Customer Prototypicality and the Effectiveness of Segment-Level Personalization: Evidence from a Field Experiment

Raghuram Iyengar
Miers-Busch W'1885 Professor and Professor of Marketing, Department of Marketing, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, riyengar@wharton.upenn.edu

Young-Hoon Park
Sung-Whan Suh Professor of Management and Professor of Marketing, Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853, yp34@cornell.edu

Ankit Sisodia
Assistant Professor of Marketing, Daniels School of Business, Purdue University, West Lafayette, IN 47907, asisodia@purdue.edu

January 2026

Customer segmentation is central to personalization, yet managerially important heterogeneity in responses remains within segments. With firms increasingly using customer embeddings as bases for segmentation, these representations offer a natural way to account for within-segment heterogeneity. Existing work has used embeddings to create micro-segments or to construct ad hoc similarity metrics, but such approaches lack a clear theoretical grounding. We propose a theory-driven, embedding-based measure of customer representativeness (prototypicality) and test whether it reliably accounts for heterogeneous responses to segment-level personalization. In collaboration with a retailer that uses embedding-based value segments, we conduct a large-scale randomized field experiment that orthogonally varies the message framing (symbolic "for-you" cues vs. generic wording) and recommendation method (algorithmic vs. popularity-based). Prototypicality systematically moderates treatment effects: among high-value customers, symbolic cues paired with algorithmic recommendations increase response for highly prototypical members but decrease engagement for peripheral members. A similar moderation pattern emerges under popularity-based recommendations, suggesting that the effects are unlikely to be driven by algorithm-specific idiosyncrasies. Finally, demographics and global similarity measures do not consistently reproduce these patterns, underscoring that within-segment representativeness is an important source of heterogeneity. Interpreted through a reference-dependent lens, our findings align with the idea that salient personalization creates expectations, yielding recognition gains when realized fit meets those expectations for prototypical customers, but misrecognition losses when it falls short for peripheral customers.

Key words: Customer Prototypicality; Embedding-based Segmentation; Segment-level Personalization; Heterogeneous Treatment Effects; Field Experiment

---

1. Introduction

Personalization is central to marketing theory and practice. Prior research has examined its many dimensions, including the value of customer information (e.g., Rossi et al. 1996), the implications of targeting rules for marketing returns (e.g., Manchanda et al. 2004), and the interplay between machine learning and targeting strategies (e.g., Ascarza 2018, Yoganarasimhan et al. 2023). Recent advances in algorithms have generated growing excitement about the potential for real-time, individual-level targeting (e.g., Jin et al. 2023, Challapalli et al. 2025). Yet one-to-one personalization at scale remains feasible primarily for firms with exceptionally rich data and substantial computational resources (e.g., Netflix, Amazon). Most firms instead rely on segment-level personalization, tailoring messages and recommendations to broad customer categories, which strikes a practical balance between mass marketing and one-to-one targeting (e.g., Wedel and Kamakura 2000). Understanding when and why segment-level personalization may enhance or diminish customer response is therefore an important question.

At the same time, the way segments are constructed has evolved, driven in large part by advances in representation learning (e.g., Jagabathula et al. 2018, Chen et al. 2024). Rather than relying on a small set of hand-picked variables (e.g., demographics), many firms now use customer embeddings, latent representations learned from transaction, browsing, or product interaction data, as inputs into segmentation and recommendation systems. These approaches summarize complex behavioral patterns as vectors in a latent space and, in practice, clustering applied to these vectors yields value- or behavior-based segments that underpin segment-level targeting. The continuous embedding space additionally creates a natural opportunity to study within-segment heterogeneity in personalization effectiveness. Extant work has leveraged embeddings as covariates in response models, to create micro-segments, or to construct ad hoc similarity metrics, but these approaches lack a theoretically grounded interpretation of why customers within the same segment should respond differently to personalization (e.g., Wedel and Kamakura 2000).

In this paper, we propose a theory-based, embedding-derived measure of within-segment representativeness (prototypicality) and test whether it systematically accounts for the heterogeneity in responses to segment-level personalization.

Our approach draws on prototype theory in categorization, which posits that categories are organized around central exemplars (prototypes) and that members vary in typicality according to their similarity to these prototypes (Rosch 1975). Extant research shows that prototypical exemplars are more responsive to category-consistent communications than peripheral ones (e.g., Mervis et al. 1981, Loken and Ward 1990, Bhattacharya and

Sen 2003). Other work formalizes prototype theory in multi-dimensional feature spaces, where category membership is characterized by conceptual distance to a prototype (e.g., Gärdenfors 2004). We apply this perspective to embedding-based customer segmentation. In this case, prototype theory would imply that customer typicality should be defined relative to their assigned segment, rather than based on the overall population. We, therefore, operationalize a customer's prototypicality based on their (normalized) distance from the centroid of their assigned segment in the embedding space. This mapping provides a parsimonious, psychologically grounded way to analyze the heterogeneity in response to segment-level personalization.

Firms typically vary their segment-level digital communications along two dimensions: message-level personalization, which makes personalization salient through symbolic cues (e.g., "Recommended for you"), and content-level personalization, which determines the recommendations shown using algorithms or heuristics inferred from past behavior (e.g., Sahni et al. 2018, Kallus and Udell 2020). We posit that, in such settings, customer responses reflect the interplay of three underlying forces at work: baseline match (the alignment between preferences and recommendations), recognition (the positive response when content is perceived as intentionally tailored), and misrecognition (the negative response when salient personalization fails to deliver on the expectations it creates). In this view, symbolic cues plausibly elevate the expectation of high fit and intentionality. When the realized content quality meets this benchmark, symbolic cues amplify engagement by converting functional accuracy into perceived personalization; when realized fit falls short, the same cues can backfire by highlighting mismatch and undermining credibility. Embedding-based prototypicality sharpens this tradeoff within a segment: prototypical customers are more likely to have a high baseline match and to experience recognition with salient cues, whereas peripheral customers face a greater risk that salient cues magnify mismatch and trigger misrecognition. A key implication is that under segment-level targeting, the same personalization design can enhance the engagement of prototypical customers and reduce it among peripheral ones.

To study these issues in a realistic setting, we collaborate with a retailer that uses an embedding-based system to segment customers into three value tiers (Top, Middle, Bottom) and relies on these segments for targeting. Within this setting, we conduct a large-scale randomized field experiment that orthogonally manipulates two dimensions of segment-level personalization: message framing (symbolic "for-you" cues vs. generic wording) and recommendation method (algorithmic vs. popularity-based recommendations). The algorithmic condition uses the retailer's proprietary recommendation system, which leverages individual purchase history and segment membership in the embedding space. The popularity-based condition draws items from a curated overall bestseller list, reflecting aggregate demand rather than individual-level tailoring. This design isolates the causal effects of message- and content-level personalization and, crucially, allows us to test whether embedding-based, segment-relative prototypicality systematically moderates these effects as predicted by our theory. The inclusion of a popularity-based benchmark provides a transparent baseline and helps assess whether the observed heterogeneity reflects behavioral responses to perceived personalization rather than idiosyncrasies of the proprietary algorithm.

It is worthwhile noting that the experimental design we employ is essential for addressing our research questions. Identifying causal effects in personalization is difficult (if not impossible) in purely observational settings for several reasons: (a) endogeneity due to which customers are targeted, (b) the simultaneous deployment of multiple personalization tactics by firms, and (c) complications in longitudinal analyses arising from time-varying shocks.

Our findings show that, consistent with our theory, customer prototypicality is an important moderator that helps explain when segment-level personalization is effective for some customers but not others. When symbolic "for-you" cues accompany algorithmic recommendations, highly prototypical, high-value customers respond positively, consistent with recognition when salient personalization is supported by strong match quality. In contrast, for less prototypical customers within the same segment, the same design attenuates or reverses effects: symbolic framing raises expectations that the delivered content is less likely to satisfy, leading to disengagement. We propose a stylized reference-dependent framework to rationalize these results (e.g., Kahneman and Tversky 1979, Tversky and Kahneman 1991).

Alternative design variations reinforce our proposed account. Removing symbolic cues while retaining algorithmic recommendations makes personalization less salient, thereby attenuating both the positive response among highly prototypical customers and the negative response among peripheral customers, consistent with a reduction in both recognition and misrecognition. A similar moderation pattern emerges under popularity-based recommendations, suggesting that the observed effects are unlikely to be driven by algorithm-specific idiosyncrasies. Finally, standard observable covariates and global embedding distance (distance to the overall customer centroid) do not consistently reproduce these patterns, underscoring the usefulness of segment-relative prototypicality to reliably account for heterogeneous treatment effects.

Our work relates to four streams of research: personalization, recommender systems, prototype theory, and reference dependence. First, prior research shows that tailoring marketing can increase engagement, yet effects vary widely across customers (e.g., Ansari and Mela 2003, Lambrecht and Tucker 2013). Sahni et al. (2018) demonstrate that even uninformative cues (e.g., customer names) can shape engagement. We complement this work by studying more informative symbolic cues and showing that their effectiveness systematically varies based on the segment-level representativeness of customers. Second, research on recommendation systems finds that algorithmic recommendations can increase sales and engagement, but that effectiveness depends on preference alignment and context (e.g., Bodapati 2008, Pathak et al. 2010, Dietvorst et al. 2015). We add a theory-based account for within-segment heterogeneity by showing that the interaction of salience and match can generate recognition and misrecognition. Third, prototype theory argues that categories are structured around exemplars rather than rigid boundaries (Rosch 1975), and consumer research documents that perceived prototypicality shapes responsiveness (e.g., Bhattacharya and Sen 2003, Reed et al. 2012). We apply this perspective to modern embedding-based segmentation by operationalizing customer prototypicality in the firm's embedding space and demonstrating its value for explaining heterogeneous treatment effects. Finally, reference dependence provides a behavioral foundation for why outcomes in many contexts are driven by comparisons between what one receives and a benchmark (e.g., Kahneman and Tversky 1979, Tversky and Kahneman 1991). We use this framework to construct a stylized model that rationalizes our empirical results.

In summary, our contributions are threefold: theoretically, we propose and provide evidence for when segment-level personalization can enhance or undermine customer engagement; methodologically, we provide a scalable way to use embeddings for causal heterogeneity analysis through an interpretable, segment-relative construct of prototypicality;

and managerially, we offer guidance to improve the effectiveness of segment-level personalization.

The remainder of the paper is organized as follows. Section 2 develops the theoretical framework. Section 3 describes the field experiment and data. Section 4 presents the results. Section 5 interprets these findings and discusses implications for embedding-based targeting and personalization design. Finally, Section 6 concludes with contributions, limitations, and directions for future research.

2. Segment-Level Personalization and Customer Prototypicality

Advances in machine learning have made one-to-one personalization technically feasible. With that said, high implementation costs, sparse individual-level data, and increasing fairness and regulatory scrutiny continue to make one-to-one personalization impractical for many firms.[1] Segment-level personalization therefore serves as a pragmatic alternative, more scalable than individual-level targeting yet more tailored than mass marketing.

In practice, segment-level targeting is often implemented by combining message- and content-level personalization. For example, travel platforms segment customers by trip purpose, pairing leisure-oriented segments with inspirational content (e.g., "Top 10 Beach Getaways") and recommendations informed by prior behavior and similar users, while business-oriented segments receive messaging emphasizing flexibility and loyalty benefits. Financial-services firms similarly differentiate communications by value tier, with premium customers receiving lifestyle- and service-oriented offers and other customers receiving savings-focused promotions. Apparel retailers use psychographic segments and highlight sustainability-oriented collections for environmentally motivated shoppers and trend-forward updates for fashion seekers. Across these settings, segment-level personalization tailors both messaging and recommendations to segment-specific preferences.

2.1. Embeddings-Based Segmentation

Recent advances in representation learning have enabled a powerful bases for segmentation: low-dimensional customer embeddings, which are latent representations learned from

[1] At scale, personalized systems can be computationally expensive to deploy and maintain (e.g., CloudZero 2024). In many consumer settings, especially those with infrequent, high-consideration purchases, customers generate limited transaction data between purchases, constraining individual-level inference (e.g., Kiseleva et al. 2016). Moreover, individualized offers and personalized pricing have attracted scrutiny on fairness and equity grounds (e.g., POLITICO 2024).

transaction, browsing, and product-interaction data.[2] These embeddings replace hand-picked features with vectors that summarize complex behavioral patterns in a latent space. Clustering applied to these vectors then yields value- or behavior-based segments.

Yet, once segments are formed, hard assignment reduces the continuous embedding space to discrete labels, obscuring consequential within-segment heterogeneity. Even conditional on segment membership, customers vary continuously in the embedding space, and this dispersion may help explain who benefits from segment-level personalization. In principle, embeddings can be used in several ways to study heterogeneity in personalization response. One approach is to include raw embedding coordinates, or low-dimensional projections thereof, directly in response or uplift models as generic covariates (e.g., Bengio et al. 2013). A second approach is to recluster the embedding space into finer micro-segments and use segment indicators as moderators, extending traditional data-driven segmentation methods to high-dimensional representations (e.g., Wedel and Kamakura 2000). A third approach is to construct ad hoc similarity measures, such as distance to a global centroid, to behavioral archetypes (e.g., "heavy buyers"), or to specific products, and use these measures to moderate treatment effects, drawing on similarity- and distance-based theories of choice (e.g., Tversky 1977, Netzer et al. 2012).

While the approaches described above may improve predictive performance, they are not grounded in a behavioral account of why customers within the same segment should respond differently to personalization. Global distance measures abstract away from segment context (e.g., Tversky 1977), and micro-clustering increases segmentation granularity without explaining systematic, graded differences in typicality within segments (e.g., Wedel and Kamakura 2000). We therefore seek an interpretation of embedding-based segments that is theoretically motivated, parsimonious, and linked to how customers perceive and respond to marketing communications.

2.2. Customer Prototypicality

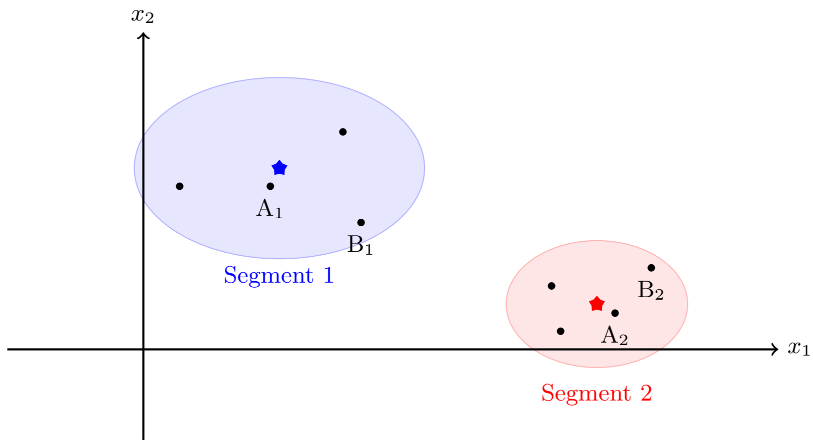
Prototype theory from cognitive psychology offers a useful (and thus far underutilized) lens for summarizing customer embeddings in a theoretically meaningful and managerially interpretable way. The theory posits that categories are organized around central exemplars

[2] For example, model-based embedding approaches have been proposed to represent customers in low-dimensional spaces suitable for segmentation in large, unstructured choice environments (e.g., Jagabathula et al. 2018). Related representation-learning approaches can produce informative customer embeddings for downstream customer understanding tasks (e.g., Chen et al. 2024).

and that members vary in typicality according to their similarity to these prototypes (Rosch 1975). We extend this perspective to embedding-based segmentation, and it implies that (a) the segment centroid provides a natural operationalization of the prototype, and (b) a customer's distance to the centroid captures how closely their preferences and behaviors align with the defining features of that segment.

In our context, the prototypicality of a customer should be measured with respect to the centroid of their assigned segment rather than to a single global centroid. Figure 1 illustrates this notion: customers are represented as points in a multidimensional feature space, and distance from the segment centroid reflects prototypicality. Those near the centroid (e.g., A_1, A_2) are more representative of the segment, whereas those near the segment boundary (e.g., B_1, B_2) are more peripheral and less aligned with the segment's core profile.

Figure 1 Customer Segments and Prototypicality



Note: Two customer segments in a two-dimensional feature space. Segment 1 (blue) and Segment 2 (red) are shown as ellipses, with stars indicating segment centroids. Black dots represent individual customers.

Building on this foundation, we use a firm's low-dimensional customer embeddings to operationalize segment-relative prototypicality. We define embedding-based prototypicality as a customer's distance from the centroid of their assigned segment in the embedding space, normalized within segment such that lower values indicate more prototypical customers and higher values indicate more peripheral ones.[3]

[3] While in the main analyses, prototypicality is computed using each customer's raw centroid distance and then normalized, the results are qualitatively robust to alternative functional transformations of distance (e.g., exponential). Appendix D reports these robustness checks.

Our mapping provides a parsimonious, theory-consistent way to translate high-dimensional embedding vectors into a psychologically grounded construct. This measure captures a customer's structural position within a segment as defined by the firm's representation space, integrating across the attributes that jointly determine segment membership. As a result, it can help explain heterogeneity in segment-level treatment effects more reliably than traditional moderators such as demographics or purchase history (e.g., Rossi et al. 1996, Ansari and Mela 2003).

2.3. Segment-level Personalization And Underlying Forces at Work

We posit that in segment-level personalization settings where personalization can be made salient, outcomes reflect the interplay of three underlying drives at work—baseline match, recognition, and misrecognition—and that customer prototypicality shapes how these forces operate. While prior research (as noted below) has separately examined the impact of each driver on personalization, we integrate these strands into a unified framework.

Baseline Match (Functional Accuracy) Baseline match is the extent to which a recommendation objectively aligns with a customer's preferences, prior to any subjective inference about the firm's intent. For example, a music-streaming service recommending a jazz playlist achieves high baseline match for a habitual jazz listener but low match for someone who primarily listens to hip-hop. Prior research on recommender systems shows that higher predictive accuracy can increase engagement when consumers evaluate content primarily on relevance (e.g., Adomavicius and Tuzhilin 2005, Ricci et al. 2015).

Recognition (Perceived Personalization) Recognition captures the positive response that can arise when customers perceive a recommendation as intentionally tailored and relevant to them. It transforms functional accuracy into a feeling of being understood, consistent with self-verification and identity-congruence effects in consumer behavior (e.g., Aaker 2010, Reed et al. 2012). Symbolic cues such as "Recommended for you" heighten this sense of personalization by signaling intentionality (e.g., Bleier and Eisenbeiss 2015, White et al. 2019). For instance, an online bookstore might recommend a new novel by a customer's favorite author with the label "For your reading list," simultaneously matching preferences and signaling recognition.

Misrecognition (Expectation Violation) Misrecognition can arise when personalization is made salient but fails to meet the expectations it creates. When symbolic cues promise individualized attention, customers may evaluate the offered content against a benchmark, and negative reactions can emerge when the claim of "for you" falls short (e.g., Brehm 1966, Oliver 1980, Kahneman and Tversky 1979). For example, when a movie platform proclaims "Movies picked just for you" yet recommends a generic blockbuster to a documentary enthusiast, the mismatch undermines credibility and signals a failure to understand customer preferences.

We suggest that prototypicality can play an important role in shaping the relative magnitude of the three forces at work. Prototypical customers may be easier for algorithms to predict and therefore tend to experience higher baseline match. They are also more likely to perceive recognition as recommendations align with their expectations. In contrast, less prototypical customers occupy the segment's periphery and, as their behavior is harder to predict, the recommendations they receive are less likely to fit. For these customers, salient personalization can amplify mismatch, triggering misrecognition when content reflects segment averages rather than their tastes. As a result, under segment-level targeting, the same personalization design can elicit recognition among prototypical customers while inducing misrecognition among peripheral ones.

3. Field Experiment

We test our prototype-based framework in a carefully designed field experiment. We first describe the experimental context and design, then lay out the implementation and data.

3.1. Experimental Context

Addressing our research questions requires a setting with several key features. First, the firm must use segment-level personalization so that treatment effects can be evaluated within segments. Second, the context must yield clear behavioral outcomes (e.g., purchases) to support rigorous causal inference. Third, segmentation must be based on customer embeddings, allowing construction of each customer's prototypicality relative to their segment centroid. Fourth, personalization tactics must vary experimentally to identify the causal effects of distinct personalization designs. Finally, the setting must be managerially relevant so that insights on customer prototypicality can inform real-world targeting decisions.

To meet these criteria, we partnered with a large Asia-based retailer specializing in outdoor recreation products (e.g., apparel, footwear, and accessories) and conducted a randomized field experiment. The company operates across offline and online channels, with brick-and-mortar stores forming the backbone of the business and online sales expanding steadily. Known for its strong brand reputation among outdoor enthusiasts and its large, loyal customer base, the retailer uses embedding-based segmentation and places significant emphasis on personalization, making it an especially appropriate setting for our theory test.

The retailer classifies customers into three value-based segments (Top, Middle, and Bottom) using a proprietary clustering algorithm applied to customer embeddings, which are constructed from rich behavioral and profile features, including RFM-based measures. This embedding-based segmentation is central to its marketing operations and guides targeted communications. The setting's scale, established embedding-based segments, and active use of personalization make it well suited for testing whether our embedding-derived measure of segment-relative prototypicality moderates the effectiveness of different personalization strategies.

3.2. Experiment Design

Our field experiment orthogonally manipulated the two key dimensions of personalization: (1) message framing, which varies symbolic cues at the message level (symbolic "for-you" versus generic wording), and (2) recommendation method, which varies content selection (algorithmic versus popularity-based recommendations). Each factor had two levels, yielding a 2×2 design. This design isolates the causal effects of each personalization approach and their interaction and, crucially for our purposes, enables a direct test of whether embedding-based customer prototypicality systematically moderates these effects as predicted by our theoretical framework.

The first factor, message framing, manipulated the presence of symbolic cues. In the symbolic condition, messages made recognition salient (e.g., "Recommended for you," "We picked this just for you"), intended to heighten perceived personalization and engagement. In the generic condition, messages used neutral phrasing such as "Check this out," providing no symbolic signal. This generic version reflects a mass-marketing baseline. It allows us to measure the incremental effect of recognition and to test whether prototypical and peripheral customers respond differently when recognition is made salient.

The second factor, recommendation method, manipulated whether featured products were selected algorithmically or based on aggregate popularity. In the algorithmic condition, items were chosen by the retailer's proprietary recommendation system, which leverages individual purchase history and segment membership in the embedding space. In the popularity-based condition, items were drawn from a curated list of overall bestsellers, reflecting broad demand rather than individual-level tailoring. Because featuring popular or trending items is a widespread industry practice, this benchmark provides a transparent and managerially relevant comparison.

Our field experiment offers an externally valid setting to test the theoretical role of prototypicality in explaining heterogeneity in segment-level treatment effects. Although the proprietary recommendation algorithm limits visibility into its internal workings, the popularity-based benchmark provides a transparent baseline. By comparing algorithmic and popularity-based recommendations within the same randomized design, we test whether embedding-derived, segment-relative prototypicality consistently organizes heterogeneous responses, even after accounting for targeting approaches, in a manner consistent with our recognition-versus-misrecognition account. Thus, our context offers a strong test for the validity of customer prototypicality.

The design yielded four experimental conditions. In G1, symbolic cues were paired with algorithmic recommendations; in G2, they were paired with popularity-based recommendations. G3 received generic messages with algorithmic recommendations, and G4 received generic messages with popularity-based recommendations, reflecting a mass-marketing approach and serving as the control condition.

3.3. Implementation

The experiment was conducted in summer 2024 among active customers, defined as those who had made at least one purchase in the preceding twelve months. From this population, we drew a stratified random sample of 96,716 customers to ensure proportional representation across the firm's value-based segments.[4] Customers were then randomly assigned to one of the four groups in Table 1, with randomization stratified by segment. This procedure minimized the risk of baseline imbalances that could confound treatment effect estimates, which is especially important because segment membership is central to our moderation analysis. Segment proportions in the final sample closely matched those in the firm's overall customer base (Table 1), and balance checks confirm comparability across conditions at both the overall and segment levels (Section 4).

[4] This focus aligned with the firm's operational practices while simplifying implementation by avoiding complications from finer-grained segmentation. Customers who had opted out of marketing communications were excluded to comply with privacy and data protection regulations.

Table 1 Experiment Design

	Recommendation Method	
Message Framing	Algorithmic	Popularity-Based
Symbolic Segment	G1 (63,961)	G2 (10,928)
Top	0.016	0.017
Middle	0.229	0.224
Bottom	0.755	0.760
Generic Segment	G3 (10,879)	G4 (10,948)
Top	0.015	0.015
Middle	0.227	0.233
Bottom	0.757	0.752

Note: Numbers in parentheses indicate the number of customers in each group. Decimal values show the proportion of customers in each segment within the corresponding group.

Although an equal split was feasible, the partner firm was particularly interested in the combined effect of symbolic cues and algorithmic recommendations (G1). Accordingly, we assigned approximately two-thirds of the sample to G1, with the remaining third evenly divided among G2–G4. This allocation ensured adequate statistical power to detect effects in the focal treatment while maintaining valid comparisons across conditions.

Customers in G1 and G2 received messages containing the symbolic phrase “Recommended for you” along with segment-specific copy tailored to their assigned segment. For example, Top-segment messages highlighted individuality and novelty (“Move seamlessly between trekking and everyday life”), Middle-segment messages emphasized practicality (“Achieve both style and practicality in your everyday life”), and Bottom-segment messages stressed value and accessibility (“Complete your outdoor style with a rational choice”). No messages referenced rank or segment tier explicitly.[5] Customers in G3 and G4 received a generic, mass-market message without symbolic cues.

Within each message, product recommendations were either personalized (G1, G3) or generic (G2, G4). In the personalized conditions, items were selected by the retailer’s proprietary algorithm, which leverages individual purchase history and the purchase patterns of other customers within the same segment. In the generic conditions, one item was randomly selected from the retailer’s curated bestseller list. Messages were delivered via mobile

[5] Although segment membership is not disclosed, our construct of representativeness can be interpreted as serving as a proxy for customers’ subjective sense of fit with a segment (e.g., Reed et al. 2012). Because messages contain no explicit rank or tier cues, the context provides a stringent test of our framework.

SMS and included product information and a link to a landing page. The landing page featured the same item shown in the message along with additional details, while all other elements (e.g., layout, visuals, and promotional text) were held constant across conditions.

All customers received the same promotion, $20 off purchases of $100 or more, to maintain comparability across groups. To maintain robust measurement of treatment effects, no other targeted promotions were sent to participants during the campaign. Although store-wide promotions were available, they were not proactively communicated to these customers. The campaign ran for two weeks, followed by a four-week post-campaign period with no additional personalized messages. This design allowed us to assess both immediate campaign responses and short-run post-campaign responses.

3.4. Data and Measures

Our dataset had three components: detailed transaction records capturing purchase behavior, demographic information such as age and gender, and customer segmentation data, including each customer's segment assignment and the embedding vector for each customer that the firm uses for segmentation.[6] Unlike studies that infer latent representations from observables (e.g., Jagabathula et al. 2018), we observe the firm's embedding space directly, allowing us to compute embedding-based prototypicality on a continuous scale without the additional measurement error from inferring embeddings ourselves. This unusually rich setting enables a direct test of the value of embedding-based prototypicality as a moderator of segment-level personalization.

We define embedding-based prototypicality as each customer's distance from the centroid of their assigned segment in the embedding space, mirroring prototype theory's view of categories as organized around central exemplars. Let $\mathbf{x}_i \in \mathbb{R}^d$ denote the embedding of customer i, assigned to segment w whose centroid $\mathbf{c}_w$ is the mean embedding vector of customers in that segment. Then:

$$p_i = d(\mathbf{x}_i, \mathbf{c}_w), \tag{1}$$

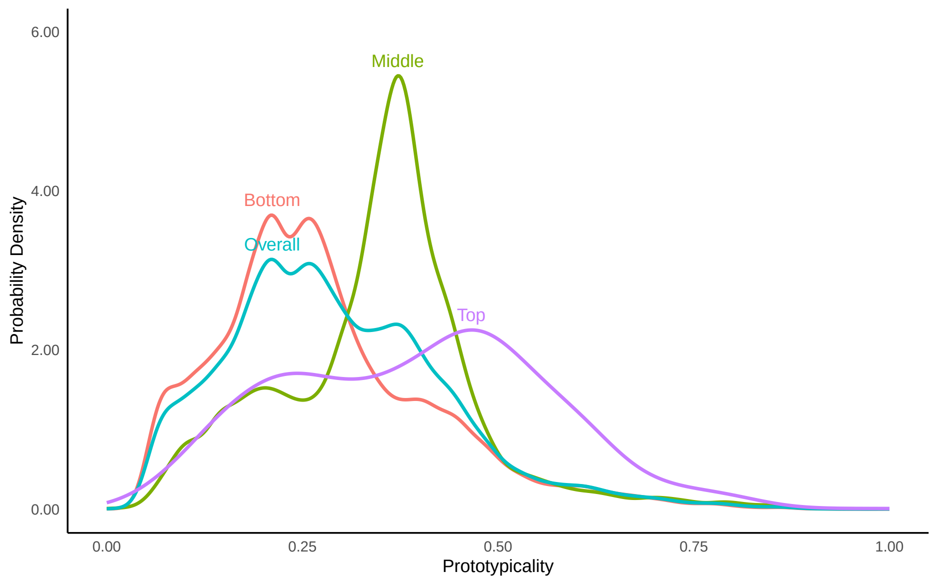
where $d(\cdot)$ denotes Euclidean distance. To ensure comparability, we min-max normalize p_i within each segment to rescale prototypicality to $[0, 1]$. Lower values indicate customers

[6] All purchase amounts were originally recorded in the local currency and converted to U.S. dollars using the average exchange rate over the data period.

closer to the segment centroid (more prototypical), whereas higher values indicate more peripheral positioning.

Figure 2 shows the distribution of embedding-based prototypicality by segment in the experiment. The figure reveals substantial within-segment dispersion, providing the variation needed to test whether personalization effects depend on prototypicality.

Figure 2 **Distribution of Embedding-Based Prototypicality**



Note: Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher values are more peripheral. Curves show probability density estimates computed separately for each segment (Top, Middle, Bottom) and for all customers combined (Overall).

For each customer, we observe three distinct periods: a twelve-month pre-campaign baseline (July 2023–June 2024) used to assess baseline purchase behavior and conduct randomization checks for comparability across experimental groups; the two-week campaign period capturing immediate treatment responses; and a four-week post-campaign window to assess persistence or decay of effects. Together, these data provide a strong empirical basis for identifying the causal impact of symbolic and algorithmic personalization and for examining how these effects vary with customer prototypicality.

4. Results

This section presents the main findings using purchase incidence as the primary outcome of customer responsiveness and as our main measure of the effectiveness of segment-level personalization.[7]

We first verify the validity of the experimental design. Table 2 reports baseline customer characteristics across the four conditions, separately for the Top, Middle, and Bottom segments. The covariates are well balanced: within each segment, pre-treatment characteristics are comparable across conditions, with differences that are small and not statistically significant at conventional levels.[8] This suggests that treatment assignment is not confounded by pre-existing differences.

Table 2 Pre-treatment Customer Characteristics by Segment

	G1	G2	G3	G4
Panel A. Top Segment				
Age	50.54	48.45	49.58	50.20
Gender (1 if male)	0.56	0.62	0.60	0.60
Tenure (months)	33.52	32.98	33.22	33.31
Recency (months)	2.13	2.33	2.40	2.04
Frequency	8.43	7.64	7.67	8.44
Monetary ($)	2305.73	1976.47	2148.25	2214.66
Panel B. Middle Segment				
Age	50.55	50.51	50.83	50.87
Gender (1 if male)	0.68	0.68	0.67	0.69
Tenure (months)	14.10	14.37	14.34	14.21
Recency (months)	6.22	6.18	6.14	6.13
Frequency	1.70	1.68	1.73	1.69
Monetary ($)	498.62	489.12	496.49	495.64
Panel C. Bottom Segment				
Age	52.90	52.99	53.01	52.73
Gender (1 if male)	0.67	0.67	0.66	0.66
Tenure (months)	14.06	14.05	13.73	13.99
Recency (months)	5.54	5.59	5.44	5.50
Frequency	1.15	1.15	1.16	1.15
Monetary ($)	144.61	145.17	148.38	144.83

Note: Tenure is measured as the elapsed time since joining the firm's loyalty program.

[7] Our partner firm has a strong offline presence but a modest online channel, so we aggregate purchases across channels to capture overall shopping behavior, recognizing that customers often evaluate products in-store before purchasing. Because take-up of any specific recommended items is sparse in a large assortment, we use purchase incidence as our primary measure of responsiveness to segment-level personalization. Later, we also describe the results for total spending as an outcome.

[8] Balance checks for the overall sample show the same pattern (Appendix A).

Next, we assess whether embedding-based prototypicality is balanced across conditions. Table 3 shows that means and standard deviations of prototypicality are nearly identical across groups, both overall and within each segment, indicating comparable baseline proximity to segment centroids. Top-segment customers are, on average, farther from the segment centroid, consistent with greater behavioral variation among high-value customers, whereas Middle and Bottom customers exhibit less variation. Within each segment, the distributions are balanced across conditions, further supporting the validity of randomization. Accordingly, subsequent differences in outcomes are unlikely to reflect pre-treatment differences in prototypicality.

Table 3 Pre-treatment Embedding-Based Prototypicality

	G1	G2	G3	G4
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Overall	0.29 (0.14)	0.29 (0.14)	0.29 (0.14)	0.29 (0.14)
Segment				
Top	0.40 (0.17)	0.38 (0.16)	0.39 (0.17)	0.39 (0.18)
Middle	0.35 (0.13)	0.35 (0.13)	0.34 (0.13)	0.34 (0.13)
Bottom	0.28 (0.14)	0.28 (0.14)	0.28 (0.14)	0.28 (0.14)

Note: Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

The remainder of this section proceeds as follows. Section 4.1 examines how customer prototypicality moderates the effects of symbolic cues. Section 4.2 then extends the analysis to the joint effects of symbolic cues, segment tier, and algorithmic personalization. Finally, Section 4.3 explores the underlying mechanism, distinguishing expectation-based misrecognition from alternative explanations.

4.1. Effects of Symbolic Cues and Prototypicality

Prior research has shown that even uninformative forms of personalization, such as including a customer's name in an email header, can influence customer behavior (e.g., Sahni et al. 2018). Building on this literature, we examine whether more informative symbolic cues that explicitly claim personalization (e.g., "for you") shape customer response, and whether these effects are moderated by customer prototypicality.

Table 4 reports logistic regressions of purchase incidence on message framing and its interactions with segment tier and prototypicality. The key independent variable, Treatment, equals 1 for symbolic messages and 0 for generic ones. Segment tier is represented by Top and Bottom indicators, with the Middle segment as the reference group. Prototypicality is measured as distance from the segment centroid, with lower values indicating greater alignment with the segment core (more prototypical). Column 1 estimates the main effect of symbolic cues, while Column 2 adds two-way and three-way interactions to test for moderation by segment tier and prototypicality. The main effect therefore corresponds to a Middle-segment customer at the segment centroid (prototypicality = 0).

Table 4 Effects of Symbolic Cues and Prototypicality

	(G1,G2) vs. (G3,G4) Symbolic vs. Generic	
Variable	(1)	(2)
Treatment	0.05	-0.43
	(0.10)	(0.28)
Segment		
Top	2.84^{***}	3.16^{***}
	(0.14)	(0.39)
Bottom	-0.45^{***}	-0.50^{*}
	(0.10)	(0.28)
Prototypicality		1.72^{***}
		(0.63)
Treatment × Segment		
Treatment × Top	0.12	1.20^{***}
	(0.16)	(0.43)
Treatment × Bottom	-0.15	0.13
	(0.12)	(0.32)
Treatment × Prototypicality		1.24^{*}
		(0.71)
Segment × Prototypicality		
Top × Prototypicality		-1.00
		(0.93)
Bottom × Prototypicality		0.49
		(0.73)
Treatment × Segment × Prototypicality		
Treatment × Top × Prototypicality		-2.78^{***}
		(1.04)
Treatment × Bottom × Prototypicality		-0.66
		(0.83)
Constant	-3.54^{***}	-4.15^{***}
	(0.09)	(0.25)
Log Likelihood	-10154.56	-10006.07
AIC	20321.13	20036.15
BIC	20378.01	20149.90
Num. Obs.	96,716	96,716

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment denotes message framing (1 = symbolic, 0 = generic) and is coded as 1 for customers in the focal treatment group (G1 or G2) and 0 for those in the control group (G3 or G4). Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Column 1 shows no statistically significant average effect of symbolic message framing on purchase likelihood. Although this result may appear to contrast with prior findings on name-based subject-line personalization (e.g., Sahni et al. 2018), our setting involves informative, header-level "for you" cues layered on top of segment-level offers, which is a conceptually distinct form of personalization that operates through perceived recognition and realized fit. Moreover, as we show in detail later, there is a nuanced interaction between symbolic cues and the recommendation method that is not captured by this main-effects model. Column 2 provides an initial indication of this interaction. Symbolic cues increase purchase propensity for Top-segment customers, as indicated by the positive two-way interaction between treatment and Top-segment status (1.20, $p < .01$), consistent with the idea that salient recognition can boost responses among high-value customers (e.g., McKinsey & Company 2021). However, the negative three-way interaction among treatment, Top-segment status, and prototypicality (-2.78, $p < .01$) implies that this benefit attenuates as Top customers become more peripheral within their segment.[9] Taken together, among the firm's most valuable customers, symbolic cues increase responses for those closely aligned with the segment core but diminish responses as customers become more peripheral.

These results indicate that prototypicality plays an important role in accounting for the effects of informative header-level cues. Because these estimates pool across recommendation methods, we next investigate how symbolic cues interact with algorithmic targeting to jointly shape customer responses.

4.2. Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting

This analysis distinguishes the effects of recognition via symbolic cues, the incremental match quality from algorithmic recommendations (relative to popularity-based recommendations), and their interaction. Specifically, the G1–G4 comparison captures the combined effect of symbolic cues and algorithmic recommendations; the G3–G4 comparison isolates the incremental effect of algorithmic recommendations in the absence of symbolic cues; and the G1–G3 comparison identifies the incremental effect of symbolic recognition when the recommendation method is held constant.

[9] As expected, Top customers purchase more than the Middle-segment reference group (3.16, $p < .01$), whereas Bottom customers purchase less (-0.50, $p < .10$), confirming baseline differences in purchase propensity. The interpretation of the Main effects are conditional on the included interactions and the coding of the variables. The key result is the slope reversal implied by the three-way interaction.

Table 5 reports the results from logistic regressions.[10] The models include segment tier, prototypicality, and their interactions. The treatment indicator equals 1 for customers in the focal treatment group and 0 for those in the corresponding control. In all models, as before, the main effect corresponds to a Middle-segment customer at the segment centroid.

Table 5 Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	-0.50	-0.18	-0.33
	(0.37)	(0.50)	(0.39)
Segment: Top	3.35^{***}	3.35^{***}	2.93^{***}
	(0.52)	(0.52)	(0.57)
Segment: Bottom	-0.57	-0.57	-0.44
	(0.39)	(0.39)	(0.41)
Prototypicality	1.57^{*}	1.57^{*}	1.89^{**}
	(0.88)	(0.88)	(0.91)
Treatment × Segment			
Treatment × Top	0.94^{*}	-0.42	1.36^{**}
	(0.57)	(0.77)	(0.61)
Treatment × Bottom	0.18	0.13	0.05
	(0.42)	(0.56)	(0.44)
Treatment × Prototypicality	1.36	0.33	1.03
	(0.95)	(1.27)	(0.98)
Segment × Prototypicality			
Top × Prototypicality	-1.20	-1.20	-0.71
	(1.28)	(1.28)	(1.35)
Bottom × Prototypicality	0.81	0.81	0.16
	(1.01)	(1.01)	(1.06)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-2.45^{*}	0.49	-2.94^{**}
	(1.38)	(1.86)	(1.45)
Treatment × Bottom × Prototypicality	-0.91	-0.65	-0.26
	(1.09)	(1.47)	(1.14)
Constant	-4.06^{***}	-4.06^{***}	-4.24^{***}
	(0.34)	(0.34)	(0.36)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Column 1 shows that the combined personalization treatment increases purchase propensity among highly prototypical Top-segment customers (0.94, $p < .10$). This pattern is consistent with the idea that symbolic cues can amplify the impact of algorithmic personalization when the message is perceived as both accurate and personally relevant. However,

[10] Results are robust to the inclusion of observed covariates (Appendix B).

the negative three-way interaction among treatment, Top-segment status, and prototypicality (-2.45, $p < .10$) indicates that this benefit diminishes as Top customers become less prototypical. Thus, symbolic recognition coupled with misaligned recommendations appears to backfire.

Column 2 isolates the impact of algorithmic targeting on purchase propensity (G3 vs. G4). When recommendations are presented without symbolic cues, personalization is less salient, so responses are more likely to reflect functional match rather than explicit recognition. The results show no significant interaction effects, indicating that algorithmic recommendations without symbolic framing neither generate strong gains among highly prototypical customers nor trigger backlash among peripheral ones.

Column 3 isolates the impact of salience on purchase propensity given algorithmic recommendation (G1 vs. G3). The presence of the negative three-way interaction among treatment, Top-segment status, and prototypicality (-2.94, $p < .05$) indicates that symbolic cues amplify both the upside and downside of personalization, heightening engagement when realized fit is high but provoking disappointment when it falls short.

Overall, the results indicate that customer prototypicality shapes the effectiveness of segment-level personalization. Without symbolic cues, algorithmic recommendations do not outperform popularity-based recommendations (Column 2). This pattern is somewhat surprising and could raise concerns about algorithm quality as a driver of our findings. However, when symbolic cues are present, the algorithm performs as expected (Column 3). Taken together, these results suggest that salience is necessary for recommendations to influence outcomes. Thus, our results not only showcase the importance of salience in marketing communications replicating prior finding (e.g., Sahni et al. 2018) but also reveal the nuanced interaction between informative symbolic cues and the recommendation method. Finally, the largest effects emerge among the firm's top customers. Although this group represents a relatively small share of the customer base, it generates a disproportionate share of revenues, highlighting the strategic value of personalizing offers where it matters most.

Next, we examine a few alternative explanations, including those related to algorithm quality, and explain why they are unlikely to account for our results.

4.3. Alternative Explanations

The results from the algorithmic conditions could reflect other drivers at work unrelated to the proposed expectation-based account. It is plausible that the proprietary recommendation algorithm fits some customers better than others or embeds other aspects that are correlated with their prototypicality. For example, if predictive accuracy declines for less-prototypical customers, the observed moderation could simply reflect variation in recommendation quality rather than behavioral responses to perceived personalization. Similarly, algorithmic recommendations may differentially expose customers to popular versus niche products, generating heterogeneous effects through composition rather than recognition. To mitigate these and other concerns, we conduct a series of falsification and robustness tests using the popularity-based (non-algorithmic) conditions (G2 vs. G4).

4.3.1. Popularity-Based Benchmark Test Table 6 reports logistic regressions comparing customers exposed to symbolic cues paired with popularity-based recommendations (G2) to those in the generic control condition (G4).[11] As before, the models include segment tier, prototypicality, and their interactions; the main effect corresponds to a Middle-segment customer at the segment centroid.

Column 1 presents the full-sample comparison, testing whether the moderating role of prototypicality persists when recommendations are based only on aggregate demand. Because popularity-based recommendations are not tailored to individual fit, any moderation in this condition cannot be due to algorithmic precision. Yet the same negative moderation by prototypicality emerges: symbolic cues are less effective for less prototypical customers even when recommendations are not personalized. Thus, the observed pattern of results is unlikely due to algorithmic quality or any difference across people in the training data. Instead, the results support our expectation-based interpretation.

[11] Results are robust to the inclusion of observed covariates (Appendix C).

Table 6 Effects of Symbolic Cues and Prototypicality in Popularity-Based Targeting

	G2 vs. G4		
Variable	(1) All	(2) Mainstream	(3) Long-Tail
Treatment	-0.62	0.11	-0.98
	(0.50)	(0.87)	(0.60)
Segment: Top	3.35^{***}	4.73^{***}	2.78^{***}
	(0.52)	(1.73)	(0.57)
Segment: Bottom	-0.57	-0.32	-0.87^{*}
	(0.39)	(0.67)	(0.47)
Prototypicality	1.57^{*}	0.94	1.48
	(0.88)	(1.67)	(0.96)
Treatment × Segment			
Treatment × Top	1.43^{*}	-2.84	1.98^{**}
	(0.75)	(2.66)	(0.83)
Treatment × Bottom	0.28	-0.05	0.24
	(0.57)	(0.95)	(0.72)
Treatment × Prototypicality	1.61	-0.19	2.39^{*}
	(1.23)	(2.34)	(1.39)
Segment × Prototypicality			
Top × Prototypicality	-1.20	-5.59	-0.88
	(1.28)	(5.08)	(1.36)
Bottom × Prototypicality	0.81	0.90	0.95
	(1.01)	(1.85)	(1.17)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-3.33^{*}	7.16	-4.45^{**}
	(1.82)	(6.73)	(1.96)
Treatment × Bottom × Prototypicality	-1.35	-0.81	-1.13
	(1.44)	(2.64)	(1.70)
Constant	-4.06^{***}	-4.32^{***}	-3.56^{***}
	(0.34)	(0.62)	(0.39)
Num. Obs.	21,876	13,424	8,452

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment is coded as 1 if the customer belongs to G2 and 0 if in the control group (G4). Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

A deeper exploration helps explain why popularity-based targeting produces the same pattern. By design, popularity-based recommendations tend to fit customers whose preferences align with mainstream demand (majority), but they are less likely to match customers with niche, long-tail preferences (minority). As a result, since the baseline match would be relatively high for mainstream customers, prototypicality may play a limited moderating role. In contrast, for customers with niche preferences, baseline match is lower, and prototypicality can be an important moderator of whether symbolic cues generate recognition or misrecognition.

To operationalize this line of thought, we classify customers as mainstream versus long-tail following the long-tail framework of Anderson (2006). The retailer's assortment exhibits a classic long-tail distribution in which a small number of products account for most sales and a large number of niche items sell infrequently (e.g., Fleder and Hosanagar 2009, Lee and Hosanagar 2019, Lichtenberg et al. 2024). We define popular products as those comprising the top 80% of cumulative unit sales during the pre-campaign period (July 2023–June 2024).[12] To improve classification reliability, we supplemented this window with transaction data from January 2022–June 2023. Customers whose purchases are concentrated in these top-selling products are coded as mainstream, whereas those with substantial purchases outside this set are coded as long-tail. This classification allows us to isolate cases in which popularity-based recommendations are likely to deliver high baseline fit from those in which they are not.

Columns 2 and 3 in Table 6 report results separately for mainstream and long-tail customers. Among mainstream customers (Column 2), popularity-based recommendations already deliver a strong baseline match, and the moderating effect of prototypicality is negligible. In contrast, among long-tail customers (Column 3), baseline match is lower, and the three-way interaction between symbolic framing, Top-segment status, and prototypicality is strongly negative and statistically significant (-4.45, $p < .05$), indicating that symbolic cues backfire precisely where expectations of personalization are most likely to be violated. Taken together, the absence of prototypicality moderation among mainstream customers and its presence among long-tail customers reinforces the interpretation that these negative effects reflect expectation-based misrecognition rather than algorithmic bias.

These results provide a stringent test of our proposed account. If the observed asymmetries merely reflected technical properties of the recommendation algorithm, they would be expected to attenuate or disappear when content is driven by aggregate popularity rather than individualized targeting. Instead, the asymmetry persists precisely among customers whose niche preferences are most likely to create large expectation–realization gaps, providing strong evidence consistent with the expectation-violation account.

4.3.2. Alternative Specifications for Distance to Segment Centroid To assess robustness of the results to how distance is specified, we re-parameterize the within-segment distance to the segment centroid using several standard monotone distance-to-similarity mappings. Let $\tilde{p}_i$ denote customer i's (normalized) distance to the centroid. We construct a prototype-proximity index $f(\tilde{p}_i)$ via: (1) exponential decay, $f_{\exp}(\tilde{p}_i) = 1 - \exp(-\lambda\tilde{p}_i)$; (2)

[12] Returned items are excluded. Results are robust to alternative popularity thresholds and are available from the authors upon request.

Gaussian decay, $f_{\text{gauss}}(\tilde{p}_i) = 1 - \exp(-\lambda \tilde{p}_i^2)$; (3) a logistic saturation, $f_{\text{logit}}(\tilde{p}_i) = 1 - \left[1 + \exp\{\lambda(\tilde{p}_i - 1)\}\right]^{-1}$; and (4) an inverse distance, $f_{\text{inv}}(\tilde{p}_i) = 1 - (1 + \tilde{p}_i)^{-1}$. Across all mappings, the moderation patterns are qualitatively unchanged, indicating that our conclusions are not driven by a particular functional form used to translate distance into proximity. Appendix D reports the full results.

4.3.3. Global Prototypicality A central tenet of our theoretical framework is that the moderation of treatment effects should depend on segment-relative positioning of customers rather than their global centrality. To test this notion, we construct a global distance measure based on each customer's proximity to the centroid of the entire customer population rather than to the centroid of their assigned segment. Specifically, we compute the Euclidean distance between each customer's embedding vector and the overall sample centroid, rescale the distances to $[0, 1]$, and substitute this global measure for segment-level prototypicality. This measure captures how representative a customer is in the firm's overall embedding space, irrespective of segment membership. If the observed moderation simply reflected global extremeness or centrality in the population, this global measure should exhibit similar interaction patterns. However, as reported in Appendix E, the interactions with symbolic framing and segment tier are statistically insignificant at conventional levels.

The null result for global prototypicality strengthens the theoretical interpretation of our findings. Conceptually, it suggests that personalization responses are organized by segment-relative identity, as implied by prototype theory, rather than by generic centrality in the overall customer base. Statistically, it alleviates concerns that the results are driven by an arbitrary distance-based measure and underscores the specificity and theoretical relevance of segment-relative prototypicality.

4.3.4. Demographics We estimate the main models replacing prototypicality with a few observable features such as age, gender, or tenure, respectively, as moderators. As reported in Appendix F, none of the demographic variables moderates the treatment effects reliably. For example, gender for top segment customers interacts with their prototypicality in one model variant but not so in others. Similarly, both age and tenure show simple effects (e.g., older customer are more likely to purchase) but no three-way interactions.

Taken together, the results for global prototypicality and for demographic moderators reinforce our theoretical claim that the heterogeneity we document arises from within-segment representativeness.

4.4. Additional Analyses

We assess the boundary conditions of our findings with two analyses. First, we examine whether message framing and prototypicality influence customer spending conditional on purchase. Second, we test whether the observed effects persist beyond the two-week campaign period.

4.4.1. Intensive Margin We investigate whether personalization also influences spending conditional on purchase (the intensive margin). Personalization could shape both incidence and spend decisions. However, once the purchase hurdle is crossed, spending may be determined more by needs than by message framing. We find no systematic effect of personalization on spending across all specifications (Appendix G). Thus, the effects documented earlier are limited to purchase incidence.

4.4.2. Post-Campaign Outcomes We examine whether the effects extend beyond the campaign period by tracking customer behavior during the four-week post-campaign window. The gains from personalization observed during the campaign largely dissipate once the campaign ends (Appendix H). This pattern confirms with intuition: a single exposure is expected to exert only a short-lived influence on customer behavior.

5. A Reference Dependent Framework for Personalization

Our findings reveal that the impact of personalization depends on the alignment between baseline match quality and recognition: customers respond positively when perceived relevance meets expectations and negatively when it falls short. Because our field setting does not allow direct measurement of perceived fit or disappointment, we present the mechanism as consistent with, rather than conclusively established by, the observed patterns.

There are three main findings. First, with no salience, algorithmic recommendations have limited incremental impact: customers are less likely to interpret content as personally tailored and respond primarily to realized match quality. Second, with salience, responses diverge by customer type: highly prototypical, high-value customers respond positively, whereas less prototypical customers disengage. Third, the pattern of results is unlikely due to the firm's proprietary algorithm: similar moderation patterns emerge under a transparent popularity-based benchmark. To interpret these patterns systematically, we propose a reference-dependent framework linking personalization salience to expected versus realized fit and, in turn, to recognition or misrecognition. This framework helps in organizing the underlying drivers and informs the design of segment-level personalization.

5.1. A Reference-Dependent Model of Personalization

We draw inspiration from prospect theory, which posits that outcomes are evaluated relative to a reference point (Kahneman and Tversky 1979). Related forms of reference dependence and asymmetry in outcomes have been documented across a range of consumer decision-making contexts (e.g., Thaler 1985, Tversky and Kahneman 1991, Heath et al. 1999). In applying this perspective, we use reference dependence solely to structure the contrast between recognition and misrecognition for prototypical versus peripheral customers. We do not make any claims on effect magnitudes, such as whether losses loom larger than gains, as in the standard prospect theory formulation.

In our setting, the reference point can be interpreted as the expected fit implied by a message claiming to be "for you." Customers are likely to experience a gain (recognition) when realized fit meets or exceeds this expectation, and a loss (misrecognition) when it falls short.

Let a customer with prototypicality p evaluate a recommendation with realized fit $m(p)$ under salience $S \in \{0, 1\}$:

$$U = m(p) + S\Big[k - \big[\tau(p) - m(p)\big]_+\Big].$$

Here, $m(p)$ denotes realized match quality, S indicates whether personalization is salient, and k captures the baseline recognition utility from salient personalization when expectations are met. When $S = 0$, utility depends solely on match quality. When $S = 1$, customers evaluate outcomes relative to an expectation threshold $\tau(p)$, the fit implied by symbolic personalization. The operator $[x]_+ = \max(x, 0)$ captures expectation violation only when realized fit falls short. Thus, salience activates reference-dependent evaluation beyond functional match.

This structure helps organize the three forces relevant for segment-level personalization. First, baseline match $m(p)$ captures the functional accuracy of recommendations. Second, recognition $S \cdot k$ captures the positive response to symbolic cues that make personalization salient, enhancing engagement when supported by adequate fit. Third, the expectation-violation penalty $-S \cdot [\tau(p) - m(p)]_+$ captures the negative reaction when salience raises expectations that realized fit fails to meet, generating misrecognition that offsets or even reverses recognition gains.

5.2. Implications for the Empirical Results

The framework accommodates our empirical results by allowing both realized fit $m(p)$ and expectation threshold $\tau(p)$ to vary with prototypicality p. The relative slopes of these two functions determine whether salience yields recognition or misrecognition: when $m(p)$ declines more steeply than $\tau(p)$, the expectation gap widens toward the periphery, producing the responses observed in the data.

When personalization is not salient ($S = 0$), utility depends solely on match quality, and algorithmic targeting may improve outcomes by increasing baseline accuracy. When salience is introduced ($S = 1$), both recognition and misrecognition become possible: recognition dominates for prototypical customers whose realized fit remains high, whereas misrecognition grows for peripheral customers when realized fit falls short of expectations. This tradeoff explains how symbolic cues amplify outcomes for some customers but undermine them for others.

6. Conclusion

Firms wish to embrace one-to-one marketing, but practical considerations around implementation costs and customer privacy often push them toward segment-level targeting. Understanding when and for whom segment-level personalization is effective is an important question for both theory and practice. With advances in machine learning, customer embeddings often form the bases of modern segmentation frameworks. These representations offer a natural way to examine the heterogeneity in response across segment members. While extant work has incorporated embeddings in their analysis e.g., as control variables in response models, most approaches lack a foundation grounded in behavioral theory to explain why consumers may respond differently to segment-level personalization. Building on prototype theory (Rosch 1975), we propose customer prototypicality as a construct that captures the representativeness of a customer within their assigned segment. We use embeddings to operationalize this construct and test whether it moderates the effectiveness of segment-level personalization.

Drawing on a large-scale field experiment that manipulated both message- and content-level personalization, we find that effectiveness of such marketing communications depends on the interplay between customer prototypicality and the way personalization is implemented. When algorithmic recommendations are delivered with symbolic cues (i.e., “for you"), responses increase among highly prototypical, high-value customers, but the same personalization loses effectiveness, and can even backfires, as customers become more peripheral to the segment. In sum, a customer's structural position within the segment is a key moderator of personalization effectiveness. A series of robustness checks support the reliability of the results. Our results are broadly consistent with a reference-dependent framework, which explains how segment-level personalization can simultaneously enhance the response of some customers and lower it for others.

For marketing theory, there are several avenues for future research that follow directly from our study. First, our analyses rely on deterministic (hard) segment assignments, which treat segment membership as discrete and certain. As noted earlier, classic theories of categorization emphasize that category membership is often graded. Additionally, they may be overlapping. Extending our framework to probabilistic segmentation would allow customers to belong to multiple segments with varying degrees of membership. In such settings, prototypicality could be defined relative to multiple segment centroids, reflecting customers' positions in cross-cutting or overlapping segments. Drawing on research on cross-cutting identities and category overlap (e.g., Blau and Schwartz 1984, Simmel 1955) our framework would predict that personalization is most effective when firms align personalization cues with the customer's most salient or dominant category position, and that misrecognition is most likely when firms act as if segment membership is unambiguous despite underlying classification uncertainty.

Second, while our empirical evidence comes from a single retail campaign, the forces we identify, i.e., baseline match, recognition, and misrecognition, are likely to operate differently across contexts with distinct personalization objectives and decision horizons. For example, in media or content recommendation settings, personalization often aims to sustain engagement and discovery rather than to prompt immediate purchase, potentially shifting the relative importance of recognition versus functional match. Similarly, in subscription or service contexts characterized by repeated interactions, customers may tolerate occasional misrecognition if long-run fit is high. Consistent with prior work showing that the effectiveness of targeting and personalization depends on context and consumption dynamics (Ansari and Mela 2003, Lambrecht and Tucker 2013), our framework suggests that salient personalization is more beneficial in settings with stable preferences and frequent feedback, but more likely to backfire in environments with heterogeneous or evolving tastes.

Third, while our study captures short-run response to a single exposure of personalization, repeated exposure is likely to shape future expectations, satisfaction, and trust. This perspective connects naturally to the expectation–disconfirmation model of satisfaction (Oliver 1980) and to consumer learning models in which beliefs are updated through experience (Erdem and Keane 1996). Extending our framework with a temporal dimension would predict that repeated accurate personalization raises the expectation benchmark and increases the returns to recognition, whereas repeated misrecognition lowers expectations or induces skepticism toward personalization claims. Studying these dynamics can help explain how short run engagement effects translate into long run customer satisfaction.

For marketing practice, our results emphasize that effective segment-level personalization requires balancing the value of symbolic recognition with the accuracy of algorithms employed for product recommendations. A key takeaway from our study is that firms should sequence rather than stack their personalization efforts: first improve match quality, then add recognition cues when fit becomes reliable. Managers should also recognize that customers with niche preferences are most vulnerable to misrecognition. For these customers, personalization should be delivered with either low salience or by employing advanced models capable of predicting their tastes. Finally, firms can proactively shape expectations by curating the message design. For instance, softer cues (e.g., "recommended items") lower the expectation benchmark and may mitigate disappointment when fit is uncertain.

In conclusion, as segment-level targeting remains the mainstay of personalization, we hope that our research contributes to a better understanding of how and when such personalization can be effective.

Funding and Competing Interests: All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

References

Aaker JL (2010) The human brand: Understanding brand personality. *Journal of Marketing Research* 47(5):961–977.

Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17(6):734–749.

Anderson C (2006) *The Long Tail: Why the Future of Business Is Selling Less of More* (New York: Hyperion), ISBN 9781401302375.

Ansari A, Mela CF (2003) E-customization. *Journal of Marketing Research* 40(2):131–145.

Ascarza E (2018) Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research* 55(1):80–98.

Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1798–1828, URL http://dx.doi.org/10.1109/TPAMI.2013.50.

Bhattacharya CB, Sen S (2003) Consumer–company identification: A framework for understanding consumers' relationships with companies. *Journal of Marketing* 67(2):76–88.

Blau PM, Schwartz JE (1984) *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations* (Orlando, FL: Academic Press).

Bleier A, Eisenbeiss M (2015) The importance of personalization in online advertising: A multi-method investigation of advertising effectiveness. *Journal of Marketing* 79(1):102–121.

Bodapati AV (2008) Recommendation systems with purchase data. *Journal of Marketing Research* 45(1):77–93.

Brehm JW (1966) *A Theory of Psychological Reactance* (New York: Academic Press).

Challapalli V, Sai KV, Singh PP, Prasad R, Maurya A, Singh A (2025) Slm4offer: Personalized marketing offer generation using contrastive learning based fine-tuning. *arXiv preprint arXiv:2508.15471* .

Chen Y, Truong QT, Shen X, Li J, King I (2024) Shopping trajectory representation learning with pre-training for e-commerce customer understanding and recommendation. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*.

CloudZero (2024) Netflix architecture: How much does netflix's aws cost? URL https://www.cloudzero.com/blog/netflix-aws/.

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144(1):114–126.

Erdem T, Keane MP (1996) Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* 15(1):1–20.

Fleder D, Hosanagar K (2009) Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science* 55(5):697–712.

Gärdenfors P (2004) *Conceptual Spaces: The Geometry of Thought* (The MIT Press).

Heath C, Larrick RP, Wu G (1999) Goals as reference points. *Cognitive Psychology* 38(1):79–109.

Jagabathula S, Subramanian L, Venkataraman A (2018) A model-based embedding technique for segmenting customers. *Operations Research* 66(5):1247–1267.

Jin W, Mao H, Li Z, Jiang H, Luo C, Wen H, Han H, Lu H, Wang Z, Li R, et al. (2023) Amazon-m2: A multilingual multi-locale shopping session dataset for recommendation and text generation. *Advances in Neural Information Processing Systems* 36:8006–8026.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.

Kallus N, Udell M (2020) Dynamic assortment personalization in high dimensions. *Operations Research* 68(4):1020–1037.

Kiseleva J, Tuzhilin A, Kamps J, Müller MJI, Bernardi L, Davis C, Kovaek I, Einarsen MS, Hiemstra D (2016) Beyond movie recommendations: Solving the continuous cold start problem in e-commercerecommendations. *ArXiv* abs/1607.07904, URL https://api.semanticscholar.org/CorpusID:7114846.

Lambrecht A, Tucker C (2013) When does retargeting work? *Journal of Marketing Research* 50(5):561–576.

Lee D, Hosanagar K (2019) How do recommender systems affect sales diversity? a cross-category investigation via randomized field experiment. *Information Systems Research* 30(1):239–259.

Lichtenberg JM, Buchholz A, Schwöbel P (2024) Large language models as recommender systems: A study of popularity bias. *arXiv preprint arXiv:2406.01285* .

Loken B, Ward JC (1990) Alternative approaches to understanding the determinants of typicality. *Journal of Consumer Research* 17(2):111–126.

Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research* 41(4):467–478.

McKinsey & Company (2021) The value of getting personalization right—or wrong—is multiplying. URL https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying.

Mervis CB, Rosch E, et al. (1981) Categorization of natural objects. *Annual Review of psychology* 32(1):89–115.

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Science* 31(3):521–543, URL http://dx.doi.org/10.1287/mksc.1120.0713.

Oliver RL (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research* 17(4):460–469.

Pathak B, Garfinkel R, Gopal RD, Venkatesan R, Yin F (2010) Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems* 27(2):159–188.

POLITICO (2024) Biden regulators set to look into unequal online pricing scheme. URL https://www.politico.com/news/2024/07/19/individualized-pricing-ftc-concerns-00167913.

Reed A, Forehand MR, Puntoni S, Warlop L (2012) Identity-based consumer behavior. *International Journal of Research in Marketing* 29(4):310–321.

Ricci F, Rokach L, Shapira B (2015) *Recommender Systems Handbook* (New York: Springer), 2nd edition.

Rosch E (1975) Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104(3):192–233.

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Science* 15(4):321–340.

Sahni NS, Wheeler SC, Chintagunta P (2018) Personalization in email marketing: The role of noninformative advertising content. *Marketing Science* 37(2):236–258.

Simmel G (1955) *Conflict and the Web of Group-Affiliations* (New York: Free Press).

Thaler R (1985) Mental accounting and consumer choice. *Marketing Science* 4(3):199–214.

Tversky A (1977) Features of similarity. *Psychological Review* 84(4):327–352, URL http://dx.doi.org/10.1037/0033-295X.84.4.327.

Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* 106(4):1039–1061.

Wedel M, Kamakura WA (2000) *Market Segmentation: Conceptual and Methodological Foundations* (New York: Springer), 2nd edition.

White K, Habib R, Hardisty DJ (2019) How to shift consumer behaviors to be more sustainable: A literature review and guiding framework. *Journal of Marketing* 83(3):22–49.

Yoganarasimhan H, Barzegary E, Pani A (2023) Design and evaluation of optimal free trials. *Management Science* 69(6):3220–3240.

Electronic Companion Supplement

Appendix A: Overall Randomization Checks

Table EC.1 reports baseline customer characteristics across conditions. Covariates are well balanced across groups, with no statistically significant differences at conventional levels. These results indicate that randomization achieved balance in the full sample, complement the segment-level checks in Table 2, and support the internal validity of the experimental design.

Table EC.1 Pre-treatment Customer Characteristics

Variable	G1	G2	G3	G4
Age	52.32	52.36	52.46	52.26
Gender (1 if male)	0.67	0.67	0.66	0.67
Tenure (months)	14.39	14.44	14.17	14.34
Recency (months)	5.64	5.67	5.55	5.60
Frequency	1.40	1.38	1.39	1.39
Monetary ($)	260.72	252.42	258.33	257.86
Num. Customers	63,961	10,928	10,879	10,948

Appendix B: Robustness Checks: Models with Controls

We estimated the models with additional covariates, including mean-centered age, gender, tenure, and RFM measures, using the specifications reported in Table 5. The results, presented in Table EC.2, remain substantively unchanged, indicating that our results are robust to the inclusion of these controls.

Table EC.2 Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting (with Controls)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	-0.61^{*}	-0.37	-0.25
	(0.37)	(0.51)	(0.41)
Segment: Top	0.74	0.92	0.70
	(0.56)	(0.58)	(0.62)
Segment: Bottom	-0.90^{**}	-0.94^{**}	-0.61
	(0.38)	(0.38)	(0.42)
Prototypicality	0.48	0.62	1.34
	(0.88)	(0.88)	(0.97)
Treatment × Segment			
Treatment × Top	1.26^{**}	-0.13	1.40^{**}
	(0.60)	(0.82)	(0.66)
Treatment × Bottom	0.28	0.30	-0.02
	(0.42)	(0.57)	(0.46)
Treatment × Prototypicality	1.71^{*}	0.78	0.94
	(0.95)	(1.31)	(1.04)
Segment × Prototypicality			
Top × Prototypicality	0.43	0.17	0.61
	(1.36)	(1.37)	(1.50)
Bottom × Prototypicality	2.00^{**}	1.96^{**}	0.87
	(1.00)	(1.00)	(1.10)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-3.30^{**}	0.27	-3.62^{**}
	(1.48)	(2.02)	(1.60)
Treatment × Bottom × Prototypicality	-1.19	-1.11	-0.10
	(1.09)	(1.48)	(1.18)
Constant	-3.28^{***}	-3.27^{***}	-3.63^{***}
	(0.36)	(0.41)	(0.40)
Controls	Yes	Yes	Yes
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Appendix C: Robustness Checks: Models with Controls

We estimated the models with additional covariates, including mean-centered age, gender, tenure, and RFM measures, using the specifications reported in Table 6. The results, presented in Table EC.3, remain substantively unchanged, indicating that our results are robust to the inclusion of these controls.

Table EC.3 Effects of Symbolic Cues and Prototypicality in Popularity-Based Targeting (with Controls)

	G2 vs. G4		
Variable	(1) All	(2) Mainstream	(3) Long-Tail
Treatment	-0.66	0.05	-1.05
	(0.51)	(0.83)	(0.65)
Segment: Top	0.79	0.62	0.74
	(0.58)	(2.05)	(0.64)
Segment: Bottom	-0.88^{**}	-0.31	-0.98^{**}
	(0.38)	(0.65)	(0.49)
Prototypicality	0.52	0.64	0.31
	(0.89)	(1.64)	(1.06)
Treatment × Segment			
Treatment × Top	1.65^{**}	-2.44	2.17^{**}
	(0.79)	(2.87)	(0.90)
Treatment × Bottom	0.28	0.06	0.21
	(0.57)	(0.91)	(0.76)
Treatment × Prototypicality	1.77	0.06	2.62^{*}
	(1.26)	(2.28)	(1.52)
Segment × Prototypicality			
Top × Prototypicality	0.22	-4.20	0.58
	(1.39)	(7.96)	(1.50)
Bottom × Prototypicality	1.96^{*}	1.57	1.92
	(1.00)	(1.80)	(1.23)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-3.27^{*}	9.13	-4.60^{**}
	(1.97)	(9.14)	(2.16)
Treatment × Bottom × Prototypicality	-1.37	-1.22	-1.11
	(1.45)	(2.55)	(1.80)
Constant	-3.44^{***}	-5.08^{***}	-2.66^{***}
	(0.41)	(0.72)	(0.51)
Controls	Yes	Yes	Yes
Num. Obs.	21,876	13,424	8,452

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment is coded as 1 if the customer belongs to G2 and 0 if in the control group (G4). Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Appendix D: Robustness Checks: Alternative Distance Mappings

This appendix reports robustness checks that re-express the same within-segment distance to the segment centroid using alternative, standard monotone distance-to-similarity mappings. The goal is to verify that the moderation results are not an artifact of any particular functional form used to translate distance into our representativeness measure.

Let $\tilde{p}_i$ denote customer i's normalized distance to the segment centroid (scaled to $[0,1]$). We construct transformed indices $f(\tilde{p}_i)$ using four commonly used monotone mappings:

1. Exponential decay: $f_{\text{exp}}(\tilde{p}_i) = 1 - \exp(-\lambda\tilde{p}_i)$
2. Gaussian decay: $f_{\text{gauss}}(\tilde{p}_i) = 1 - \exp(-\lambda\tilde{p}_i^2)$
3. Logistic saturation: $f_{\text{logit}}(\tilde{p}_i) = 1 - \left[1 + \exp\{\lambda(\tilde{p}_i - 1)\}\right]^{-1}$
4. Inverse distance: $f_{\text{inv}}(\tilde{p}_i) = 1 - (1 + \tilde{p}_i)^{-1}$

Each mapping is strictly decreasing in $\tilde{p}_i$ (for $\lambda > 0$), so $f(\tilde{p}_i)$ decreases as a customer moves closer to the centroid. We interpret lower values of $f(\tilde{p}_i)$ as greater prototypicality (closer to the centroid) and higher values as greater peripherality. Importantly, because the transformations are monotone, they preserve the within-segment rank ordering implied by the underlying distance measure.

For the exponential, Gaussian, and logistic mappings, we select the sensitivity parameter λ by grid search to minimize the Akaike Information Criterion (AIC) for the corresponding regression specification. This procedure yields $\lambda = 0.1$ for $f_{\text{exp}}(\cdot)$, $f_{\text{gauss}}(\cdot)$, and $f_{\text{logit}}(\cdot)$.

Tables EC.4–EC.7 report the estimated moderation models using each alternative index. Across all mappings, the qualitative moderation pattern is unchanged: the direction and statistical significance of the key interaction terms remain consistent with the main results. These findings indicate that our conclusions are robust to how within-segment distance is mapped into the index and are not driven by a particular functional form.

Table EC.4 Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting (Exponential Decay)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	-1.35	-0.39	-0.97
	(0.94)	(1.26)	(0.98)
Segment: Top	4.26^{***}	4.26^{***}	3.92^{***}
	(1.02)	(1.02)	(1.09)
Segment: Bottom	-0.84	-0.84	-0.34
	(0.98)	(0.98)	(1.03)
Prototypicality	13.19^{*}	13.19^{*}	16.01^{**}
	(7.42)	(7.42)	(7.71)
Treatment × Segment			
Treatment × Top	1.96^{*}	-0.35	2.31^{**}
	(1.11)	(1.50)	(1.17)
Treatment × Bottom	0.78	0.49	0.29
	(1.06)	(1.42)	(1.10)
Treatment × Prototypicality	11.55	2.82	8.73
	(7.98)	(10.70)	(8.25)
Segment × Prototypicality			
Top × Prototypicality	-9.19	-9.19	-2.67
	(12.76)	(12.76)	(13.65)
Bottom × Prototypicality	6.86	6.86	1.44
	(8.57)	(8.57)	(8.98)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-23.57^{*}	6.52	-30.08^{**}
	(13.78)	(18.68)	(14.60)
Treatment × Bottom × Prototypicality	-7.60	-5.42	-2.18
	(9.25)	(12.41)	(9.63)
Constant	-5.03^{***}	-5.03^{***}	-5.41^{***}
	(0.87)	(0.87)	(0.91)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Table EC.5 Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting (Gaussian Decay)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	−0.63	−0.19	−0.44
	(0.48)	(0.64)	(0.49)
Segment: Top	3.52^{***}	3.52^{***}	3.28^{***}
	(0.55)	(0.55)	(0.57)
Segment: Bottom	−0.62	−0.62	−0.41
	(0.49)	(0.49)	(0.52)
Prototypicality	5.28^{*}	5.28^{*}	6.21^{**}
	(2.94)	(2.94)	(3.02)
Treatment × Segment			
Treatment × Top	1.04^{*}	−0.24	1.28^{**}
	(0.59)	(0.79)	(0.61)
Treatment × Bottom	0.30	0.21	0.09
	(0.53)	(0.71)	(0.55)
Treatment × Prototypicality	4.36	0.93	3.42
	(3.15)	(4.22)	(3.22)
Segment × Prototypicality			
Top × Prototypicality	0.67	0.67	5.44
	(10.34)	(10.34)	(10.72)
Bottom × Prototypicality	4.12	4.12	1.60
	(3.48)	(3.48)	(3.64)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-18.44^{*}	4.77	-23.21^{**}
	(11.17)	(14.90)	(11.52)
Treatment × Bottom × Prototypicality	−2.91	−2.53	−0.39
	(3.75)	(5.04)	(3.89)
Constant	-4.25^{***}	-4.25^{***}	-4.45^{***}
	(0.44)	(0.44)	(0.46)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Table EC.6 Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting (Logistic Saturation)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	-20.33	-4.93	-15.40
	(14.14)	(18.95)	(14.60)
Segment: Top	18.69	18.69	6.33
	(23.34)	(23.34)	(24.93)
Segment: Bottom	-13.11	-13.11	-3.42
	(15.23)	(15.23)	(15.93)
Prototypicality	46.25^{*}	46.25^{*}	55.87^{**}
	(25.99)	(25.99)	(26.95)
Treatment × Segment			
Treatment × Top	42.94^{*}	-12.36	55.29^{**}
	(25.20)	(34.15)	(26.68)
Treatment × Bottom	13.27	9.68	3.59
	(16.42)	(22.03)	(17.07)
Treatment × Prototypicality	40.17	9.63	30.54
	(27.93)	(37.44)	(28.83)
Segment × Prototypicality			
Top × Prototypicality	-30.59	-30.59	-5.33
	(47.31)	(47.31)	(50.53)
Bottom × Prototypicality	25.83	25.83	6.42
	(30.16)	(30.16)	(31.54)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-86.49^{*}	25.26	-111.74^{**}
	(51.09)	(69.23)	(54.08)
Treatment × Bottom × Prototypicality	-26.43	-19.41	-7.02
	(32.52)	(43.64)	(33.81)
Constant	-26.89^{**}	-26.89^{**}	-31.82^{**}
	(13.15)	(13.15)	(13.65)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Table EC.7 Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting (Inverse Distance)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	-3.09	-0.92	-2.17
	(2.03)	(2.73)	(2.14)
Segment: Top	5.96^{***}	5.96^{***}	5.84^{***}
	(2.01)	(2.01)	(2.15)
Segment: Bottom	-0.91	-0.91	0.15
	(2.07)	(2.07)	(2.20)
Prototypicality	5.85^{*}	5.85^{*}	7.41^{**}
	(3.39)	(3.39)	(3.59)
Treatment × Segment			
Treatment × Top	3.81^{*}	-0.12	3.94^{*}
	(2.18)	(2.94)	(2.32)
Treatment × Bottom	1.97	1.06	0.91
	(2.26)	(3.02)	(2.38)
Treatment × Prototypicality	5.59	1.56	4.02
	(3.68)	(4.93)	(3.86)
Segment × Prototypicality			
Top × Prototypicality	-5.29	-5.29	-4.52
	(4.02)	(4.02)	(4.33)
Bottom × Prototypicality	1.54	1.54	-0.58
	(3.81)	(3.81)	(4.04)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-7.74^{*}	0.77	-8.52^{*}
	(4.36)	(5.90)	(4.65)
Treatment × Bottom × Prototypicality	-3.72	-2.13	-1.59
	(4.14)	(5.55)	(4.35)
Constant	-6.71^{***}	-6.71^{***}	-7.63^{***}
	(1.87)	(1.87)	(1.99)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Appendix E: Robustness Check: Global Prototypicality

We replicated the analyses using a global measure of prototypicality, defined as each customer's distance to the overall population centroid, applying the specifications reported in Table 5. The results, reported in Table EC.8, are not statistically significant at conventional levels, indicating that heterogeneity is driven by segment-relative prototypicality rather than global centrality.

Table EC.8 Effects of Symbolic Cues and Global Prototypicality in Algorithmic Targeting

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	0.05	−0.02	0.07
	(0.63)	(0.85)	(0.66)
Segment: Top	3.17	3.17	2.89
	(3.44)	(3.44)	(3.47)
Segment: Bottom	2.91^{***}	2.91^{***}	2.92^{***}
	(0.62)	(0.62)	(0.65)
Prototypicality	7.59^{***}	7.59^{***}	7.51^{***}
	(0.94)	(0.94)	(0.98)
Treatment × Segment			
Treatment × Top	−1.26	−0.28	−0.98
	(3.70)	(4.88)	(3.73)
Treatment × Bottom	−0.36	0.01	−0.37
	(0.67)	(0.90)	(0.70)
Treatment × Prototypicality	−0.03	−0.08	0.06
	(1.01)	(1.36)	(1.06)
Segment × Prototypicality			
Top × Prototypicality	−3.07	−3.07	−2.97
	(3.94)	(3.94)	(3.95)
Bottom × Prototypicality	-2.77^{**}	-2.77^{**}	-3.39^{**}
	(1.28)	(1.28)	(1.36)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	1.39	0.10	1.28
	(4.24)	(5.58)	(4.25)
Treatment × Bottom × Prototypicality	0.70	−0.62	1.32
	(1.39)	(1.87)	(1.46)
Constant	-7.73^{***}	-7.73^{***}	-7.75^{***}
	(0.59)	(0.59)	(0.62)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality refers to global prototypicality scaled 0–1; lower values are closer to the overall customer centroid (more prototypical), higher are more peripheral.

Appendix F: Robustness Check: Demographics

To assess whether demographic variables could account for the heterogeneity attributed to prototypicality, we estimated a series of logistic regressions in which prototypicality was replaced by age, gender, or tenure as moderators of treatment effects. The results, reported in Tables EC.9–EC.11, provide no consistent evidence that demographics explain variation in response to personalization treatments. Overall, these analyses support the view that prototypicality, rather than demographics, offers the more fundamental lens for understanding heterogeneity in personalization.

Table EC.9 Effects of Symbolic Cues in Algorithmic Targeting (Age)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	0.03	−0.06	0.09
	(0.13)	(0.17)	(0.13)
Segment: Top	2.84^{***}	2.84^{***}	2.70^{***}
	(0.21)	(0.21)	(0.22)
Segment: Bottom	-0.43^{***}	-0.43^{***}	-0.52^{***}
	(0.14)	(0.14)	(0.15)
Age	0.02^{**}	0.02^{**}	0.02^{**}
	(0.01)	(0.01)	(0.01)
Treatment × Segment			
Treatment × Top	0.02	−0.14	0.16
	(0.22)	(0.30)	(0.23)
Treatment × Bottom	−0.20	−0.10	−0.10
	(0.16)	(0.21)	(0.16)
Treatment × Age	−0.01	0.00	−0.01
	(0.01)	(0.01)	(0.01)
Segment × Age			
Top × Age	-0.05^{***}	-0.05^{***}	-0.03^{*}
	(0.02)	(0.02)	(0.02)
Bottom × Age	−0.01	−0.01	−0.01
	(0.01)	(0.01)	(0.01)
Treatment × Segment × Age			
Treatment × Top × Age	0.02	0.01	0.00
	(0.02)	(0.02)	(0.02)
Treatment × Bottom × Age	0.02	0.00	0.01
	(0.01)	(0.02)	(0.01)
Constant	-3.50^{***}	-3.50^{***}	-3.55^{***}
	(0.12)	(0.12)	(0.12)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Age is mean-centered.

Table EC.10 Effects of Symbolic Cues in Algorithmic Targeting (Gender)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	-0.16	-0.27	0.11
	(0.19)	(0.25)	(0.20)
Segment: Top	2.55^{***}	2.55^{***}	2.46^{***}
	(0.31)	(0.31)	(0.33)
Segment: Bottom	-1.01^{***}	-1.01^{***}	-0.62^{***}
	(0.23)	(0.23)	(0.23)
Gender	-0.76^{***}	-0.76^{***}	-0.41
	(0.24)	(0.24)	(0.25)
Treatment × Segment			
Treatment × Top	0.13	-0.09	0.22
	(0.33)	(0.45)	(0.35)
Treatment × Bottom	0.21	0.39	-0.18
	(0.24)	(0.32)	(0.25)
Treatment × Gender	0.32	0.36	-0.04
	(0.26)	(0.34)	(0.27)
Segment × Gender			
Top × Gender	0.64	0.64	0.45
	(0.41)	(0.41)	(0.42)
Bottom × Gender	0.96^{***}	0.96^{***}	0.19
	(0.29)	(0.29)	(0.30)
Treatment × Segment × Gender			
Treatment × Top × Gender	-0.25	-0.19	-0.05
	(0.44)	(0.59)	(0.45)
Treatment × Bottom × Gender	-0.65^{**}	-0.77^{*}	0.12
	(0.32)	(0.42)	(0.33)
Constant	-3.05^{***}	-3.05^{***}	-3.31^{***}
	(0.17)	(0.17)	(0.19)
Num. Obs.	$74,909$	$21,827$	$74,840$

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Gender equals 1 if the customer is male and 0 otherwise.

Table EC.11 Effects of Symbolic Cues in Algorithmic Targeting (Tenure)

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	0.04	0.00	0.03
	(0.15)	(0.19)	(0.14)
Segment: Top	2.74^{***}	2.74^{***}	2.44^{***}
	(0.21)	(0.21)	(0.22)
Segment: Bottom	-0.58^{***}	-0.58^{***}	-0.76^{***}
	(0.17)	(0.17)	(0.18)
Tenure	0.02^{***}	0.02^{***}	0.02^{***}
	(0.01)	(0.01)	(0.01)
Treatment × Segment			
Treatment × Top	−0.04	−0.30	0.26
	(0.23)	(0.30)	(0.23)
Treatment × Bottom	−0.13	−0.18	0.05
	(0.18)	(0.25)	(0.19)
Treatment × Tenure	0.00	0.01	−0.01
	(0.01)	(0.01)	(0.01)
Segment × Tenure			
Top × Tenure	-0.02^{**}	-0.02^{**}	-0.02^{*}
	(0.01)	(0.01)	(0.01)
Bottom × Tenure	-0.01^{*}	-0.01^{*}	-0.02^{***}
	(0.01)	(0.01)	(0.01)
Treatment × Segment × Tenure			
Treatment × Top × Tenure	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)
Treatment × Bottom × Tenure	0.00	−0.01	0.01
	(0.01)	(0.01)	(0.01)
Constant	-3.29^{***}	-3.29^{***}	-3.29^{***}
	(0.13)	(0.13)	(0.13)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Tenure is mean-centered.

Appendix G: Effects on Customer Spending

Table EC.12 reports results from models estimating the effects on customer spending, conditional on purchase. The dependent variable is log spending, allowing us to interpret coefficients in proportional terms. Across all models, we find no systematic effect of either treatment or its interaction with prototypicality on customer spending levels. The coefficients are small in magnitude and statistically indistinguishable from zero. This pattern aligns with the expectation that personalization treatments influence the likelihood of purchase (the extensive margin) but not the amount spent once a purchase occurs (the intensive margin).

Table EC.12 Effects on Customer Spending

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	-0.51^{**}	-0.48	-0.02
	(0.21)	(0.38)	(0.34)
Segment: Top	-0.39	-0.39	0.11
	(0.26)	(0.27)	(0.44)
Segment: Bottom	-1.02^{***}	-1.02^{***}	-0.52
	(0.22)	(0.23)	(0.36)
Prototypicality	-0.92^{*}	-0.92^{*}	0.25
	(0.51)	(0.52)	(0.86)
Treatment × Segment			
Treatment × Top	0.35	0.49	-0.15
	(0.29)	(0.52)	(0.46)
Treatment × Bottom	0.67^{***}	0.50	0.17
	(0.25)	(0.43)	(0.38)
Treatment × Prototypicality	1.00^{*}	1.17	-0.17
	(0.57)	(1.01)	(0.90)
Segment × Prototypicality			
Top × Prototypicality	1.30^{**}	1.30^{*}	-0.06
	(0.66)	(0.67)	(1.07)
Bottom × Prototypicality	1.95^{***}	1.95^{***}	0.71
	(0.60)	(0.61)	(0.95)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	-0.77	-1.36	0.59
	(0.73)	(1.27)	(1.11)
Treatment × Bottom × Prototypicality	-1.58^{**}	-1.24	-0.33
	(0.67)	(1.13)	(0.99)
Constant	12.36^{***}	12.36^{***}	11.88^{***}
	(0.19)	(0.19)	(0.33)
Num. Obs.	1,906	552	1,870

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.

Appendix H: Post-Campaign Outcomes

Table EC.13 reports results for the four-week post-campaign window. These estimates show that personalization effects dissipate after the campaign, suggesting that the effects are short-lived during the campaign period.

Table EC.13 Effects of Symbolic Cues and Prototypicality in Algorithmic Targeting: Post-Campaign Period

Variable	G1 vs. G4 (1)	G3 vs. G4 (2)	G1 vs. G3 (3)
Treatment	0.68^{*}	0.79	-0.10
	(0.38)	(0.50)	(0.38)
Segment: Top	4.11^{***}	4.11^{***}	2.73^{***}
	(0.54)	(0.54)	(0.57)
Segment: Bottom	-0.14	-0.14	-0.64^{*}
	(0.40)	(0.40)	(0.39)
Prototypicality	3.33^{***}	3.33^{***}	1.42
	(0.84)	(0.84)	(0.90)
Treatment × Segment			
Treatment × Top	-0.96	-1.38^{*}	0.43
	(0.58)	(0.79)	(0.61)
Treatment × Bottom	-0.41	-0.50	0.09
	(0.43)	(0.56)	(0.42)
Treatment × Prototypicality	-1.52^{*}	-1.91	0.39
	(0.91)	(1.23)	(0.97)
Segment × Prototypicality			
Top × Prototypicality	-3.77^{***}	-3.77^{***}	-0.84
	(1.28)	(1.28)	(1.37)
Bottom × Prototypicality	0.05	0.05	1.26
	(0.96)	(0.96)	(1.01)
Treatment × Segment × Prototypicality			
Treatment × Top × Prototypicality	2.09	2.93	-0.84
	(1.39)	(1.88)	(1.48)
Treatment × Bottom × Prototypicality	0.74	1.21	-0.47
	(1.04)	(1.39)	(1.09)
Constant	-4.78^{***}	-4.78^{***}	-3.99^{***}
	(0.35)	(0.35)	(0.35)
Num. Obs.	74,909	21,827	74,840

Note: Standard errors are shown in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Treatment equals 1 for customers in the focal treatment group and 0 for customers in the corresponding control group: Column (1) G1 vs. G4; Column (2) G3 vs. G4; Column (3) G1 vs. G3. Prototypicality scaled 0–1; lower values are closer to the segment centroid (more prototypical), higher are more peripheral.