table reports the predicted and empirical statistics of our three key behavioral responses: bunching, incidence in the dominated region, and missing mass in the upper tail of the spending distribution. These moments were not targeted by our empirical estimation that largely leveraged coupon redemption errors. Column 1 reports the empiricially observed value of each moment, while column 2 shows the predicted value from the model.

6.4 Welfare Loss from Incomplete Representations

We assess the welfare costs of partial representations. We recover consumers' hedonics, their utility from b, by assuming the quasilinear form:

$$u_j(b) = u(b) + j \cdot b, \quad j \sim F_j \tag{19}$$

the higher discount more contrasting compared to consumers who spend less.

Marginal utility is *additively separable* in consumer type and b. The consumer's FOC given linear prices is $u'(b) + j = p$, so j shifts demand horizontally. We estimate these parameters by comparing Control with Linear (which are correctly represented). Bodoh-Creed et al. (2023) show, using techniques from D'Haultfoeuille and Février (2011), that our linear price treatments can nonparametrically identify u and F_θ. We offer a formal explanation in Appendix K.

We ask: How much does salience harm consumers? To what extent does the firm benefit? In Nonlinear treatment $t_{M,\delta}$ with salience β a consumer j obtains average surplus $CS(j)$ and the firm obtains revenue from j of $Pi(j)$ equal to:

$$CS(j) = \sum_{R=D,MOV,F} [u(b_R^j) + j \cdot b_R^j - p(b_R^j)] \cdot \widetilde{\mu}_\beta(\boldsymbol{\alpha}_R) \tag{20}$$

$$\Pi(j) = \sum_{R=D,MOV,F} p(b_R) \cdot \widetilde{\mu}_\beta(\boldsymbol{\alpha}_R) \tag{21}$$

where b_R^j is j's optimal basket under representation R, $\widetilde{\mu}_\beta(\boldsymbol{\alpha}_R)$ is the probability of such representation (after adjusting to M-messages), and $p(b_R^j) \equiv (1-\delta) \cdot b_R^j \cdot \mathbf{1}_{b_R^j \geq M} + b_R^j \cdot \mathbf{1}_{b_R^j < M}$ is the price of the basket. Aggregate consumer surplus and revenue are then

$$CS = \int_j CS(j)\, \mathrm{d}F_j(j), \quad \Pi = \int_j \Pi(j)\, \mathrm{d}F_i(j). \tag{22}$$

Figure 10 reports the loss/gain in consumer surplus and firm revenues under salience *compared to rationality*. The welfare loss depends on attention and on the distribution of consumer types. Salience reduces surplus between 5% and 14%, depending on the promotion. These are sizable losses that cannot be ignored in welfare analysis.

Consumer welfare losses are large when i) the discount is high, because the cost of foregoing the promotion is large, and ii) M is low, because more consumers are near M,

so more people fail to bunch. Placing M on the banner does not necessarily improve welfare, and in fact it tends to reduce it (by fostering neglect of δ), reducing consumers' ability to benefit from the promotion.

The implications for the firm are more ambiguous. Under most promotions, the firm loses revenue because consumers spend too little (in contrast to the conventional wisdom that firms benefit from biases). Our firm usually runs promotions that give a discount around 10%. In our estimates, salience reduces firm revenue in three out of four of them, compared to the same promotion with rational types. Multi-dimensional schedules can thus trigger significant losses to consumers and the firm.

Figure 10: Welfare Effects of Salience



Notes: This figure reports the effect of consumer misoptimization on consumer surplus (CS) and revenue, separately for each promotion.

7 Conclusions

In a field experiment randomizing Nonlinear and Linear promotions we document context dependent price elasticity, which produces significant violations of revealed preference and rejects standard frictional accounts of anomalies in the field as well as conventional behavioral biases. We show that the data can be explained by salience-driven attention to the features of Nonlinear schedules. The presence of a minimum order value in Nonlinear promotions cues consumers to to think about total spending, triggering focus on the pain of paying and reducing sensitivity to discounts.

Our results have significant implications for applied economics. Context-dependent price elasticity implies that choice-based statistics may not be sufficient to recover invariant structural parameters because they depend on features of context, including normatively irrelevant ones, and on their salience. As an example, the response of consumers to health insurance premia may depend on whether deductible or health risks are more or less salient in the context or choice, and the elasticity of labor supply to taxes may depend on how the salience of different brackets, as shaped by media or the social environment (Chetty et al., 2013). Critically, one key property of salience is interference: making a piece of information more prominent can interfere with attention to previously attended to features, causing a suboptimal reaction.

Going forward, it may be useful to develop methods that measure and correct for salience driven context effects. This is feasible, because attention and context can be formalized and measured, and the use of such non-choice data can help explain behavior in a more robust and unified way. Recent work shows the promise of this approach for explaining context dependence in laboratory statistical problems and risky choices (Bordalo et al., 2023; Bohren et al., 2024). Here we offer a first pass at this approach in a field context, where we proxied for bottom up attention using data on coupon error messages.

Future work should develop more extensive measures of choice context, including but not limited to visual layout, and for attention. These methods may may help explain why informing taxpayers about nonlinearities in the tax schedule does not improve choices, while encouraging them to take advantage of incentives does (Chetty and Saez, 2013), why information on energy costs works in different directions depending on whether it cues energy expenses or savings (d'Adda et al. 2022, Rodemeier and Löschel 2025), and why asking people to set a time for a flu shot (Milkman et al., 2011) can be more effective than information about the benefits of a vaccine (Nyhan and Reifler, 2015). Our findings advocate for taking salience into account in economic analysis and policy design.

References

Abaluck, J. and Adams-Prassl, A. (2021). What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses. *The Quarterly Journal of Economics*, 136(3):1611–1663.

Abaluck, J. and Gruber, J. (2011). Choice inconsistencies among the elderly: evidence from plan choice in the medicare part d program. *American Economic Review*, 101(4):1180–1210.

Allcott, H. and Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *The American Economic Review*, 105(8):2501–2538.

Allen, E. J., Dechow, P. M., Pope, D. G., and Wu, G. (2017). Reference-dependent preferences: Evidence from marathon runners. *Management Science*, 63(6):1657–1672.

Ariely, D., Loewenstein, G., and Prelec, D. (2003). "coherent arbitrariness": Sta-

ble demand curves without stable preferences. *The Quarterly journal of economics*, 118(1):73–106.

Baicker, K., Mullainathan, S., and Schwartzstein, J. (2015). Behavioral hazard in health insurance. *The Quarterly Journal of Economics*, 130(4):1623–1667.

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104.

Bartels, D. M., Herzog, N., and Sussman, A. B. (2023). Distinguishing between anchors and targets. *Available at SSRN 4521038*.

Bhargava, S., Loewenstein, G., and Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, 132(3):1319–1372.

Bodoh-Creed, A., Hickman, B. R., List, J. A., Muir, I., and Sun, G. (2023). Stress testing a structural model of nonlinear pricing: Robust inference on intensive-margin consumer demand. Working paper.

Bohren, J. A., Hascher, J., Imas, A., Ungeheuer, M., and Weber, M. (2024). A cognitive foundation for perceiving uncertainty. Working paper, National Bureau of Economic Research.

Bordalo, P., Conlon, J., Gennaioli, N., Kwon, S., and Shleifer, A. (2023). How people use statistics. Working paper, National Bureau of Economic Research.

Bordalo, P., Gennaioli, N., Lanzani, G., and Shleifer, A. (2025). A cognitive theory of reasoning and choice. Working paper, Becker Friedman Institute for Economics.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2020). Memory, attention, and choice. *The Quarterly Journal of Economics*, 135(3):1399–1442.

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The quarterly journal of economics*, 126(2):749–804.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review*, 103(7):2683–2721.

Chetty, R. and Saez, E. (2013). Teaching the tax code: Earnings responses to an experiment with eitc recipients. *American Economic Journal: Applied Economics*, 5(1):1–31.

de Clippel, G., Oprea, R., and Rozen, K. (2025). As if. Working paper.

D'Haultfoeuille, X. and Février, P. (2011). The provision of wage incentives: A structural estimation using contracts variation. *Mimeo, CREST*, 2011-29.

Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2009). Do switching costs make markets less competitive? *Journal of Marketing research*, 46(4):435–445.

d'Adda, G., Gao, Y., and Tavoni, M. (2022). A randomized trial of energy cost information provision alongside energy-efficiency classes for refrigerator purchases. *Nature Energy*, 7(4):360–368.

Feldman, N. E. and Katuščák, P. (2006). *Should the Average Tax Rate be Marginalized?* CERGE-EI.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gars, J., Holz, J., Rodemeier, M., and Villa, J. M. (2025). Gambling for Retirement: The Economics of Savings Lotteries. Working paper.

Grubb, M. D. and Osborne, M. (2015). Cellular service demand: Biased beliefs, learning, and bill shock. *American Economic Review*, 105(1):234–71.

Haaland, I. K., Roth, C., Stantcheva, S., and Wohlfart, J. (2024). Measuring what is top of mind. Working paper, National Bureau of Economic Research.

Handel, B. and Schwartzstein, J. (2018). Frictions or mental gaps: what's behind the information we (don't) use and when do we care? *Journal of Economic Perspectives*, 32(1):155–178.

Handel, B. R. and Kolstad, J. T. (2015). Health insurance for humans: Information frictions, plan choice, and consumer welfare. *American Economic Review*, 105(8):2449–2500.

Hortaçsu, A. and Syverson, C. (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *The Quarterly journal of economics*, 119(2):403–456.

Ito, K. (2010). How do consumers respond to nonlinear pricing? Evidence from household electricity demand. Working paper, Energy Institute at Haas.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 47(2):263–291.

Kamenica, E. (2008). Contextual inference in markets: On the informational content of product lines. *American Economic Review*, 98(5):2127–2149.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8:435–464.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *The quarterly journal of economics*, 128(2):669–723.

Kőszegi, B. and Szeidl, A. (2013). A model of focusing in economic choice. *The Quarterly journal of economics*, 128(1):53–104.

Kreps, D. (1988). *Notes on the Theory of Choice.* Westview Press.

Li, X. and Camerer, C. F. (2022). Predictable effects of visual salience in experimental decisions and games. *The Quarterly Journal of Economics*, 137(3):1849–1900.

Liebman, J. B. and Zeckhauser, R. J. (2004). Schmeduling. Working paper.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, 108(26):10415–10420.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208.

Moffitt, R. (1990). The econometrics of kinked budget constraints. *Journal of Economic Perspectives*, 4(2):119–139.

Nyhan, B. and Reifler, J. (2015). Does correcting myths about the flu vaccine work? an experimental evaluation of the effects of corrective information. *Vaccine*, 33(3):459–464.

Rees-Jones, A. and Taubinsky, D. (2020). Measuring "schmeduling". *The Review of Economic Studies*, 87(5):2399–2438.

Rodemeier, M. (2024). Buy baits and consumer sophistication: Field evidence from instant rebates. *American Economic Journal: Economic Policy*.

Rodemeier, M. and Löschel, A. (2025). Information nudges, subsidies, and crowding out of attention: Field evidence from energy efficiency investments. *Journal of the European Economic Association*.

Saez, E. (2010). Do taxpayers bunch at kink points? *American economic Journal: economic policy*, 2(3):180–212.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. In *The Foundations of Price Theory Vol 4*, pages 101–116. Routledge.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.

Tversky, A. and Thaler, R. H. (1990). Anomalies: preference reversals. *Journal of Economic Perspectives*, 4(2):201–211.

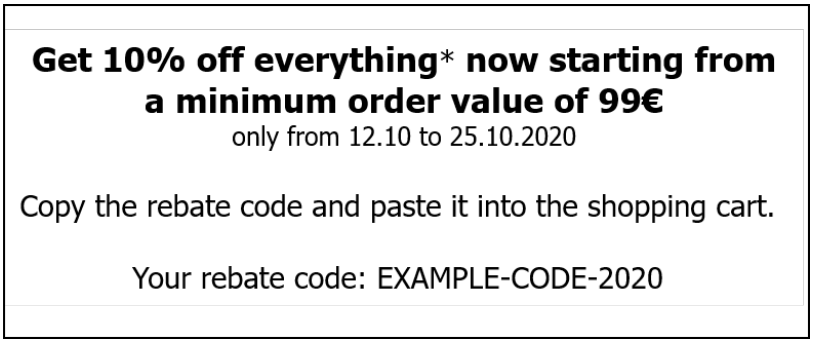
Varian, H. R. (2006). Revealed preference. *Samuelsonian economics and the twenty-first century*, pages 99–115.

Woodford, M. (2012). Prospect theory as efficient perceptual distortion. *American Economic Review*, 102(3):41–46.

Online Appendix

A Additional Figures

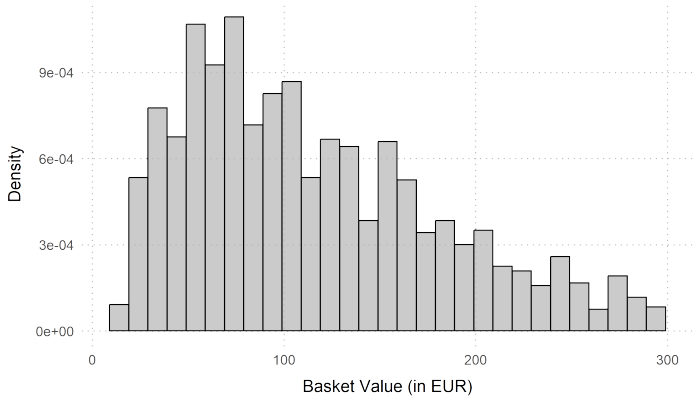
Figure 11: Promotion page for 10% starting from 99 EUR after subject clicked on banner.



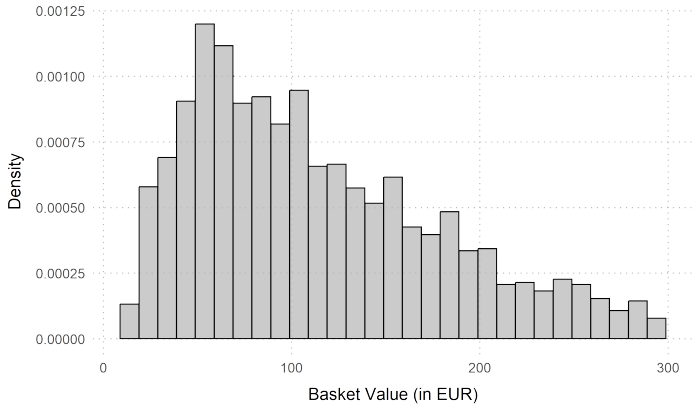
Notes: This Figure shows an example of the Terms & Conditions page which subjects can reach by clicking on the promotion banner in Figure 2, Panel B.

Figure 12: Spending Distributions in Control and Linear

(a) Control

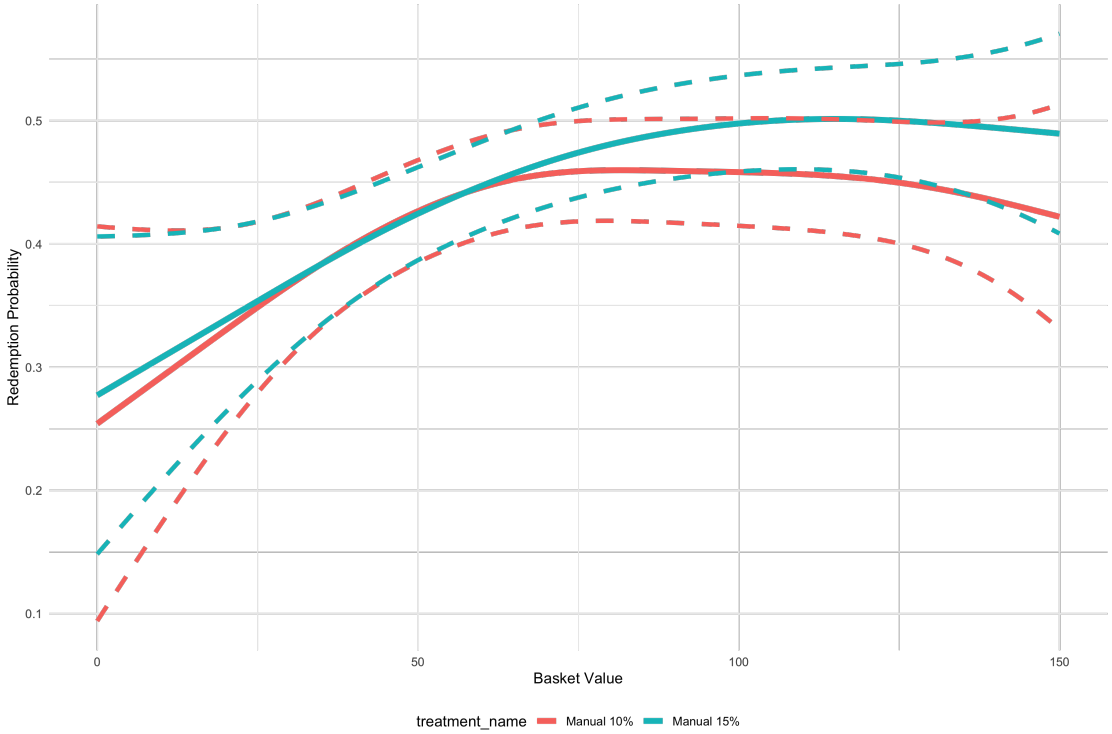


(b) Linear



Notes: This Figure plots the density of baskets in Control (Panel a) and Linear (Panel b). Blue bars indicate the MOVs from the Nonlinear treatments, which are immaterial in Control and Linear. We exclude the mass $b = 0$ and $b > 300$. The first bin starts 10 and ends at 19, and bins continue in 10 EUR increments.

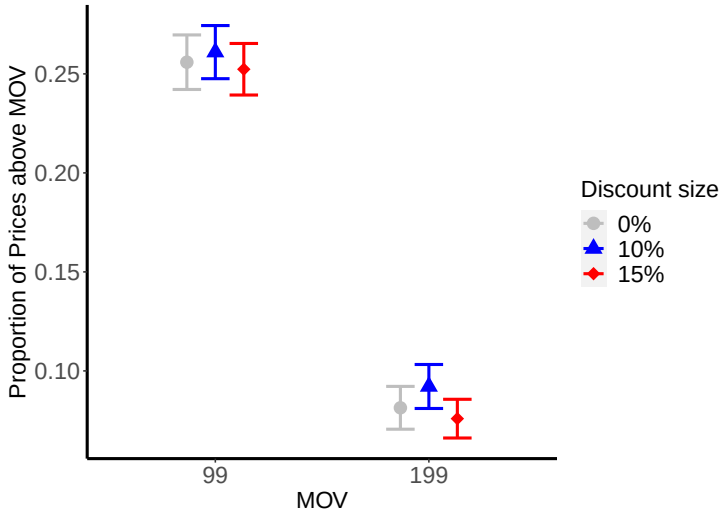
Figure 13: Hassle Costs across the Spending Distribution



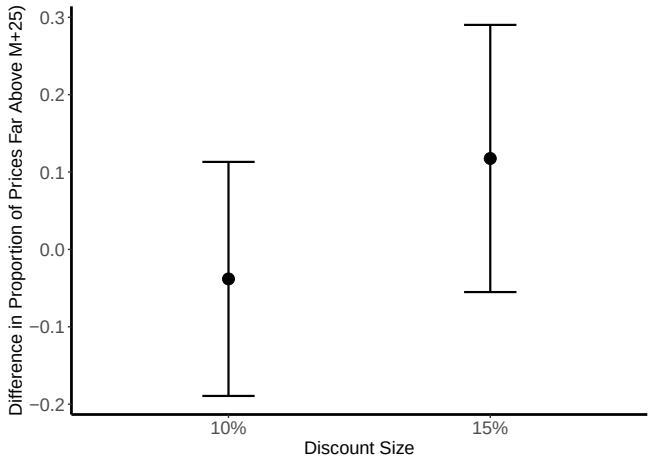
Notes: This Figure shows the coupon redemption probability as a function of basket value, separately for the 10% and 15% Linear promotion. This illustrates the importance of hassle costs that keep consumers from redeeming the coupon. Since all consumers receive a redemption reminder during checkout, inattention in the form of *forgetting* to redeem should be mitigated in our experiment. Dashed lines are 95% confidence intervals.

Figure 14: Salience Model Prediction 6: Price Anchoring above Threshold

(a) Share of Prices above M

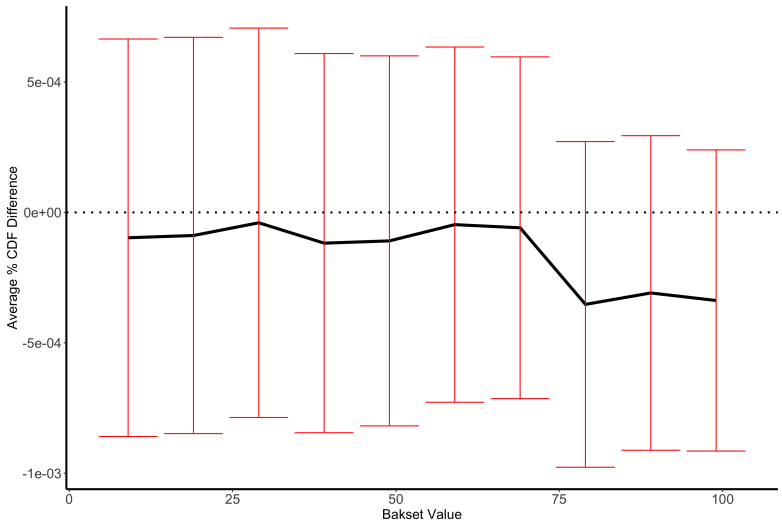


(b) Comparison of Prices far above M



Notes: This figure illustrates price anchoring across treatment conditions. Panel A shows the proportion of prices that are above the threshold M, separetly for each MOV and discount combination. Panel B shows the difference in proportions for prices far above M (defined as M+25). We plot the differences for the 10% and 15% discount separetly. The bars represent 95%-CIs.

Figure 15: Difference in Share of Prices above M across Discount Size



Notes: This figure plots the percent difference between the Control and MOV CDFs for basket values below 99 Euros.

B Additional Tables

Table 5 reports means and standard deviations of observables across conditions: i) the number of sessions, ii) the share of sessions in which a purchase occurs, iii) the share of subjects purchasing from Germany, iv) gender, v) age, vi) the share of "business to business" transactions, vii) the share of transactions cancelled ex-post, and ix) the internet browser used. Characteristics iii) to vii) are only observed for subjects who make a purchase.

Table 5: Summary Table

Variable	Control	10%	10%,99€, salient	10%,99€	10%,199€, salient	10% 199€
Number of sessions	1.847	1.839	1.824	1.825	1.823	1.830
	(2.024)	(2.007)	(1.988)	(1.954)	(2.075)	(2.004)
Purchase (Yes=1)	0.016	0.017	0.016	0.016	0.016	0.016
	(0.127)	(0.131)	(0.126)	(0.124)	(0.124)	(0.125)
Customer from Germany (Yes=1)	0.918	0.916	0.912	0.909	0.907	0.920
	(0.275)	(0.278)	(0.283)	(0.288)	(0.291)	(0.272)
Female (Yes=1)	0.504	0.505	0.489	0.519	0.503	0.490
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Age (in years)	52.513	52.272	52.207	52.051	52.260	51.183
	(14.863)	(14.417)	(14.038)	(14.703)	(14.805)	(14.388)
B2B sales channel (Yes=1)	0.039	0.043	0.040	0.039	0.039	0.045
	(0.194)	(0.203)	(0.197)	(0.195)	(0.193)	(0.208)
Canceled transaction ex-post (Yes=1)	0.013	0.017	0.014	0.016	0.012	0.014
	(0.114)	(0.130)	(0.117)	(0.125)	(0.109)	(0.117)
Internet explorer (Yes=1)	0.012	0.012	0.013	0.012	0.012	0.013
	(0.111)	(0.111)	(0.111)	(0.110)	(0.110)	(0.112)
Firefox (Yes=1)	0.108	0.104	0.110	0.107	0.108	0.104
	(0.310)	(0.305)	(0.313)	(0.309)	(0.311)	(0.305)
Chrome (Yes=1)	0.363	0.368	0.369	0.366	0.364	0.367
	(0.481)	(0.482)	(0.483)	(0.482)	(0.481)	(0.482)
Safari (Yes=1)	0.349	0.349	0.339	0.352	0.346	0.347
	(0.477)	(0.477)	(0.473)	(0.478)	(0.476)	(0.476)
Observations	120,550	121,068	119,807	119,175	119,559	119,804

Variable	15%	15%,99€,salient	15%,99€	15%,199€,salient	15%,199€
Number of sessions	1.839	1.834	1.835	1.836	1.835
	(2.007)	(1.997)	(2.000)	(2.028)	(2.020)
Purchase (Yes=1)	0.017	0.017	0.018	0.016	0.017
	(0.131)	(0.129)	(0.132)	(0.126)	(0.129)
Customer from Germany (Yes=1)	0.916	0.917	0.913	0.916	0.918
	(0.278)	(0.276)	(0.282)	(0.277)	(0.275)
Female (Yes=1)	0.505	0.487	0.499	0.498	0.502
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Age (in years)	52.272	52.587	52.512	51.970	52.093
	(14.417)	(14.453)	(14.191)	(14.565)	(14.563)
B2B sales channel (Yes=1)	0.043	0.043	0.041	0.039	0.042
	(0.203)	(0.202)	(0.197)	(0.193)	(0.200)
Canceled transaction ex-post (Yes=1)	0.017	0.014	0.017	0.017	0.019
	(0.130)	(0.117)	(0.131)	(0.130)	(0.136)
Internet explorer (Yes=1)	0.012	0.013	0.012	0.013	0.013
	(0.111)	(0.115)	(0.110)	(0.114)	(0.111)
Firefox (Yes=1)	0.104	0.107	0.111	0.105	0.106
	(0.305)	(0.309)	(0.314)	(0.307)	(0.308)
Chrome (Yes=1)	0.368	0.363	0.363	0.370	0.369
	(0.482)	(0.481)	(0.481)	(0.483)	(0.483)
Safari (Yes=1)	0.349	0.348	0.350	0.347	0.342
	(0.477)	(0.476)	(0.477)	(0.476)	(0.475)
Observations	121,068	119,798	120,207	120,066	120,653

This table presents the mean of observable variables in different treatment conditions. Standard deviations are reported in parentheses.

Table 6: Differences in Outcomes Across Promotions

	Control	Linear	Nonlinear
Proportion of Subjects Buying	0.0288	0.0313	0.0291
Average Basket Value	2.1746	2.5362	2.3361
Average Basket Value | Purchased	155.2724	157.6103	158.2792
N	65,169	131,527	523,705

Notes: This Table shows key outcomes in Control, Nonlinear, and Linear treatments. The last row shows the sample size for each. In Linear treatments we pool $\delta = 10, 15$, in Nonlinear also $M = 99, 199$.

Table 7: Upper Tail Test at $\theta = 0$, By Discount

Group	$X(0)$	One-sided p-value
10% Off	-0.084	0.023
	(0.042)	
15% Off	0.009	1
	(0.045)	

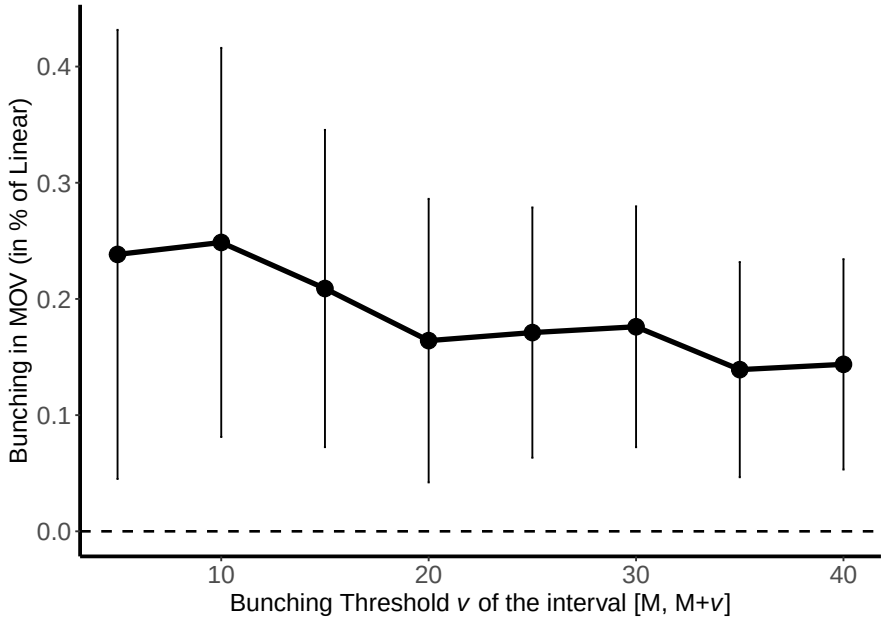
C Bunching Near M

This violation of IIA suggests that the visually detectable "missing mass" to the left of M and bunching in Figure 5 may be partly due to excess mass to the right of M coming from above. To see this we compare Nonlinear and Linear locally, by defining excess mass at $[M, M + \upsilon]$ under Nonlinear compared to Linear as:

$$BU(\upsilon) = \frac{\Pr(b \in [M, M+\upsilon] | t_{M,\delta}) - \Pr(b \in [M, M+\upsilon] | t_{\delta})}{\Pr(b \in [M, M+\upsilon] | t_{\delta})}. \tag{23}$$

Figure 16 reports $BU(\upsilon)$ for υ varying at 5 EUR increments. There is indeed excess mass in right neighborhoods of M, amounting to a 0.1 to 0.2% higher density relative to Linear. In the immediate vicinity of M this excess mass is entirely consistent with bunching from below, and even at $M + 20$ EUR, due to lumpy prices. Notice, though that excess mass extends stays fairly stable until 40 EUR above M, which is a large distance. This seems more indicative of mass coming from above.

Figure 16: Rational Prediction 3: Bunching



Notes: This figure shows bunching at or above M in Nonlinear. Bunching mass (the ordinate) is computed as 23 divided by Linear mass in the same region. The abscissa varies v. The bars represent 95%-CIs.

D Appendices on methods

Bootstrap That is, the right tail $b \geq b'$ is strictly smaller under Nonlinear than Linear (i.e. the mass to its left is larger $B(b|t_{M,\delta}) - B(b|t_\delta) > 0$). Let $\{b_i\}_{i=1}^{100}$ be the percentiles of baskets. We construct our test statistic as follows:

$$\hat{S} = \sum_{M=99,199} \sum_{d=0.1,0.15} \sum_{b_i > M} \left(\max\{0, \hat{B}(b_i|t_{M,\delta}) - \hat{B}(b_i|t_\delta)\} \right)^2. \tag{24}$$

where $\hat{B}$ is the empirical CDF. The value of this statistic in our data is $\hat{S} = 0.000076$. Under the rational null, this positive value should be consistent with mere sampling error.

To do inference, following Barrett and Donald (2003) we draw $K = 2000$ fictitious Nonlinear and Linear samples from the same DGP: the empirical Linear distribution $\hat{B}(b_i|t_\delta)$. Each sample k has one subsample for $t_{M\delta}$, another for t_δ. Using these subsam-

ples we compute our test statistic for k, $\hat{S}_k$.[27]. The p-value for our test across all samples is equal to the share of times the bootstrap statistic is larger than the observed $\hat{S}$.

E Proofs

Proposition 1 and Rational Predictions 1, 2, 3 and 4. A consumer j with continuous, increasing and concave utility maximizes (1) subject to (2) by finding: i) her optimal basket b^j_{M0} and outside good $x^j_{M0} = w_j - b^j_{M0}$ in the control portion of her budget $b + x \leq w_j$ and $b < M$, and ii) her optimal basket $b^j_{M\delta}$ and outside good $x^j_{M\delta} = w_j - b^j_{M\delta}$ under the Linear portion of her budget $(1-\delta) \cdot b + x \leq w_j$ and b. These baskets exist and are unique. She then chooses solves:

$$\max_{b \in \left\{b^j_{M0}, b^j_{M\delta}\right\}} \left\{u_j\left(b^j_{M0}, x^j_{M0}\right), u_j\left(b^j_{M\delta}, x^j_{M\delta}\right)\right\}.$$

Let b^j_C be the optimal basket in t_C and b^j_δ the optimal basket in t_δ. We then create two consumer partitions. The first consists of sets H and L, which respectively identify types who in t_δ prefer to spend not less than M or above it:

$$\begin{aligned} H &= \left\{j | b^j_\delta \geq M\right\}, \\ L &= \left\{j | b^j_\delta < M\right\}, \end{aligned}$$

The second consists of sets Y and N, which identify types who prefer to be in the linear $b \geq M$ or control $b < M$ portion of the $t_{M\delta}$ budget, respectively:

[27] So, with one subsample we construct $\hat{B}(b_i|t_\delta, k)$, with the other $\hat{B}(b_i|t_{M\delta}, k)$. The size of each subsample is equal to the number of baskets in t_δ). Following Barrett and Donald (2003) we test for the original null (which entails a "$\geq$" condition) with a data-generating process obeying the null hypothesis that is easiest to reject, i.e. an "$=$" condition

$$
\begin{aligned}
Y &= \left\{ j | u_j \left(b^j_{M\delta}, x^j_{M\delta} \right) \geq u_j \left(b^j_C, x^j_C \right) \right\}, \\
N &= \left\{ j | u_j \left(b^j_{M\delta}, x^j_{M\delta} \right) < \left(b^j_C, x^j_C \right) \right\}.
\end{aligned}
$$

Note that $H \cap N = \emptyset$. A member of H in fact chooses - in the Linear portion - her optimal Linear basket. Thus, $(b^j_{M\delta}, x^j_{M\delta}) = (b^j_\delta, x^j_\delta) \succ_j (b^j_C, x^j_C) \succ_j (b^j_{M0}, x^j_{M0})$, where the second step in follows because control allocations also belong to the Linear budget. An L member belongs to Y and sets $b^j_{M\delta} = M$ (by convexity of indifference curves she prefers M to any $b > M$ in her Linear budget) or belongs to N and picks her control basket $b^j_{M\delta} = b^j_C$.

We can partition L into "bunchers" B and "non-bunchers" NB as follows:

$$
\begin{aligned}
B &= \left\{ j | j \in L \; and \; u_j \left(M, w_j \left(1 - \delta \right) \cdot M \right) \geq u_j \left(b^j_0, x^j_0 \right) \right\}, \\
NB &= \left\{ j | j \in L \; and \; u_j \left(M, w_j - \left(1 - \delta \right) \cdot M \right) < u_j \left(b^j_0, x^j_0 \right) \right\}.
\end{aligned}
$$

Let $\underline{b}_{M,\delta} = \max_{j \in NB} b^j_0$. The cdf of in $t_{M\delta}$ is Pr$(b^j_C \leq b \; and \; j \in NB)$ for $b \leq \underline{b}_{M,\delta}$, then Pr$(b^j_C \leq \underline{b_{M,\delta}} \; and \; j \in ND)$ for $b \in (\underline{b}_{M,\delta}, M)$, then there is a jump to Pr$(b \leq M | t_{M\delta})$ = Pr$(b^j_C \leq \underline{b}_{M,\delta} \; and \; j \in ND) + Pr(j \in B)$ + Pr$(b_\delta = M$ *and* $j \in H)$, which (given that the sum of the first two terms is equal to Pr$(b_\delta < M)$) is equal to Pr$(b_\delta \leq M)$, and finally the cdf is Pr$(b_\delta \leq b) = B(b|t_\delta)$ for $b > M$.

Consider Rational prediction 1. The above distribution immediately implies that for any $\underline{b}_{M,\delta} < b < M$, we have that $B(b|t_{M\delta}) - B(\underline{b}_{M,\delta}|t_{M\delta}) = 0$.

Consider Prediction 2. The probability mass of the baskets at $b \geq M$ obeys Pr$(b \geq M | t_{M\delta})$ = Pr$(b = M|t_{M\delta})$ + Pr$(b > M|t_\delta)$ = Pr(B) + Pr$(b = M|t_\delta)$ + Pr$(b > M|t_\delta)$ = Pr(L) − Pr(NB) + Pr$(b \geq M|t_\delta)$, which is of course larger than Pr$(b \geq M|t_\delta)$.

Consider Rational Prediction 3. By the previous analysis, for $b > M$ the cdf under $t_{M,\delta}$ is $B(b|t_\delta)$. The result then immediately follows.

Finally, we derive the spending distribution under quasilinear utility and single crossing. As a first step, note that $j' > j \Rightarrow b_C^{j'} > b_C^j$. If, by contradiction, $u_j(b_C^j) > u_j(b_C^{j'})$ and $u_{j'}(b_C^j) < u_{j'}(b_C^{j'})$, which violates single crossing given that $j' > j$. It is also the case that for all types j and j' for which we have that $b_C^{j'} > b_C^j$, then we must have $j' > j$. If, in fact, $j' < j$, this pair of types would again violate single crossing because the higher type j likes the smaller basket b_C^j better than the larger $b_C^{j'}$ while the lower type $j\prime$ does the opposite. Thus, $j' > j \Leftrightarrow b_C^{j'} > b_C^j$.

The previous argument implies that $j \leq j_{NB} \equiv \arg\max_{j \in NB} b_C^j$ for all $j \in NB$. Suppose now that there is a $j^* \in B$ featuring $j^* < j_{NB}$. This would mean, by single crossing, that $b_C^{j^*} < b_C^j$. Yet, by membership in NB and B we know that:

$$\begin{aligned} u_{j_{NB}}(b_C^{j_{NB}}) + w_{j_{NB}} - b_C^{j_{NB}} &\geq u_{j_{NB}}(M) + w_{j_{NB}} - (1-\delta)\, M, \\ u_{j^*}(b_C^{j^*}) + w_{j^*} - b_C^{j^*} &< u_{j^*}(M) + w_{j^*} - (1-\delta)\, M, \end{aligned}$$

which can be rewritten as:

$$\begin{aligned} u_{j_{NB}}(M) - u_{j_{NB}}(b_C^{j_{NB}}) &\leq (1-\delta)\, M - b_C^{j_{NB}}, \\ u_{j^*}(M) - u_{j^*}(b_C^{j^*}) &> (1-\delta)\, M - b_C^{j^*}, \end{aligned}$$

where the inequalities can be rewritten as:

$$\begin{aligned} u_{j_{NB}}(M) - u_{j_{NB}}(b_C^{j_{NB}}) &\leq (1-\delta)\, M - b_C^{j_{NB}}, \\ u_{j^*}(M) - u_{j^*}(b_C^{j_{NB}}) &> (1-\delta)\, M - b_C^{j_{NB}} + [u_{j^*}(b_C^{j^*}) - b_C^{j^*} - u_{j^*}(b_C^{j_{NB}}) + b_C^{j_{NB}}]. \end{aligned}$$

because by assumption $j^* < j_{NB}$, and by definition of the optimum under t_C we have that the term in square brackets in positive, the two inequalities imply:

$$u_{j^*}(M) - u_{j^*}(b_C^{j_{NB}}) > u_{j_{NB}}(M) - u_{j_{NB}}(b_C^{j_{NB}}),$$

which violates single crossing. This implies that all types $j \in NB$ are lower than all types $j \in B$. As a result, NB contains all types whose basket under control is below $\underline{b}$. As a result, $B\left(b|t_{M\delta}\right) = B\left(b|t_C\right)$ for all $b \leq \underline{b}$. □

Proposition 2. Partition baskets in three regions: i) lower tail $\mathcal{L} \equiv [0, M \cdot (1-\delta)]$, ii) dominated $\mathcal{D} \equiv (M \cdot (1-\delta), M)$, and iii) upper tail $\mathcal{U} \equiv [M, \infty)$. Let $\omega_{\mathcal{X}}^C$, $\omega_{\mathcal{X}}^{M\delta}$ and $\omega_{\mathcal{X}}^{\delta}$ be the mass located in $\mathcal{X} = \mathcal{L}, \mathcal{D}, \mathcal{U}$ in Control, Nonlinear, and Linear, respectively. Let $\nu_{\mathcal{X}\mathcal{Y}}$ be the mass of consumers in $\mathcal{X}$ under t_C and $\mathcal{Y}$ under $t_{M\delta}$. Let $\lambda_{\mathcal{X}\mathcal{Y}}$ be the respective mass in t_δ. The joint distribution of baskets in the three regions is:

$C \setminus M\delta$	$\mathcal{L}$	$\mathcal{D}$	$\mathcal{U}$	
$\mathcal{L}$	$\nu_{\mathcal{L}\mathcal{L}}$	$\nu_{\mathcal{L}\mathcal{D}}$	$\nu_{\mathcal{L}\mathcal{U}}$	$\omega_{\mathcal{L}}^C$
$\mathcal{D}$	$\nu_{\mathcal{D}\mathcal{L}}$	$\nu_{\mathcal{D}\mathcal{D}}$	$\nu_{\mathcal{D}\mathcal{U}}$	$\omega_{\mathcal{D}}^C$
$\mathcal{U}$	$\nu_{\mathcal{U}\mathcal{L}}$	$\nu_{\mathcal{U}\mathcal{D}}$	$\nu_{\mathcal{U}\mathcal{U}}$	$\omega_{\mathcal{U}}^C$
	$\omega_{\mathcal{L}}^{M\delta}$	$\omega_{\mathcal{D}}^{M\delta}$	$\omega_{\mathcal{U}}^{M\delta}$	

,

$C \setminus \delta$	$\mathcal{L}$	$\mathcal{D}$	$\mathcal{U}$	
$\mathcal{L}$	$\lambda_{\mathcal{L}\mathcal{L}}$	$\lambda_{\mathcal{L}\mathcal{D}}$	$\lambda_{\mathcal{L}\mathcal{U}}$	$\omega_{\mathcal{L}}^C$
$\mathcal{D}$	$\lambda_{\mathcal{D}\mathcal{L}}$	$\lambda_{\mathcal{D}\mathcal{D}}$	$\lambda_{\mathcal{D}\mathcal{U}}$	$\omega_{\mathcal{D}}^C$
$\mathcal{U}$	$\lambda_{\mathcal{U}\mathcal{L}}$	$\lambda_{\mathcal{U}\mathcal{D}}$	$\lambda_{\mathcal{U}\mathcal{U}}$	$\omega_{\mathcal{U}}^C$
	$\omega_{\mathcal{L}}^{\delta}$	$\omega_{\mathcal{D}}^{\delta}$	$\omega_{\mathcal{U}}^{\delta}$	

. (25)

With no frictions, no consumer moves to or stays into $\mathcal{D}$ under $t_{M,\delta}$, $\nu_{\mathcal{X}\mathcal{D}} = \omega_{\mathcal{D}}^{M\delta} = 0$ for all $\mathcal{X}$, and nobody moves from higher to lower regions under either promotion, $\nu_{\mathcal{X}\mathcal{Y}} = \lambda_{\mathcal{X}\mathcal{Y}} = 0$ for $\mathcal{X} > \mathcal{Y}$ (because any consumer has a weakly positive incentive to increase her basket).

The cost $A_j(b, b_C^j) + \epsilon$ of increasing the basket to $b > b_C^j$ creates rigidity around b_C^j

for each j, dampening her reaction to the promotion, but does not prompt a reduction of the basket. Thus, we still have the restriction $\nu_{\mathcal{XY}} = \lambda_{\mathcal{XY}} = 0$ for $\mathcal{X} > \mathcal{Y}$, which gives:

$C \setminus M\delta$	$\mathcal{L}$	$\mathcal{D}$	$\mathcal{U}$	
$\mathcal{L}$	$\nu_{\mathcal{LL}}$	0	$\nu_{\mathcal{LU}}$	$\omega^C_{\mathcal{L}}$
$\mathcal{D}$	0	$\nu_{\mathcal{DD}}$	$\nu_{\mathcal{DU}}$	$\omega^C_{\mathcal{D}}$
$\mathcal{U}$	0	0	$\nu_{\mathcal{UU}}$	$\omega^C_{\mathcal{U}}$
	$\omega^{M\delta}_{\mathcal{L}}$	$\omega^{M\delta}_{\mathcal{D}}$	$\omega^{M\delta}_{\mathcal{U}}$	

,

$C \setminus \delta$	$\mathcal{L}$	$\mathcal{D}$	$\mathcal{U}$	
$\mathcal{L}$	$\lambda_{\mathcal{LL}}$	$\lambda_{\mathcal{LD}}$	$\lambda_{\mathcal{LU}}$	$\omega^C_{\mathcal{L}}$
$\mathcal{D}$	0	$\lambda_{\mathcal{DD}}$	$\lambda_{\mathcal{DU}}$	$\omega^C_{\mathcal{D}}$
$\mathcal{U}$	0	0	$\lambda_{\mathcal{UU}}$	$\omega^C_{\mathcal{U}}$
	$\omega^{\delta}_{\mathcal{L}}$	$\omega^{\delta}_{\mathcal{D}}$	$\omega^{\delta}_{\mathcal{U}}$	

, (26)

where $\nu_{\mathcal{LD}} = 0$ because a consumer choosing a basket in $\mathcal{L}$ under t_C will never move to $\mathcal{D}$ under $t_{M\delta}$, because $b^j_C \in \mathcal{L}$ is revealed preferred to any $b \in \mathcal{D}$ under the control budget and hence also under $t_{M\delta}$.

To obtain our test, we derive the mass of transition away from $\mathcal{D}$ in t_δ, i.e. $\lambda_{\mathcal{DU}}$, and in $t_{M\delta}$, i.e. $v_{\mathcal{DU}}$, under adjustment costs. A consumer j with $b^j_C \in \mathcal{D}$ chooses to $b^j_{\delta,A} \in \mathcal{U}$ under t_δ if and only if she prefers $b^j_{\delta,A}$ to any $b^j_C < b < M$:

$$u_j(b^j_{\delta,A}) - u(b) + (b - b^j_{\delta,A})(1-\delta) > A_j(b^j_{\delta,A}, b^j_C) - A_j(b, b^j_C), \quad b^j_C < b < M. \tag{27}$$

Let $J_\delta(\mathcal{DU})$ the possibly empty set of types fulfilling $b^j_C \in \mathcal{D}$ and (27). Let $V_j(b^j_{\delta,A}, b^j_C)$ be difference between $j \in J_\delta(\mathcal{DU})$ welfare under $b^j_{\delta,A}$ and b^j_C:

$$V_j(b^j_{\delta,A}, b^j_C) \equiv u_j(b^j_{\delta,A}) - u(b^j_C) + (b^j_C - b^j_{\delta,A})(1-\delta) - A_j(b^j_{\delta,A}, b^j_C). \tag{28}$$

Equation (28) sets a threshold $\epsilon^j_\delta = V_j(b^j_{\delta,A}, b^j_C)$ so that a consumer $j \in J_\delta(\mathcal{DU})$ moves to $b^j_{\delta,A}$ with probability $\lambda^j_{\mathcal{DU}} = \Pr(\epsilon \leq \epsilon^j_\delta)$, and $\lambda_{\mathcal{DU}} = \sum_{j \in J_\delta(\mathcal{DU})} \pi_j \cdot \lambda^j_{\mathcal{DU}}$.

A consumer j with $b^j_C \in \mathcal{D}$ chooses to $b^j_{M\delta,A} \in \mathcal{U}$ under $t_{M\delta}$ if and only if she prefers $b^j_{M\delta,A}$ to any $b^j_C < b < M$:

$$u_j(b^j_{M\delta,A}) - u(b) + b - b^j_{M\delta,A}(1-\delta) > A_j(b^j_{M\delta,A}, b^j_C) - A_j(b, b^j_C), \;\; b^j_C < b < M. \quad (29)$$

Let $J_{M\delta}(\mathcal{DU})$ the possibly empty set of types fulfilling $b^j_C \in \mathcal{D}$ and (29). Let $V_j(b^j_{M\delta,A}, b^j_C)$ be difference between $j \in J_{M\delta}(\mathcal{DU})$ welfare under $b^j_{M\delta,A}$ and b^j_C:

$$V_j(b^j_{M\delta,A}, b^j_C) \equiv u_j(b^j_{M\delta,A}) - u(b^j_C) + b^j_C - b^j_{M\delta,A}(1-\delta) - A_j(b^j_{M\delta,A}, b^j_C). \quad (30)$$

Equation (30) sets a threshold $\epsilon^j_{M\delta} = V_j(b^j_{M\delta,A}, b^j_C)$ so that a consumer $j \in J_{M\delta}(\mathcal{DU})$ moves to $b^j_{M\delta,A}$ with probability $\nu^j_{\mathcal{DU}} = \Pr(\epsilon \leq \epsilon^j_{M\delta})$, and $\nu_{\mathcal{DU}} = \sum_{j \in J_{M\delta}(\mathcal{DU})} \pi_j \cdot \nu^j_{\mathcal{DU}}$.

The mass in $\mathcal{D}$ in $t_{M\delta}$ is positive but smaller compared to t_C, $0 < \omega^{M\delta}_{\mathcal{D}} < \omega^C_{\mathcal{D}}$. With adjustment costs, this reveals that $J_{M\delta}(\mathcal{DU}) \neq \emptyset$. It is also true that $J_{M\delta}(\mathcal{DU}) \supseteq J_\delta(\mathcal{DU})$. Indeed, $j \in J_\delta(\mathcal{DU}) \Rightarrow j \in J_{M\delta}(\mathcal{DU})$ because (29) can be written as:

$$u_j(b^j_{M\delta,A}) - u(b) + (b - b^j_{M\delta,A})(1-\delta) + \delta b > A_j(b^j_{M\delta,A}, b^j_C) - A_j(b, b^j_C), \; b^j_C < b < M, \quad (31)$$

which is easier to fulfill for any $b^j_{M\delta,A} \in \mathcal{U}$ than (27). Furthermore, for any type $j \in J_\delta(\mathcal{DU})$, the probability that she moves to $\mathcal{U}$ is larger in $t_{M\delta}$ than in t_δ. To see this, note as a first step that such type chooses the same basket under $t_{M\delta}$ and t_δ, $b^j_{M\delta,A} = b^j_{\delta,A}$, by an argument equivalent to the one we used in the proof of Proposition 1. As a second step, note that for this type we can write:

$$V_j(b^j_{M\delta,A}, b^j_C) = V_j(b^j_{\delta,A}, b^j_C) + \delta \cdot b^j_C > V_j(b^j_{\delta,A}, b^j_C), \quad (32)$$

which implies $\epsilon^j_{M\delta} > \epsilon^j_\delta$ and hence $\nu^j_{\mathcal{DU}} > \lambda^j_{\mathcal{DU}}$. Overall, then the evidence of shrinkage of $\mathcal{D}$ when going from t_C to $t_{M\delta}$ implies, under adjustment costs, that $\nu_{\mathcal{DU}} > \lambda_{\mathcal{DU}}$.

Let $J_i(\mathcal{LU})$ be the set of types having $b^j_C \in \mathcal{L}$ and $b^j_{i,A} \in \mathcal{U}$ under t_i, $i = M\delta, \delta$. The same logic of (31) for $0 \leq b \leq M(1-\delta)$ implies that $j \in J_\delta(\mathcal{LU}) \Rightarrow j \in J_{M\delta}(\mathcal{LU})$. The same logic of (32) implies that a type $j \in J_\delta(\mathcal{LU})$ is strictly more likey to move to $\mathcal{U}$ under $t_{M\delta}$ than under t_δ. However, we have no evidence to rule out $J_{M\delta}(\mathcal{LU}) = \emptyset$. As a result, for these moves we can only say $\nu_{\mathcal{LU}} \geq \lambda_{\mathcal{LU}}$.

The final step of the proof uses $\nu_{\mathcal{DU}} > \lambda_{\mathcal{DU}}$ and $\nu_{\mathcal{LU}} \geq \lambda_{\mathcal{LU}}$ to derive an implication of the adjustment cost model. From (26) we get:

$$\begin{aligned}
\nu_{\mathcal{DU}} &= \omega^C_{\mathcal{D}} - \omega^{M\delta}_{\mathcal{D}}, \\
\lambda_{\mathcal{DU}} &= \omega^C_{\mathcal{D}} - \omega^{\delta}_{\mathcal{D}} + \lambda_{\mathcal{LD}}, \\
\lambda_{\mathcal{LD}} &= \omega^C_{\mathcal{L}} - \omega^{\delta}_{\mathcal{L}} - \lambda_{\mathcal{LU}},
\end{aligned}$$

which implies:

$$\lambda_{\mathcal{DU}} = \omega^C_{\mathcal{D}} - \omega^{\delta}_{\mathcal{D}} + \omega^C_{\mathcal{L}} - \omega^{\delta}_{\mathcal{L}} - \lambda_{\mathcal{LU}},$$

The rational model with frictions implies:

$$\nu_{\mathcal{DU}} > \lambda_{\mathcal{DU}} \Leftrightarrow \omega^C_{\mathcal{D}} - \omega^{M\delta}_{\mathcal{D}} > \omega^C_{\mathcal{D}} - \omega^{\delta}_{\mathcal{D}} + \omega^C_{\mathcal{L}} - \omega^{\delta}_{\mathcal{L}} - \lambda_{\mathcal{LU}},$$

which is equivalent to:

$$\omega^{\delta}_{\mathcal{D}} + \omega^{\delta}_{\mathcal{L}} > \omega^{M\delta}_{\mathcal{D}} + \omega^C_{\mathcal{L}} - \lambda_{\mathcal{LU}}. \tag{33}$$

Due to the other rational restriction $\nu_{\mathcal{LU}} \geq \lambda_{\mathcal{LU}}$, Equation (33) implies:

$$\omega_{\mathcal{D},\delta} + \omega_{\mathcal{L},\delta} > \omega_{\mathcal{D},M\delta} + \omega_{\mathcal{L},C} - \nu_{\mathcal{LU}}. \tag{34}$$

But then, from Table (26) we know that $\nu_{\mathcal{LU}} = \omega_{\mathcal{L},C} - \omega_{\mathcal{L},M\delta}$, which implies:

$$\omega_{\mathcal{D},\delta} + \omega_{\mathcal{L},\delta} > \omega_{\mathcal{D},M\delta} + \omega_{\mathcal{L},M\delta} \Leftrightarrow \tag{35}$$

$$\omega_{\mathcal{U}}^{\delta} < \omega_{\mathcal{U}}^{M\delta}. \tag{36}$$

Equation (36) states that the tail $b \geq M$ is larger under $t_{M\delta}$ than under t_{δ}. Rejecting this condition rejects rational consumers with frictions.

A richer model allows for noise in the, so the optimum $b_{i,A}^{j}$ is perturbed into $b_{i,A}^{j}+$, where ϕ is iid. In this model all cells are positive and all types reach any region (the support of ϕ is large enough). Now the joint distribution is as in (25). By the arguments in (31) and (32), though, we still have that $\nu_{\mathcal{LU}} \geq \lambda_{\mathcal{LU}}$, $\nu_{\mathcal{UU}} \geq \lambda_{\mathcal{UU}}$ and $\nu_{\mathcal{DU}} > \lambda_{\mathcal{DU}}$, which again implies $\omega_{\mathcal{U}}^{M\delta} > \omega_{\mathcal{U}}^{\delta}$. □

Proposition 3. Each type j represents $t_{M\delta}$ as t_C with probability $\mu(\boldsymbol{\alpha}_M)$, as t_{δ} with probability $\mu(\boldsymbol{\alpha}_{\delta})$, as $t_{M,\delta}$ otherwise. Let $j_M \equiv \min_{j \in H} j$. Under $\boldsymbol{\alpha}_M$ the consumer checks out with b_C^j, under $\boldsymbol{\alpha}_{\delta}$ with b_{δ}^j and tries to redeem, under $\boldsymbol{\alpha}_{M\delta}$ with b_C^j if $j \leq j_{ND}$, with M if $j_{ND} < j \leq j_M$, with b_{δ}^j if $j > j_M$. In the latter two cases she tries to redeem.

Under $\boldsymbol{\alpha}_M$ the consumer buys b_C^j. Under $\boldsymbol{\alpha}_{\delta}$ she also does it with b_{δ}^j if $j \geq j_M$. If $j < j_M$ she gets an error message and: i) with probability η_M she switches to $\boldsymbol{\alpha}_M$ and buys b_C^j, ii) with probability $(1 - \eta_M)$ she switches to $\boldsymbol{\alpha}_{M\delta}$. If $j \leq j_{ND}$ she buys b_C^j if $j_{ND} < j < j_M$ she buys $b = M$. Under $\boldsymbol{\alpha}_{M\delta}$ the consumer buys the same.

In $b < \underline{b}_{M,\delta}$ under all representations consumers buy b_C^j, so the CDF consists of $j \leq j_{ND}$ with $b_C^j < j$. Hence, here $B(b|t_{M,\delta} = B(b|t_C)$. In $\underline{b}_{M,\delta} \leq b < M$ there are no people $j > j_{ND}$ using a correct representation (initially or upon receiving a coupon

error), only types $j_{ND} < j < j_M$ using a control representation and having $b_C^j \leq b$. A control representation arises with probability $\widetilde{\mu}_M \equiv \mu(\boldsymbol{\alpha}_M) + \mu(\boldsymbol{\alpha}_\delta) \cdot \eta_M$, so the CDF in this range is $B(b|t_{M,\delta} = B(b|t_C) \cdot \widetilde{\mu}_M + B(\underline{b}_{M,\delta}|t_C)\dot{(}1 - \widetilde{\mu}_M)$. For any $b \geq M$ types with a correct or discount representation have $b_\delta^j \leq b$ and types with a control representation have $b_C^j \leq b$. Thus, $B(b|t_{M\delta}, \boldsymbol{\alpha}) = B(M|t_C) \cdot \mu(\boldsymbol{\alpha}_M) + B(M|t_\delta) \cdot [1 - \mu(\boldsymbol{\alpha}_M)]$., which completes the proof. □

Proof of Proposition 4. We can trace the effect of higher δ through an increase in bottom up discount contrast σ_D. By deriving Equation (11) it is immediate to find:

$$\begin{aligned} \frac{\partial \mu_\beta(\alpha_M)}{\partial \sigma_D} &= -\mu_\beta(\alpha_M) \cdot [1 - \mu_\beta(\alpha_M)], \frac{\partial \mu_\beta(\alpha_\delta)}{\partial \sigma_D} = \mu_\beta(\alpha_\delta) \cdot \mu_\beta(\alpha_M), \\ \frac{\partial \mu_\beta(\alpha_{M,\delta})}{\partial \sigma_D} &= \mu_\beta(\alpha_{M,\delta}) \cdot \mu_p(\alpha_M). \end{aligned}$$

Similarly, we can show that:

$$\begin{aligned} \frac{\partial \mu_\beta(\alpha_M)}{\partial v_M} &= \sigma_M \cdot \mu_\beta(\alpha_M) \cdot \mu_p(\alpha_\delta), \frac{\partial \mu_p(\alpha_\delta)}{\partial \sigma_D} = -\sigma_M \cdot \mu_\beta(\alpha_\delta) \cdot [1 - \mu_\beta(\alpha_\delta)], \\ \frac{\partial \mu_\beta(\alpha_{M,\delta})}{\partial \sigma_D} &= \sigma_M \cdot \mu_\beta(\alpha_{M,\delta}) \cdot \mu_\beta(\alpha_\delta). \end{aligned}$$

□

Consider finally the cdf for $b > M$. Here there are three groups. First, consumers ex ante choosing α_M and choosing $b > M$, which amount to mass $F(b|t_C) \cdot \mu_i(\alpha_M)$. Second, consumers ex-ante choosing $\alpha_{M\delta}$, whose mass below b is $F(b|t_\delta) \cdot \mu_i(\alpha_{M\delta})$. Third, consumers ex-ante choosing α_δ, who now do not receive an error messgae, whose mass below b is $F(b|t_\delta) \cdot \mu_i(\alpha_\delta)$. Thus, the cdf in this range is equal to:

$$F(b|t_C) \cdot \mu_i(\alpha_M) + F(b|t_\delta) \cdot [1 - \mu_i(\alpha_M)].$$

Proof of Model Predictions 1-5. Model prediction 1 follows directly from the equilibrium cdf. For $b < M$:

$$F\left(b|t_{M\delta},\alpha\right) - F\left(M\left(1-\delta\right)|t_{M\delta},\alpha\right) = \left[F\left(b|t_C\right) - F\left(M\left(1-\delta\right)|t_C\right)\right]\cdot\widetilde{\mu}_M,$$

which pins down the mass in the prediction.

Model prediction 2 follows from the fact that the probability of receiving an error message and jumping ambove M is is equal to the share of consumers settled on profile α_δ who spend in the bunching region $\mathcal{B}$ and then adjust to the correct profile $\alpha_{M\delta}$:

$$\Pr\left(b \geq M, error\right) = \mu_i\left(\alpha_\delta\right)\cdot\Pr\left(\mathcal{B}\right)\cdot\left(1-\eta_M\right),$$

while the probability of receiving an error and being in the bunching region is equal to:

$$\Pr\left(b \geq M, error\right) = \mu_i\left(\alpha_\delta\right)\cdot\Pr\left(\mathcal{B}\right)\cdot\eta_M.$$

The ratio of these probabilities directly yields Prediction 2.

To get model prediction 3, consider the difference between the mass to the right of $b > M$ under $t_{M\delta}$ and under t_δ:

$$\begin{aligned}
&\left[1 - F\left(b|t_{M\delta},\alpha\right)\right] - \left[1 - F\left(b|t_\delta\right)\right] \\
&\quad = F\left(b|t_\delta\right) - F\left(b|t_{M\delta},\alpha\right) \\
&\quad = -\left[F\left(b|t_C\right) - F\left(b|t_\delta\right)\right]\cdot\mu_i\left(\alpha_M\right) < 0.
\end{aligned}$$

This expression gets closer to zero as δ increases because $\mu_i\left(\alpha_M\right)$ decreases with δ. The logic is the same for model prediction 4. Model prediction 5 is evident by inspection.

□

Proof of Model Prediction 6. Under $t_{M\delta}$, the total mass of baskets with at least one good $p_k \geq M$ (call this event $\boldsymbol{B}_{SQ}$) and $b \geq M$ is equal to:

$$\begin{aligned}\Pr\left(\boldsymbol{B}_{SQ}|t_{M\delta},\alpha\right) = &\left[\mu_i\left(\alpha_Q\right)+\mu_i\left(\alpha_{SQ}\right)\right]\cdot\Pr\left(\boldsymbol{B}_{SQ}|t_C\right)\\ &+\left[\mu_i\left(\alpha_{M\delta}\right)+\mu_i\left(\alpha_\delta\right)\right]\cdot\Pr\left(\boldsymbol{B}_{SQ}|t_\delta\right)\\ &+\mu_i\left(\alpha_{SQ\delta}\right)\cdot\Pr\left(b\geq M|t_{SQ}\right).\end{aligned}$$

where α_Q and α_{SQ} are the profiles in which only Q or SQ are attended to, $\alpha_{M\delta}$ are profiles in which both Q and P are attended to, α_δ is the profile in which only P is attended to, and $\alpha_{SQ\delta}$ is the profile in which only SQ and P are attended to. We restrict attention to these profiles, eliminating the redundant one in which only Q and SQ are attended to. Here $\Pr\left(b\geq M|t_{SQ}\right)$ denotes the rational distribution for a treatment in which the promotion is based on SQ rather than Q.

The probability of $b \geq M$ is then equal to:

$$\begin{aligned}&\Pr\left(\boldsymbol{B}_{SQ}|t_{M\delta},\alpha\right)+\Pr\left(b\geq M\cap\sim\boldsymbol{B}_{SQ}|t_{M\delta},\alpha\right)\\ &=\left[\mu_i\left(\alpha_Q\right)+\mu_i\left(\alpha_{SQ}\right)\right]\cdot\Pr\left(b\geq M|t_C\right)\\ &\quad+\left[\mu_i\left(\alpha_{M\delta}\right)+\mu_i\left(\alpha_\delta\right)\right]\cdot\Pr\left(b\geq M|t_\delta\right)\\ &\quad+\mu_i\left(\alpha_{SQ\delta}\right)\cdot\Pr\left(b\geq M|t_{SQ}\right).\end{aligned}$$

Thus,

$$\frac{\Pr\left(\boldsymbol{B}_{SQ}|t_{M\delta},\alpha\right)}{\Pr\left(b\geq M|t_{M\delta},\alpha\right)} = \frac{\left[\mu_i\left(\alpha_Q\right)+\mu_i\left(\alpha_{SQ}\right)\right]\cdot\Pr\left(\boldsymbol{B}_{SQ}|t_C\right)+\left[\mu_i\left(\alpha_{M\delta}\right)+\mu_i\left(\alpha_\delta\right)\right]\cdot\Pr\left(\boldsymbol{B}_{SQ}|t_\delta\right)+\mu_i\left(\alpha_{SQ\delta}\right)\cdot\Pr\left(b\geq M|t_{SQ}\right)}{\left[\mu_i\left(\alpha_Q\right)+\mu_i\left(\alpha_{SQ}\right)\right]\cdot\Pr\left(b\geq M|t_C\right)+\left[\mu_i\left(\alpha_{M\delta}\right)+\mu_i\left(\alpha_\delta\right)\right]\cdot\Pr\left(b\geq M|t_\delta\right)+\mu_i\left(\alpha_{SQ\delta}\right)\cdot\Pr\left(b\geq M|t_{SQ}\right)}.$$

□

Under t_δ we have:

$$\frac{\Pr\left(\boldsymbol{B}_{SQ}|t_\delta,\alpha\right)}{\Pr\left(b\geq M|t_\delta,\alpha\right)}=\frac{\Pr\left(\boldsymbol{B}_{SQ}|t_\delta\right)}{\Pr\left(b\geq M|t_\delta\right)}.$$

Thus, the share of baskets with $b > M$ that contain expensive items increases under MOV provided:

$$\begin{aligned}&\left[\Pr\left(b\geq M|t_\delta\right)-\Pr\left(\boldsymbol{B}_{SQ}|t_\delta\right)\right]\mu_i\left(\alpha_{SQ\delta}\right)\cdot\Pr\left(b\geq M|t_{SQ}\right)\\&>\left[\mu_i\left(\alpha_Q\right)+\mu_i\left(\alpha_{SQ}\right)\right]\cdot\Pr\left(b\geq M|t_C\right)\cdot\Pr\left(b\geq M|t_\delta\right)\\&\quad\cdot\left[\frac{\Pr\left(\boldsymbol{B}_{SQ}|t_\delta\right)}{\Pr\left(b\geq M|t_\delta\right)}-\frac{\Pr\left(\boldsymbol{B}_{SQ}|t_C\right)}{\Pr\left(b\geq M|t_C\right)}\right].\end{aligned}$$

Given that $\Pr\left(b\geq M|t_\delta\right)-\Pr\left(\boldsymbol{B}_{SQ}|t_\delta\right)>0$, this is always fulfilled provided baskets with expensive goods are not more prevalent among baskets with $b > M$ under t_δ than under t_C.

F Bootstrap Test For Upper Tail

In this section, we describe our bootstrap test of upper tail violations in more detail. As stated in the body of the paper, the main idea behind the test is that we wish to "count" the number of times that $X(\theta) < 0$ in the upper tail. Specifically, let $b_1, \ldots, b_K$ index a grid of K basket values. The test statistic we construct is defined by

$$\hat{S}=\sum_{k=1}^{K}\sum_{\delta=0.1,0.15}\sum_{M=99,199}\mathbb{1}\{b_j>M\}(\hat{B}(b_j|t_\delta)-\hat{B}(b_j|t_{M\delta}))^2. \tag{37}$$

Intuitively, $\hat{S}$ measures the number of basket values b_j above M that the upper tail mass above b_j in $t_{M\delta}$ is *smaller* than the corresponding upper tail mass in t_δ, placing higher weight on larger larger violations.

We need to compare the value of $\hat{S}$ in our data with its distribution under the null hypothesis. To do this we start by stating the null hypothesis precisely: $H_0 : X_{M\delta}(\theta) \geq 0$ for all M, δ and all $\theta > 0$. The alterantive hypothesis is then that there exists some M, δ, θ such that $X_{M\delta}(\theta) < 0$. The form of the null hypothesis raises a technical challenge: it is not a "sharp" null, so the sampling distribution of $\hat{S}$ is not fixed within the null. Thus, there is a question of *which* null distribution we should compare $\hat{S}$ to.

Following ideas contained in Barrett and Donald (2003), our test is motivated by the observation that the data-generating process which is consistent with our null hypothesis which is *closest* to the alternative hypothesis is one where uniformly, $X_{M\delta}(\theta) = 0$. But this is precisely the data-generating process that would arise if $B(b|t_{M\delta}) = B(b|t_\delta)$ for all b. We can sample directly from this DGP. Let N_δ be the number of observations in t_δ, and define $N_{M\delta}$ analogously. Then for each bootstrap iteration $b = 1, \ldots, 2000$, we implement the following procedure:

1. Within the Linear treatment groups t_δ, re-sample a sample size N_δ of basket values from the data in t_δ with replacement.

2. Within the Nonlinear treatment groups $t_{M\delta}$, re-sample a sample size N_δ of basket values from the data in t_δ with replacement.

3. Using this re-sampled dataset, compute $\hat{S}_b$ using Equation (37)

Asymptotically, the distribution of $\hat{S}_b$ thus approximates the asymptotic distribution of $\hat{S}$. Thus, our bootstrap p-value measures how extreme $\hat{S}$ is relative the bootstraps: $p \equiv \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{\hat{S} > \hat{S}_b\}$.

G Elasticity Estimates

In this appendix, we describe how we compute the various elasticities when comparing our experimental elasticities to what would have been implied by a standard bunching analysis.

We begin by describing the bunching-based estimate of elasticity. For fixed M and δ, our estimation proceeds in two stages. In the first stage, we attempt to identify the location of where the bunching region begins, and in the second stage, we translate this bunching region start point into an implied elasticity. In the first stage, we proceed as follows. For a given guess of the bunching region start point $\underline{b}$, we bin basket values into bins of size 1, and estimate a polynomial regression of bin size on bin location, using only basket values below $\underline{b}$ or above $M+25$. This gives us an estimate of the "counterfactual" density, assuming that $\underline{b}$ is the beginning of the bunching region. We then use this counterfactual density to estimate excess mass just above M (which is taken to be true mass in $[M, M+25)$ minus the regression-implied mass in that region), and missing mass just below M (again, taken to be the true mass in $[\underline{b}, M)$ minus its regression-implied counterpart). We estimate $\underline{b}$ as the value of $\underline{b}$ that sets the excess mass equal to the missing mass. Given this estimate of $\underline{b}$, the implied elasticity is $\varepsilon = -\underline{b}/(e+1) - (\underline{b}^{-1/e}/(1+1/e) \cdot M^{1+1/e} - (1-\delta)M)$.

Next, we describe the quantile effects estimate. This estimate allows elasticity to flexibly vary across the basket value distribution, but uses experimental variation for identification. For fixed M and δ, we take basket values $b \in \{M-10, M-9, \ldots, M+10\}$ and compute a quantile-effects elasiticty of treatment around b by taking $(B^{-1}(B(b|t_C)|t_\delta) - b)/b/\delta$, and taking the average of these estimates across the different values of b.

Finally, if we assume that demand curves are linear with equal slopes at the individual level, we can get a slightly higher-powered estimate of elasticity. Under the linearity

assumption, we can regress basket value on δ. Let the coefficient in front of δ, divided by the total proportion of buyers, be given by $\hat{\beta}_1$. Since price in control was normalized to be 1, this slope approximates the individual-level *semi-elasticity* of demand: $\frac{\partial q}{\partial p}p$. To get an elasticity, we just have to divide by the baseline level. Since we are estimating elasticities around M, our estimate of elasticity is therefore $\hat{\beta}_1/M$, since the baseline price was normalized to $p = 1$.

In Table 1, we average these estiamtes across all values of M and across all values of δ. For the bunching and quantile effects elasticities, standard errors are obtained by bootstrap: we re-draw our entire dataset with replacement and estimate the elasticity on the re-drawn dataset.

H Price Targeting

We now allow consumers to focus on good-specific features that vary with the promotion. One intuitive feature, which we discussed, is whether the basket contains a good whose price is above M. It is an "easy to check" and salient condition for determining whether $b \geq M$ without computing basket value. We cal this feature "Strongly Qualify for Promotion", $F = S$, where $S(\mathbf{x}_S) = 1$ if the basket contains at least one unit of a good k with regular price $p_{rk} \geq M$ and 0 otherwise. We denote this feature by the indicator $\mathbf{1}_S$. Under linear discount or control this feature takes value 1 for all baskets (and cannot draw attention).

Denote by α_S the consumer's attention to this feature. Let $\mathbb{P}(E)$ denote the probability a discount applies or not in event E. We can write the representation of the promotion as the perceived probability the basket qualifies times the perceived discount:

$$\Delta(b;\alpha) = \mathbb{P}(\mathbf{1}_{b\geq M}|\mathbf{1}_S)^{\alpha_{MOV}}\mathbb{P}(\mathbf{1}_S)^{\alpha_S} \cdot \alpha_D \cdot \delta \cdot b, \tag{38}$$

The representation does not depend on S if the consumer attends to basket value, $\alpha_{MOV} = 1$. In this case, Equation (38) yields $\Delta(b;\alpha) = \alpha_D \cdot \delta \cdot b$ as in (**??**). If however $\alpha_{MOV} = 0$, but $\alpha_S = 1$, the perceived discount is $\Delta(b;\alpha) = (\mathbf{1}_S) \cdot \alpha_D \cdot \delta \cdot b$. That is, the consumer perceives a discount if her basket contains a good whose price is above M, not otheriwse. [28]

Possible attention to S creates a new profile in which the consumer neglects b but attends to the regular price of expensive items, $\alpha_S = 1$ and $\alpha_{MOV} = 0$. By also boosting attention to regular prices, attention to S can interfere with a "Linear Pricing" representation $\alpha_D = 1$ and $\alpha_{MOV} = \alpha_S = 0$. In addition, attention to the S feature creates an under-inclusive representation whereby the consumer attends to S and to the discount, $\alpha_S = \alpha_D = 1$, but not to the basket value, $\alpha_{MOV} = 0$. This consumer represents the goal of "obtaining the discount" as "buy an expensive item". Thus, she compares her maximum utility $u_j(\mathbf{x}_S) + x_O$ in the set of feasible discounted baskets with expensive items $\{b|\mathbf{1}_S = 1\}$, with the maximal utility she obtains in the set of feasible non discounted baskets with only cheap items, featuring $p_k < M$.

As a result, $(\alpha_D, \alpha_{MOV}, \alpha_S) = (1, 0, 1)$ entails a violation of Hicksian separability. Consumers seeking the discount tilt their budget toward expensive items $p_k \geq M$, other consumers do not do so. Due to this distortion, Nonlinear Pricing reduces - compared to Linear - the utility from spending and it also has the following testable prediction.

Model Prediction 4 *$t_{M\delta}$ changes the composition of baskets compared to t_δ for $b \geq M$. In particular, if the share of baskets $b \geq M$ that contain at least one good with $p_k \geq M$ is not larger under t_δ than under t_C, then $t_{M\delta}$ unambiguously increases the share of baskets $b \geq M$ that contain at least one good with $p_k \geq M$ comparedto t_δ.*

By focusing attention on regular prices above M, $t_{M\delta}$ changes the composition of baskets having $b > M$ compared to t_δ. As consumers focus on the discount, on S but not

[28] If the consumer neither attends to MOV nor S, the perceived discount is the same as in (**??**).

on MOV, they tilt expenditure toward expensive goods so as to meet M. If they focus on S but neglect δ, they behave as in control, which however does not alter basket composition with respect to t_δ if Hicksian separability (or its weakening stated in Prediction 4) holds between these conditions.

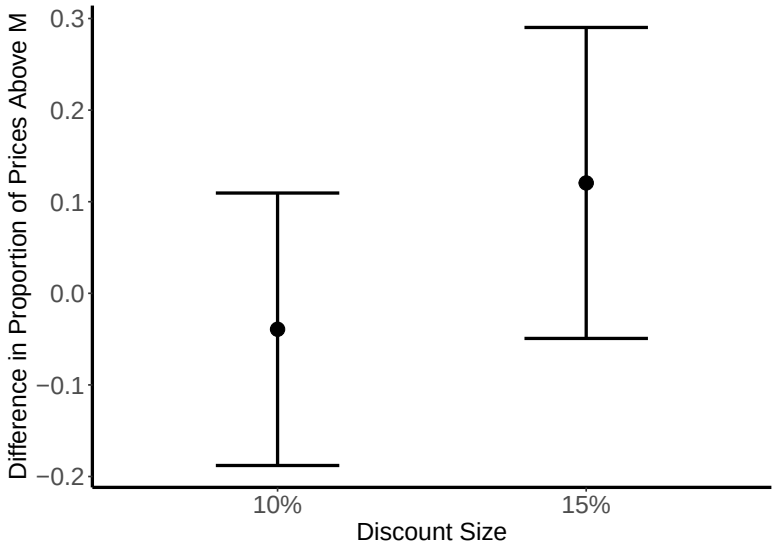
We test this prediction in Figure 17. In Panel (a), we plot the proportion of prices above M conditional on $b \geq M$ in Control and Linear Pricing. We do this separately for $M = 99$ and $M = 199$. It is roughly the case that this quantity is not substantially larger under Linear than Control. Thus, Prediction 4 implies that we should expect that the share of baskets $b \geq M$ with prices $p_k \geq M$ to be higher in Nonlinear than in Linear treatments. We test this prediction in Panel (b). Splitting 10% vs and 15% off, we plot the difference in proportion of prices above M conditional on purchasing above the M in the Nonlinear vs Linear promotions. We find a positive extra mass for the 15% off treatments, but not the 10% off treatments. One might wonder whether our results for the 15% off MOVs are driven by anchoring near M. This is not the case: the increased mass appears to be across the entire upper tail, not just prices near M, which provides evidence against an explanation appealing purely on price anchoring.

I Ex-Ante High Spenders

In our model, missing right tail mass arises also among high type j consumers, who are likely to choose $b >> M$ under linear. Upon focusing on M, in fact, even these consumers behave more thrifty under Nonlinear. As a result, we should also observe missing upper tail mass among such high types. Under rational inattention, by contrast, consumers who are ex-ante more likely to spend above M have a stronger incentive to attend to the discount δ. Hence, missing right tail mass should be smaller among them.

We sort consumers based on a predictor of spending. For each website visitor, we

Figure 17: Difference in Share of Prices above M across Discount Size



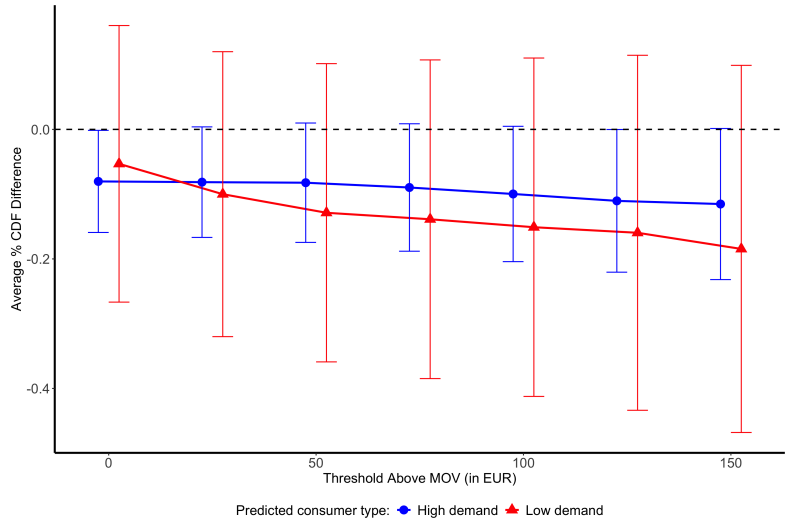
Notes: This figure illustrates price anchoring across treatment conditions. The figure shows the proportion of prices that are above the threshold M for the 10% and 15% discount, respectively. The bars represent 95%-CIs. The bars represent 95%-CIs.

observe the first page they land on. Randomization occurs subsequently, so the landing page is orthogonal to treatment assignment. Using data from Control, we calculate the share of consumers who spend above M conditional on the page they land on. We classify the top half of pages by this metric as "high spender" pages and the bottom half as "low spender" pages. We then classify consumers based on whether they started browsing in a high/low spender page.

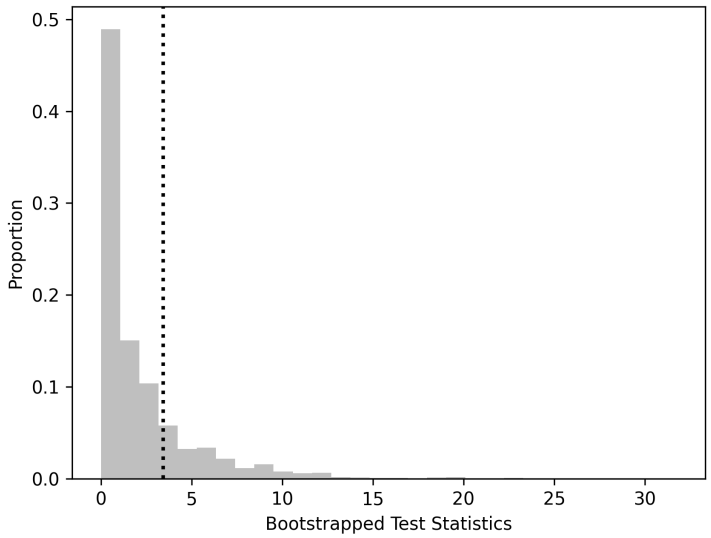
Figure 18 reports, in Panel (a), excess mass for these consumer groups in the 10% Nonlinear treatments, where missing mass is present. Missing right tail mass is statistically significant for high spenders, while it is not for other consumers. This is consistent with our mechanism: the MOV feature in Nonlinear focuses even some high willingness to pay consumers on the pain of spending, reducing their sensitivity to δ compared to similar consumers under Linear.

Figure 18: Right Tail Test, by Consumer Type

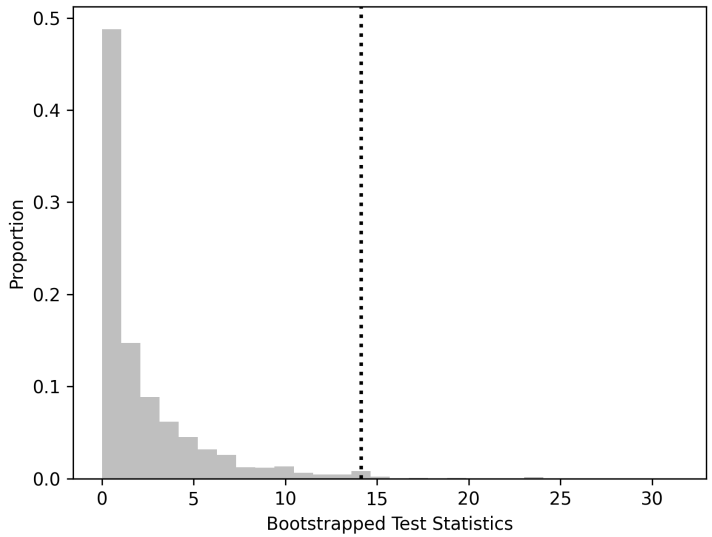
(a) Missing Mass



(b) Low Bootstrap



(c) High Bootstrap



Notes: This figure compares missing mass in the upper tail for people predicted to spend more than M, as described in more detail in the text. Panel (a) plots the excess mass for the two groups at different thresholds, with vertical bars reporting 95% CIs. Panels (b) and (c) report the bootstrap tests for the

J Estimation of Elasticity And Basket Value Thresholds

In this appendix, we provide details about how to obtain $\underline{b}_{M\delta}$ from the price elasticity, ξ. Recall that $\underline{b}_{M\delta}$ is defined by an indifference condition: the basket value in control consumed by a consumer who is just indifferent between purchasing M to obtain the discount and just purchasing the control quantity. Mathematically, this indifference states that $u(M) - M(1-\delta) = u(\underline{b}_{M\delta}) - \underline{b}_{M\delta}$. Rearranging, we have $u(M) - u(\underline{b}_{M\delta}) = M - \underline{b}_{M\delta} - M\delta$. Taking a second order Taylor expansion of the LHS, this equality is approximated by $u'(\underline{b}_{M\delta})(M - \underline{b}_{M\delta}) + \frac{1}{2}u''(\underline{b}_{M\delta})(M - \underline{b}_{M\delta})^2$. But note that by the consumer's FOC in control, $u'(\underline{b}_{M\delta}) = 1$, while differenting this FOC, we have that $\frac{\partial b}{\partial p} = 1/u''(\underline{b}_{M\delta})$. This implies that the elasticity, $\xi \equiv \frac{\partial b}{\partial p}\Big/\underline{b}_{M\delta} = 1/(\underline{b}_{M\delta}u''(\underline{b}_{M\delta}))$, so $u''(\underline{b}_{M\delta}) = 1/(\underline{b}_{M\delta}\xi)$. Substituting the above expressions for u' and u'' in the indifference condition and rearranging then gives $(M - \underline{b}_{M\delta})^2 = -2\xi M\delta\underline{b}_{M\delta}$, which is a quadratic equation in $\underline{b}_{M\delta}$. The smaller of the two roots for this equation is precisely the desired indifference point $\underline{b}_{M\delta}$.[29] We report these indifference points in Table 8 below.

Table 8: Basket Value Thresholds

	$b^*_{M\delta}$	$\underline{b}_{M\delta}$
10% Off, 99 MOV	89.11	63.55
15% Off, 99 MOV	84.16	67.64
10% Off, 199 MOV	179.12	127.73
15% Off, 199 MOV	169.18	115.87

K Estimation of Preference Types

In this Appendix, we discuss how we identify the utility function in the structural estimation. To gain some intuition for how u is identified, fix some quantile $r \in (0,1)$,

[29] Note that the larger of the two roots of the equation exceeds M and corresponds to an indifference condition for a *maximum* order value, where one gets the discount only if consumption is below M.

and let b_C^r be the basket-value chosen by the r^{th} quantile individual in control. A given guess of the utility function u implies, via the consumer's first-order condition, that this individual must have type $\theta_r = 1 - u'(q_r^C)$. Moreover, the model also implies that the r^{th} quantile consumer at *all* treatments all have the same type. Thus, holding the same guess of the utility function, the consumption choice of the r^{th} quantile in the linear treatment with discount d must be $(u')^{-1}((1-d) - \theta_r)$.[30]. While this argument in principle shows why u is nonparametrically identified, in practice, we find that assuming linear demand at the individual level fits the data well, and is considerably more transparent to estimate than more complicated specifications. Linear demand implies utility to be quadratic, i.e. $U(b, \theta) = -\frac{ab^2}{2} + \theta b$. The parameter characterizing the shape of the utility function is thus a. Under this quadratic specification, the derivative of aggregate demand with respect to price is $D' = \frac{1}{a}\Pr[b > 0]$. Thus, we estimate a according to $\hat{a} = 1/(\hat{\beta}\widehat{\Pr}[b > 0])$, where $\hat{\beta}$ is the regression coefficient when regressing basket value on $(1-d)$ using data from the linear pricing treatments. Here, by convention, $d = 0$ in control. This yields a highly statistically significant point estimate of $\hat{a} = 0.0078$ with standard error 0.0020.

With an estimated utility function, estimation of F_θ is also straightforward and can be done nonparametrically. As argued earlier, once the utility function is known, θ can be inferred from an individual's consumption choice via the relationship $\theta = (1-d) - u'(b)$, where d is the discount given in the treatment group of the individual. We can thus directly estimate the θ distribution via the empirical CDF of these inferred values. When computing counterfactuals, it is convenient to represent this CDF as a B-spline as well. We found that a B-spline with 6 knots was sufficiently rich to approximate the distribution of θ well.

[30]Here again, we used the consumer's FOC

L Model Fit Heterogeneity

In Table 9, Heterogeneity by discount.

Table 9: Comparative Statics for Discount Size

	Empirical Value	Predicted Value
Bunching in MOV (% of Linear), 10% Off	0.086	0.138
	(0.062)	
Bunching in MOV (% of Linear), 15% Off	0.194	0.193
	(0.066)	
Incidence in Dominated Region (% Of Control), 10% Off	0.739	0.660
	(0.060)	
Incidence in Dominated Region (% Of Control), 15% Off	0.901	0.623
	(0.061)	
Upper Tail Missing Mass Ratio, 10% Off	0.819	0.595
	(0.326)	
Upper Tail Missing Mass Ratio, 15%	0.061	0.549
	(0.392)	
Upper Tail Missing Mass Ratio, 10% Off,$\theta = 0$	0.490	0.595
	(0.198)	
Upper Tail Missing Mass Ratio, 15%,$\theta = 0$	-0.055	0.549
	(0.228)	

Notes: This table reports the predicted and empirical statistics of our three key behavioral responses, separetly for the 10% and 15% discount. Column 1 reports the empiricially observed value of each moment, while column 2 shows the predicted value from the model.