Personalization and Bundling: How Message Fit and Product Framing Influence Spending

Maia Kennedy
University of California, Berkeley MIDS

Indri Adisoemarta
University of California, Berkeley MIDS

Jane Li
University of California, Berkeley MIDS

Jordan Andersen
University of California, Berkeley MIDS

13 August 2025

1 Introduction

In digital commerce, personalization and product bundling are widely used strategies aimed at increasing consumer engagement and spending. This study investigates whether identity-based personalization enhances the effectiveness of product bundling in driving purchase behavior. Using a 3x2 factorial design, we randomly assigned 272 US-based participants into one of 6 conditions varying in personalization and bundling. We hypothesized that both personalization and bundling would independently increase total spending and that their combination would also produce a positive effect on spending. Contrary to expectations, we found no statistically significant differences in total purchase amounts across any of the treatment conditions. The findings contribute to the need for more nuanced understanding of the contextual and psychological factors that shape the effectiveness of digital marketing strategies.

1.1 Research question

How does identity-based personalization influence the effectiveness of product bundling in driving consumer spending in an online retail context? Specifically, we investigate whether the combination of personalization and bundling strategies increase total purchase amounts or influence individuals to make any purchase compared to non-personalized, non-bundled options.

1.2 Justification

In today's competitive digital commerce landscape, personalization and product bundling are two widely adopted strategies to enhance customer engagement and increase sales. While both tactics are effective independently, their combined impact on consumer behavior remains underexplored. Prior research has shown that personalization can increase perceived relevance and trust in recommendations (XinYun & Chun, 2024), while bundling strategies that align with consumer goals can increase overall purchase value (Rao et al., 2018). However, few studies have examined whether personalization enhances the perceived value and performance of product bundles. This research addresses that gap by testing whether identity-driven personalization amplifies the effectiveness of bundling in influencing real purchase behavior. The findings have important implications for marketing strategy, product recommendation systems, and the design of personalized user experiences in e-commerce platforms.

1.3 Hypothesis

We test the following hypotheses:

- H1: Identity-based personalization (accurate matching to consumer profiles) increases total purchase amounts, regardless of bundling condition
- H2: Product bundling increases total purchase amounts, regardless of personalization condition
- H3 (Interaction Effect): The combination of accurate personalization and product bundling leads to higher purchase totals than either strategy alone, indicating a synergistic effect

2 Experiment

2.1 Comparison of Potential outcomes

The outcome measurements for this experiment will compare total simulated purchase amounts and whether a subject made a purchase (a binary outcome). To simulate real-world trade-offs in our outcome variables, each participant was given a fixed virtual budget of $300 to spend or save during the shopping task. They were instructed that this was their total budget, and any combination of purchases and savings had to fit within it. They could choose to purchase any of the items appearing in the simulated environment, or, alternatively, they could choose to not purchase any of the items and save the $300. The only variation across subjects is whether three of these products are bundled and the type of personalization messaging applied.

We are comparing these outcomes across six randomly assigned groups, as indicated by the ROXO diagram in Table 1. These groups represent various combinations of the treatment conditions (Bundled, Accurate Personalization, Inaccurate Personalization), and the outcomes will be compared to determine how these treatments impact the likelihood and total amount of purchase. The groups will be as follows:

- Group 1 (R):No bundle, no personalization
- Group 2 (R, X1): Bundled, no personalization
- Group 3 (R, X2): No bundle, Accurate Personalization
- Group 4 (R, X3): No bundle, Inaccurate Personalization
- Group 5 (R, X1,X2): Bundled, Accurate Personalization
- Group 6 (R, X1, X3): Bundled, Inaccurate Personalization

The key comparison is between these groups to assess how different combinations of bundling and personalization treatments influence purchasing behavior.

Table 1. ROXO Diagram

Randomized Groups	Treatment	Outcome Measures
R		Total Purchase Amount $, Any purchase (1=yes, 0=no)
R	X1	Total Purchase Amount $, Any purchase (1=yes, 0=no)
R	X2	Total Purchase Amount $, Any purchase (1=yes, 0=no)
R	X3	Total Purchase Amount $, Any purchase (1=yes, 0=no)
R	X1, X2	Total Purchase Amount $, Any purchase (1=yes, 0=no)
R	X1, X3	Total Purchase Amount $, Any purchase (1=yes, 0=no)

2.2 Treatment

This study employed a 3×2 factorial design, resulting in six total experimental conditions: three levels of personalization (accurate match, inaccurate match, and no personalization) crossed with two levels of bundling (bundled vs. non-bundled). Five of these conditions constituted treatment groups, while the sixth, no personalization and no bundling, served as a control. Participants were randomly assigned to one of the six conditions using Qualtrics' built-in randomization logic, with approximately equal allocation across groups. There were no technical deviations from the intended assignment. All participants received the correct product display and personalization message according to their assigned condition.

Product Selection

The shopping task featured six consumer technology products, selected for their broad appeal, gender neutrality, and price variability (ranging from \$24 to \$129). These products included:

- Apple AirPods 4 (\$129.00)
- Anker Portable Charger (\$54.99)
- Roku Streaming Stick 4K (\$46.99)
- Tile by Life360 Pro (\$34.99)
- HD Mini Projector (\$59.99)
- JBL Clip 5 Speaker (\$79.95)

For all participants, product listings included a brief description and use cases. In personalized conditions, messaging was framed to reflect the participant's assigned identity profile (e.g., family-oriented, frequent traveler). In bundled conditions, three of the six products, AirPods, Portable Charger, and Streaming Stick, were grouped together with a promotional "bundle" label and accompanying message, while the other three were shown individually. In the non-bundled condition, all six products were presented as individual options with no grouping. For samples of these product listings, refer to appendix A.

2.3 Experimental Conditions

The six experimental groups were structured as follows:

Table 2. Bundling X Personalization

	Bundled	Non-Bundled
Personalized Accurate	Bundle (3) + 3 individual items	6 items
Personalized Inaccurate	Bundle (3) + 3 individual items	6 items
Non-Personalized	Bundle (3) + 3 individual items	6 items

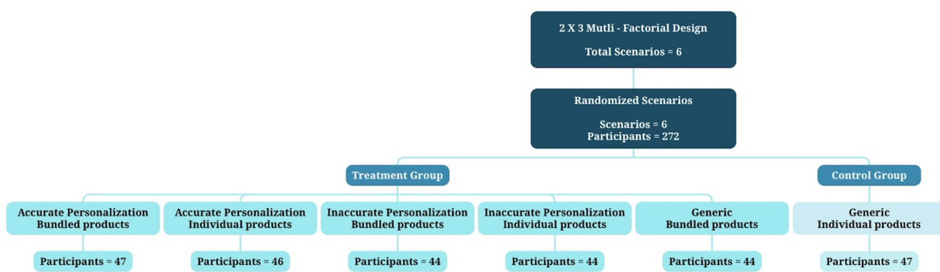


Figure 1: Observation Tracking Flow Diagram

Validation of Treatment Perception & Diagnostic Checks

To assess whether participants perceived the personalization manipulation as intended, we included a post-task perception check asking how personalized the product messaging felt. This served as a validation step to evaluate alignment between perceived and assigned personalization conditions.

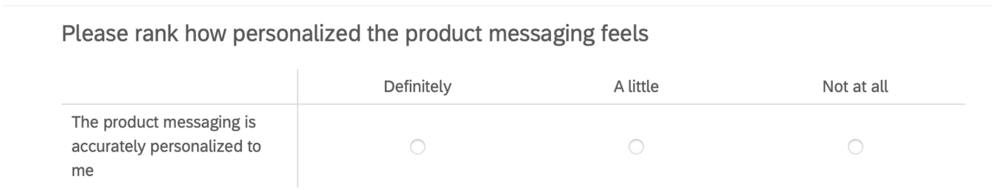


Figure 2: Treatment perception check

Additionally, we included a post-task perception check to evaluate whether participants had read and internalized the product messaging. This took the form of an optional question asking participants to confirm they read product messaging.

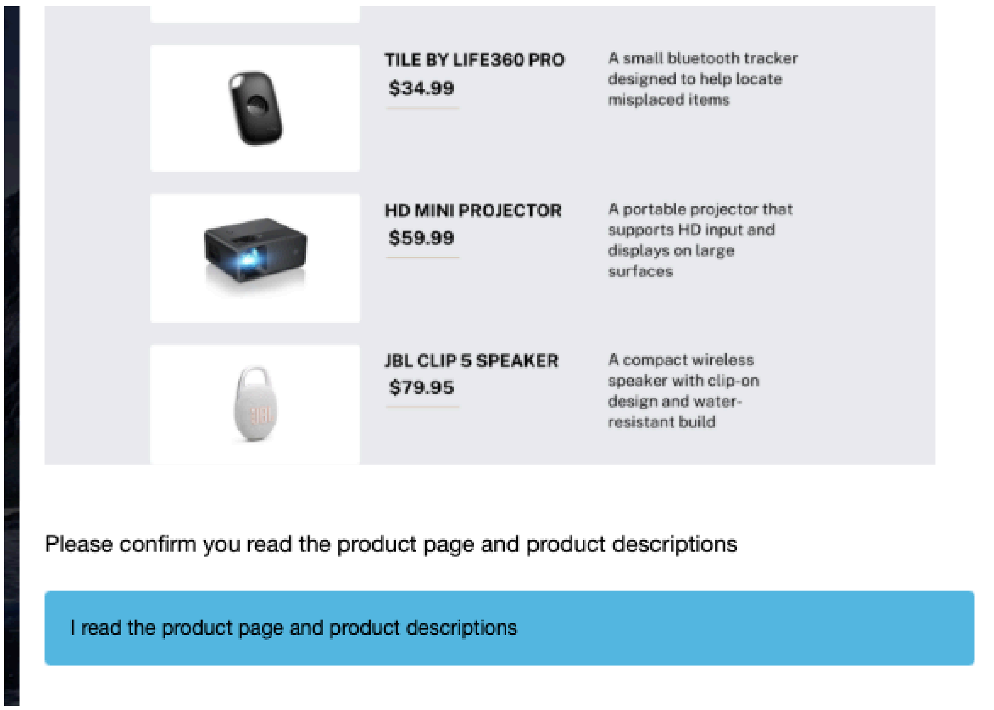


Figure 3: Compliance confirmation check

Pre-Treatment

Before exposure to any treatment, all participants completed a brief pre-task questionnaire designed to assess their lifestyle and consumer profile. The survey included a combination of text- and image-based questions, intended to simulate the kind of multidimensional data that modern e-commerce platforms often use to personalize recommendations. Participant responses were used to assign each individual to one of five predefined identity categories:

- Travel
- Family
- Fitness
- Work & Study
- Leisure

These categories were determined using backend logic, which scored each participant's responses. In cases where participants matched equally across multiple profiles, one of the top-scoring categories was selected at random.

Sample Profiling Question

"When you have free time on a weekend, how are you most likely to spend it?"

- Exploring new places or taking short trips
- Exercising, going to the gym, or outdoor activities
- Spending time at home with family
- Catching up on work, studying, or learning something new
- Watching shows, playing games, or relaxing

This profiling process established the foundation for delivering identity-based personalization treatments.

2.3.1 Personalization Treatment

Following profile assignment, participants were randomly assigned to one of three personalization conditions:

- Accurate Personalization: Messaging matched the participant's assigned profile (e.g., a Work & Study profile received messaging emphasizing productivity use cases).
- Inaccurate Personalization: Messaging contradicted the participant's assigned profile (e.g., a Family profile received Leisure-themed messaging).
- No Personalization (Control): Messaging was generic and did not reference any identity profile.

All personalized messages were applied at the product messaging level, with blurbs tailored to frame the use cases of the products in ways that aligned (or did not align) with the participant's profile. To validate whether participants perceived the personalization as intended, a post-task perception check asked them to rate how personalized the messaging felt.

2.3.2 Bundling Treatment

Participants were also assigned to one of two bundling conditions:

- Bundled: Three of the six products were grouped together and presented with a promotional bundle label and messaging. The remaining three items were shown individually.
- Non-Bundled: All six products were presented individually, with no grouping or bundling cues.

Bundling was designed to test whether framing certain items as a package deal would influence purchase behavior relative to the same products shown separately.

2.3.3 Participants and Recruitment

The unit of analysis in this study is individual participants. A total of 272 valid and complete responses were collected, with a 0% dropout rate. All participants were adults, primarily located in the United States, and were recruited using a mix of personal networks and paid sampling. Specifically:

- 70 participants were recruited through friends, family, and MIDS classmates, and
- 202 participants were recruited via the online survey platform PureSpectrum.

2.4 Power Analysis

We test the power of our experimental design using a randomized simulation where participants are randomly assigned to one of the 6 groups, with the no-bundle, no-personalization group representing the control and the accurate personalization, bundle group being the main outcome of interest. Our sharp null hypothesis states that the outcome variable (purchase amount) will remain unchanged regardless of group assignment. This simulation requires pre-determined assumptions of the outcome variable. We define a base outcome under the sharp null hypothesis, assuming a purchase amount of $30. We then add random noise drawn from a normal distribution with mean 0 and standard deviation 5. The baseline amount of $30 reflects the original pricing of low-budget tech gadgets designed to appeal to a broad range of ages and genders. For those in the treatment group, we assumed an increase in purchase amount equivalent to the product of the base purchase amount and the estimated effect size increased by 1:

outcome = purchase amount x (1 + effect size)

We then compare the effects of three hypothetical increases in purchase amounts (5%, 10%, and 15%) when customers are exposed to accurate personalization and bundling, relative to a baseline with no personalization and no bundling. The test range of 5-15% is based on prior research indicating that personalization can increase revenue by 10-15% (Arora et al., 2021). Evidence for the impact of bundling is mixed, ranging from modest gains (1-2%) to negative effects.

The regression model for the randomized data shows significant results for the accurate personalization, bundle group, with purchase amount increasing by $4. Results of the power analysis indicate that an effect size of 15% could be detected with 250 participants with the study powered at 89.7%, which is sufficiently above the 80% threshold required for a well-designed experiment. Smaller effect sizes of 5% will be more difficult to detect, requiring 2,000 participants to reach 77.8% power. However, given that prior literature tends to report larger effects in the range of 10–15% and considering budget limitations, we determined that a sample size of 250 participants would be adequate for detecting meaningful effects within this expected range, while maintaining a high level of statistical power.

Table 3: Results: Simulated Randomization Inference Regression

	Dependent variable:
	Purchase Amount
bundle	0.323
	(0.398)
personalizationaccurate	0.825**
	(0.403)
personalizationinaccurate	0.689*
	(0.390)
bundle:personalizationaccurate	3.665***
	(0.567)
bundle:personalizationinaccurate	−0.088
	(0.547)
Constant	29.397***
	(0.290)
Observations	2,000
R^2	0.087
Adjusted R^2	0.085
Residual Std. Error	5.050 (df = 1994)
F Statistic	37.956*** (df = 5; 1994)
Note:	*p<0.1; **p<0.05; ***p<0.01

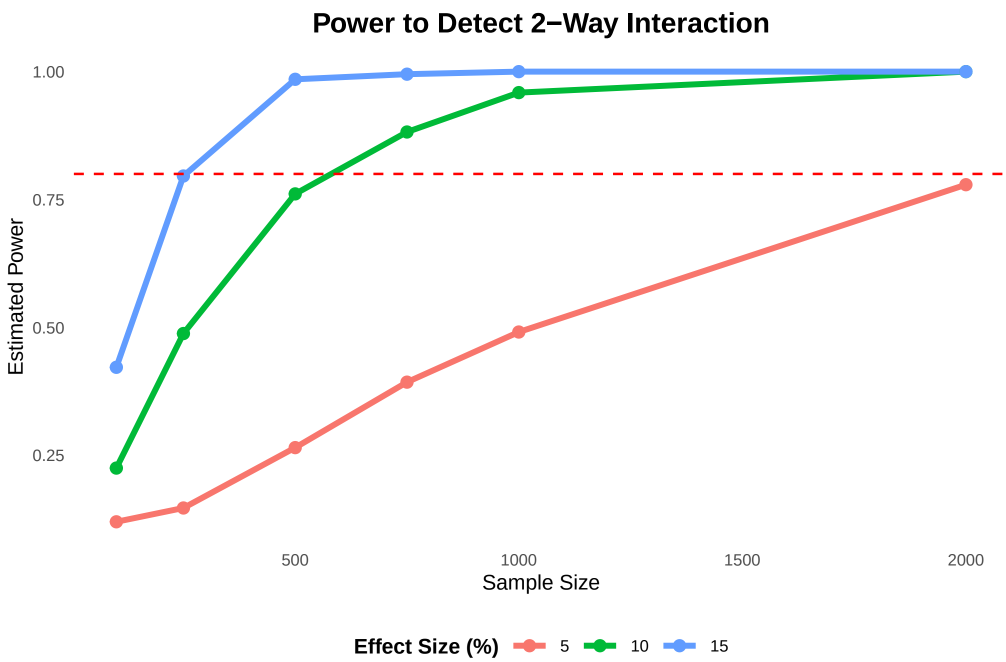


Figure 4: Power to Detect 2-Way Interaction

3 Analysis

3.1 Data

3.1.1 Outcomes Measured

Purchase Amount

In our experiment, one of our measured outcomes is the simulated total purchase amount from our survey participants. As described earlier, participants are given a theoretical $300 they can spend on any or all of the six tech-focused products they are shown. Depending on whether or not they fall into the bundling treatment, they are shown either a bundle option (which totals $230.38) with 3 more options that are each $34.99, $59.99, or $79.95 respectively, or 6 individual options. Both treatments have an option for the participant to spend $0.00. The items are the same in each treatment and have the same prices. Examples of the purchase screen in our Qualtrics survey are shown in appendix B.

The mean purchase amount in our responses was $109.83, with a median of $69.97. The min and max were $0 and $298.93 respectively. The standard deviation was $100.89. The resulting purchase amounts are detailed in the histogram below.

Binary Purchase Outcome

A second outcome we measured was whether or not a participant made a purchase. If they made any purchase (greater than $0.00 spent), they were marked as 1 for "any purchase". Otherwise, we marked them as 0 if they made no purchases. There were 201 participants who made any purchase, and 71 participants who made no purchase.

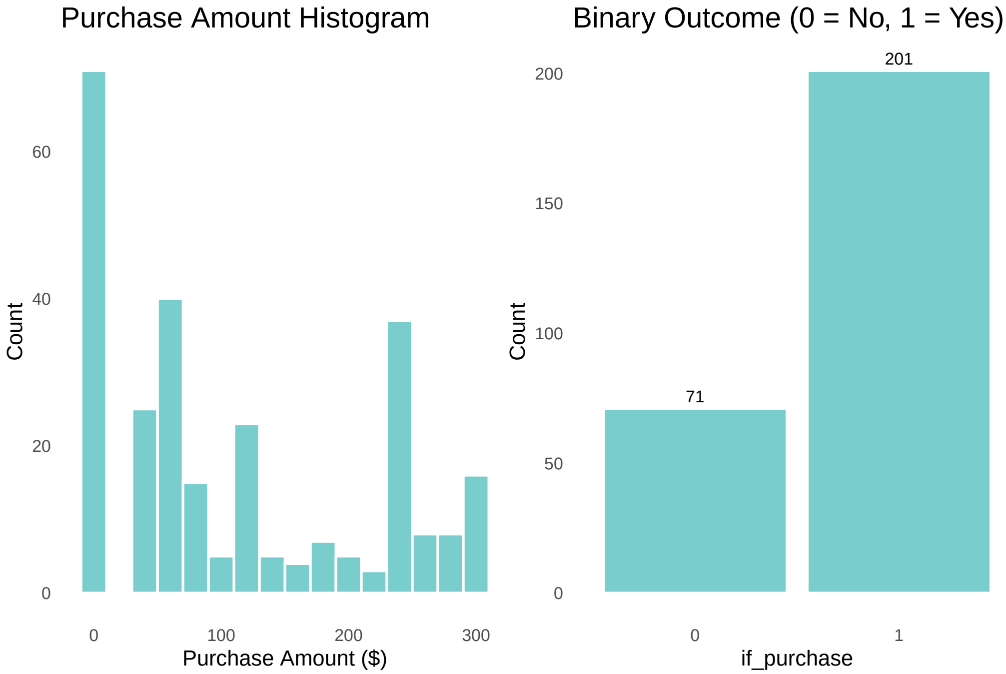


Figure 5: Outcome distribution plots

Compliance Measure

As a proxy for compliance in our treatment, we asked participants the following question: "Please confirm you read the product page and product descriptions". This question was optional, so participants did not have to answer this question with a confirmation. We used this as an attention check to identify the participants that "complied" in our experiment. "Compliance" means they fully read the descriptions, which implies they were making choices in the simulation that aligned with their real world buying choices.

In our survey, there were 224 who answered our proxy compliance question with a confirmation, and 48 who did not confirm they read the descriptions.

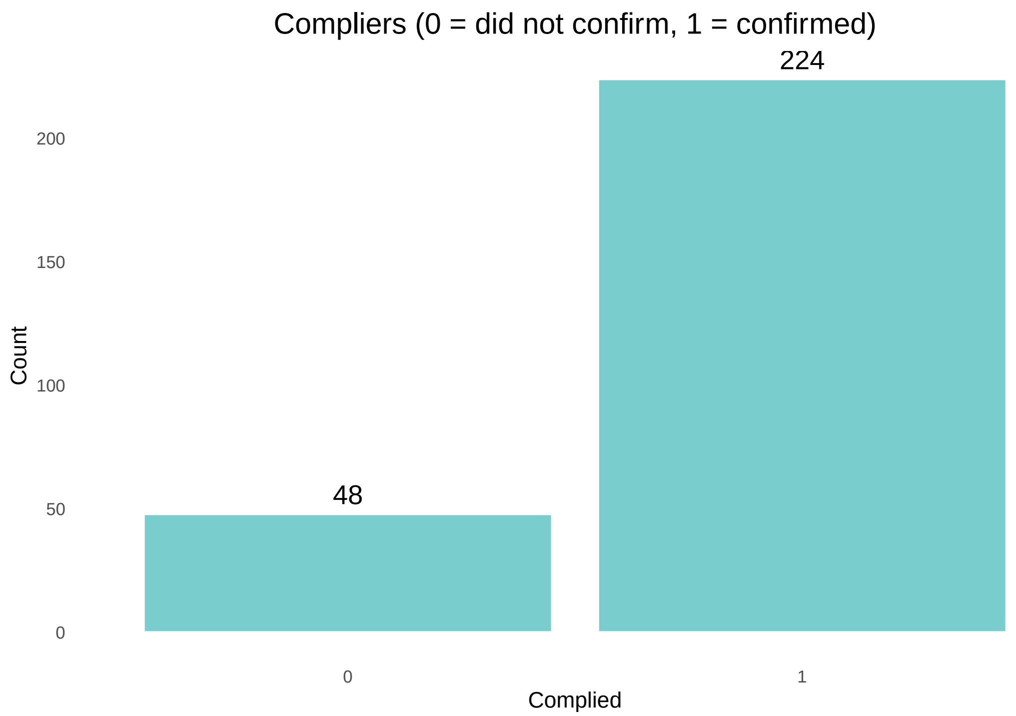


Figure 6: Compliance Measure

Perceived Personalization Measure

Additionally, we ran a check to confirm that the treatment assignment was consistent with participant perceptions. The result of this question indicated that 38% of those in the accurate personalization group reported that the messaging felt "definitely personalized," compared to 27% in the generic (no-personalization) group and 24% in the inaccurate personalization group. While not a dramatic separation, these differences suggest that the manipulation was modestly successful in shaping perceived personalization.

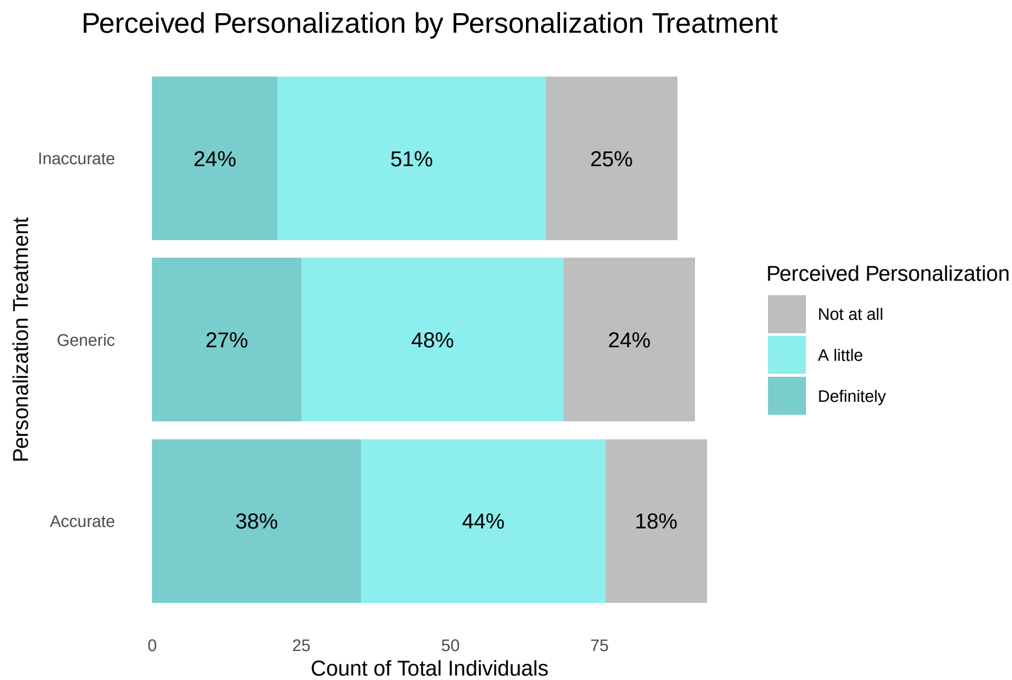


Figure 7: Perceived Personalization by Personalization Treatment

3.1.2 Covariate Features

In our survey, our covariate features are age, gender, and income level. These are collected via questions asked before the randomization treatment occurs. The distribution of our covariate features are displayed in the following figure.

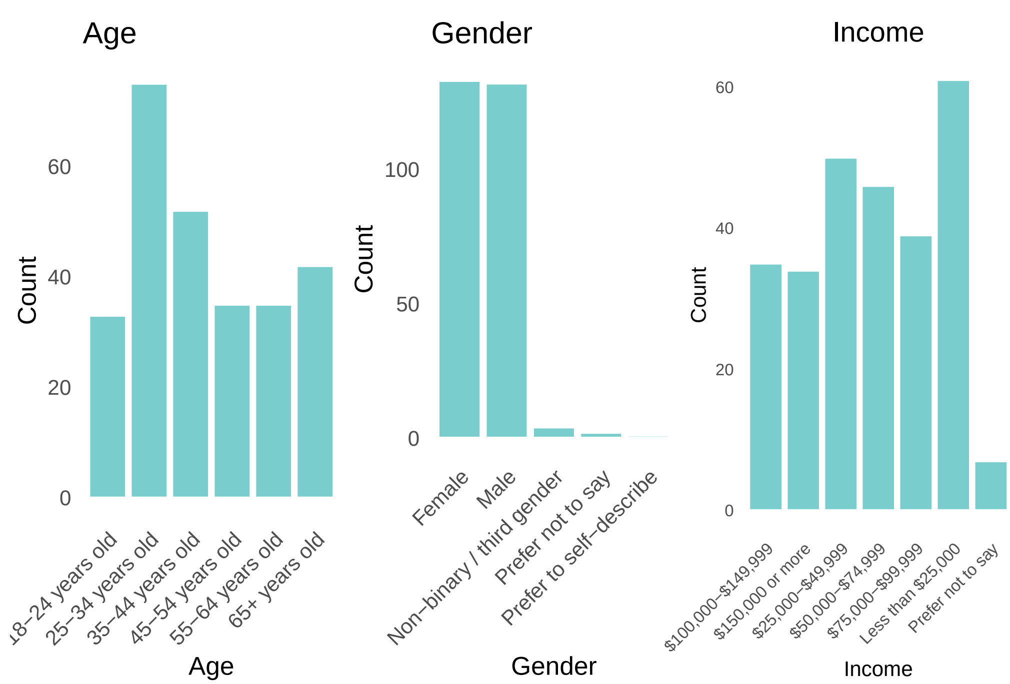


Figure 8: Covariate Features Plots

3.1.3 Models

For the analysis, we fitted four models for comparison of the purchase amount outcome:

Models	Formula	Outcome
Bundle treatment linear model	`lm(purchase_amount ~ bundle)`	Y1 = Purchase amount ($)
Personalization treatment linear model	`lm(purchase_amount ~ Personalization_Treatment)`	Y1 = Purchase amount ($)
Basic multifactorial linear model	`lm(purchase_amount ~ bundle * Personalization_Treatment)`	Y1 = Purchase amount ($)
Multifactorial linear model with covariates	`lm(purchase_amount ~ bundle * Personalization_Treatment + Age + Gender)`	Y1 = Purchase amount ($)

We also fitted four models for the comparison of the binary purchase outcome:

Models	Formula	Outcome
Bundle treatment logistic model	`glm(if_purchase ~ bundle, family = binomial(link = "logit"))`	Y2 = if_purchase (0/1)
Personalization treatment logistic model	`glm(if_purchase ~ Personalization_Treatment, family = binomial(link = "logit"))`	Y2 = if_purchase (0/1)
Basic multifactorial logistic model	`glm(if_purchase ~ bundle * Personalization_Treatment, family = binomial(link = "logit"))`	Y2 = if_purchase (0/1)
Multifactorial logistic model with covariates	`glm(if_purchase ~ bundle * Personalization_Treatment + Age + Gender, family = binomial(link = "logit"))`	Y2 = if_purchase (0/1)

Note: if_purchase (Y2) ==0: an individual's purchase amount is $0; ==1: purchase amount > $0.

Table 6: Linear Regression on Purchase Amount with Robust SE

	Dependent variable:			
	purchase_amount			
	(1)	(2)	(3)	(4)
bundle1	−20.88*		−18.11	−9.15
	(−44.78, 3.02)		(−61.96, 25.73)	(−55.13, 36.83)
Personalization_TreatmentAccurate		−14.52	−10.05	−1.47
		(−43.09, 14.05)	(−48.01, 27.91)	(−40.33, 37.40)
Personalization_TreatmentInaccurate		−12.28	−12.33	−2.19
		(−43.36, 18.80)	(−56.01, 31.36)	(−47.12, 42.75)
bundle1:Personalization_TreatmentAccurate			−8.07	−16.72
			(−65.18, 49.04)	(−74.74, 41.31)
bundle1:Personalization_TreatmentInaccurate			0.69	−13.23
			(−61.65, 63.03)	(−76.81, 50.36)
Constant	120.20***	118.77***	127.53***	117.01***
	(103.80, 136.60)	(96.94, 140.61)	(97.74, 157.32)	(56.31, 177.72)
Observations	272	272	272	272
R^2	0.01	0.004	0.01	0.10
Adjusted R^2	0.01	−0.003	−0.004	0.03
Residual Std. Error	100.53 (df = 270)	101.06 (df = 269)	101.07 (df = 266)	99.34 (df = 251)
F Statistic	2.93* (df = 1; 270)	0.55 (df = 2; 269)	0.81 (df = 5; 266)	1.43 (df = 20; 251)

Note: *p<0.1; **p<0.05; ***p<0.01

Robust (HC1) SEs in parentheses; 95% CIs shown. Model 4 repeats Model 3's interaction with controls for age, gender, and income.

3.1.3.1 Regression Models- Purchase Amount Outcome

The models in Table 6 estimate the effects of the bundle and personalization conditions on purchase amount.

Effect of Bundling

As we ran a multi-factor experiment, we were able to look at the effects of the individual treatments on the outcome variable. We see that adding a bundle reduced the purchase amount by about -20.88 compared to the baseline of no bundle. This effect could be considered significant at a significance level of 0.1.

Effect of Personalization

Neither accurate nor inaccurate personalization resulted in a significantly different treatment effect compared to our baseline of no personalization (generic personalization.) The effects and standard errors are all inclusive of 0.

Interaction Effects: Bundling X Personalization

The third model represents the interaction between bundling and personalization. It estimates that those in each of the treatment groups spent less than the control generic personalization, no bundling group. However, none of the coefficients report statistically significant differences, and the confidence intervals are all inclusive of zero, indicating uncertainty about the presence of a treatment effect and meaning we are unable to reject the null hypothesis that there is no difference between treatment and control groups.

The survey controlled for demographics, and the inclusion of demographics in Model 4 does raise the R^2 from .01 to .10, but the adjusted R^2 remains close to zero and an ANOVA test (Appendix C) confirms the addition of age and gender does not significantly improve explanatory power. Overall, these results suggest that, within this sample, there is no significant difference in purchase amounts between the treatment and control groups.

Table 7: Logistic Regression Results on Binary Outcome with Robust SE

	Dependent variable:			
	Any purchase (1=yes, 0=no)			
	(1)	(2)	(3)	(4)
bundle1	−0.52*		−0.32	−0.12
	(−1.07, 0.03)		(−1.26, 0.63)	(−1.18, 0.94)
Personalization_TreatmentAccurate		0.21	0.71	0.85
		(−0.47, 0.88)	(−0.39, 1.82)	(−0.37, 2.07)
Personalization_TreatmentInaccurate		−0.16	−0.20	−0.06
		(−0.81, 0.50)	(−1.16, 0.75)	(−1.14, 1.03)
bundle1:Personalization_TreatmentAccurate			−0.82	−1.01
			(−2.25, 0.60)	(−2.60, 0.58)
bundle1:Personalization_TreatmentInaccurate			0.10	−0.06
			(−1.23, 1.42)	(−1.59, 1.46)
Constant	1.31***	1.03***	1.19***	2.10**
	(0.90, 1.73)	(0.56, 1.50)	(0.50, 1.87)	(0.46, 3.74)
Observations	272	272	272	272
Log Likelihood	−154.41	−155.59	−152.83	−139.27
Akaike Inf. Crit.	312.83	317.18	317.66	320.53

Note: *p<0.1; **p<0.05; ***p<0.01

Model 4 represents same interaction as Model 3 between bundling and personalization with the addition of controls for age, gender and income.

3.1.3.2 Regression models - Binary Purchase Outcome Logistic regression models from Table 7 were estimated to examine the effects of bundling and personalization on the likelihood of making a purchase.

In Model 1, bundling is the only predictor of purchase likelihood. The intercept of 1.3148354 represents the predicted log-odds of making a purchase for participants in the unbundled (reference) group. Converting this to a probability, the unbundled group is estimated to have a 79% chance of purchase. The bundling coefficient indicates that, holding all else equal, being shown bundled products reduces the log-odds of purchase by -0.5199055 In odds ratio terms, this translates to multiplying the odds of purchase by approximately 0.59, meaning the bundled group has about 41% lower odds of making a purchase than the unbundled group. This effect is marginally significant at the 10% level, suggesting a potential, but not statistically strong, negative association between bundling and purchase likelihood.

Model 2 added main effects for accurate and inaccurate personalization (relative to a generic message) alongside bundling. None of the personalization effects were statistically significant, and the bundling effect was similar in size and significance to Model 1. Model 3 introduced interactions between bundling and personalization type. Neither interaction term was statistically significant, and the main effects remained non-significant. Model 4 included the same predictors and interactions as Model 3 but added demographic controls (age, gender, income). The inclusion of controls did not meaningfully change the coefficients, and no variables reached conventional levels of statistical significance.

Overall, across all four models, there is no strong statistical evidence that bundling, personalization, or their interaction significantly affect the likelihood of purchase. The direction of the coefficients suggests that bundling may slightly reduce purchase likelihood and that accurate personalization may slightly increase it, but these effects are small and statistically uncertain.

Logistic Regression Model Performance (Diagnostic Plots)

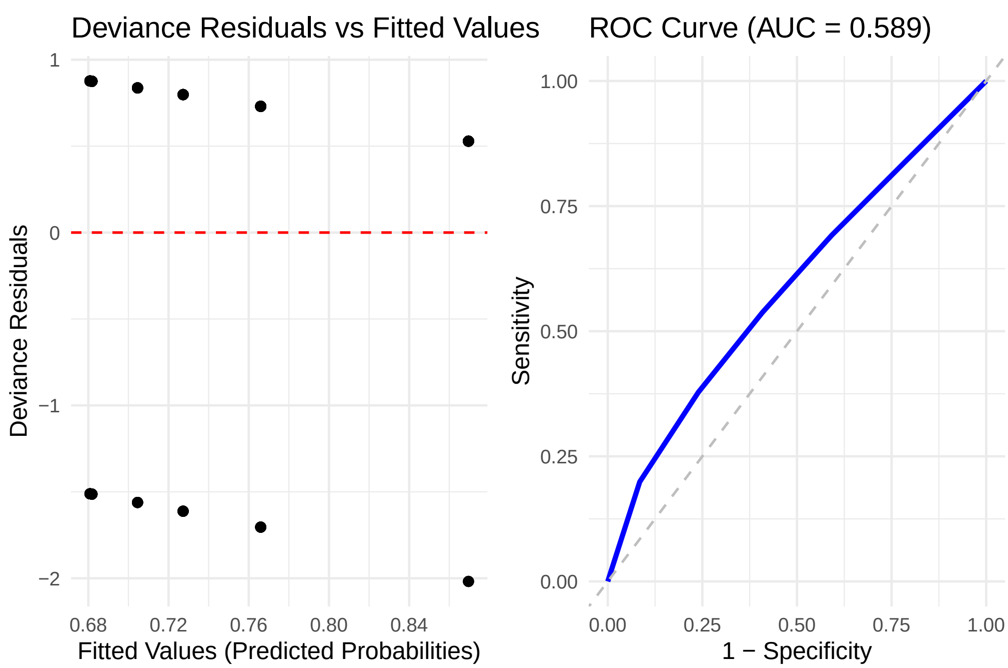


Figure 9: Logistic Regression Model Diagnostic Plots

We modeled purchase behavior as a binary outcome (purchase vs. no purchase) using logistic regression with the two treatment factors as predictors. The ROC curve yielded an AUC of 0.59, indicating the model's predictive ability is only slightly better than random chance (AUC = 0.5). The smooth, gently rising curve above the diagonal suggests weak discrimination between purchasers and non-purchasers. The deviance residuals versus fitted values plot shows two distinct parallel lines running diagonally from upper left to lower right, reflecting the categorical nature of the predictors—one binary and one with three levels. This pattern suggests that residuals cluster by treatment groups, which aligns with the factor structure but also highlights limited variability explained by the model.

Overall, these diagnostics confirm the logistic regression model's limited predictive power in this context.

Table 8: Regression on Purchase Amount and Binary Purchase Outcome

	Dependent variable:							
	purchase_amount				if_purchase			
	OLS Purchase amount (OLS)				*logistic* Any purchase (logit)			
bundle1	−20.88* (−44.78, 3.02)		−18.11 (−61.96, 25.73)	−9.15 (−55.13, 36.83)	−0.52* (−1.07, 0.03)		−0.32 (−1.26, 0.63)	−0.12 (−1.18, 0.94)
Personalization_TreatmentAccurate		−14.52 (−43.09, 14.05)	−10.05 (−48.01, 27.91)	−1.47 (−40.33, 37.40)		0.21 (−0.47, 0.88)	0.71 (−0.39, 1.82)	0.85 (−0.37, 2.07)
Personalization_TreatmentInaccurate		−12.28 (−43.36, 18.80)	−12.33 (−56.01, 31.36)	−2.19 (−47.12, 42.75)		−0.16 (−0.81, 0.50)	−0.20 (−1.16, 0.75)	−0.06 (−1.14, 1.03)
bundle1:Personalization_TreatmentAccurate			−8.07 (−65.18, 49.04)	−16.72 (−74.74, 41.31)			−0.82 (−2.25, 0.60)	−1.01 (−2.60, 0.58)
bundle1:Personalization_TreatmentInaccurate			0.69 (−61.65, 63.03)	−13.23 (−76.81, 50.36)			0.10 (−1.23, 1.42)	−0.06 (−1.59, 1.46)
Constant	120.20*** (103.80, 136.60)	118.77*** (96.94, 140.61)	127.53*** (97.74, 157.32)	117.01*** (56.31, 177.72)	1.31*** (0.90, 1.73)	1.03*** (0.56, 1.50)	1.19*** (0.50, 1.87)	2.10** (0.46, 3.74)
Observations	272	272	272	272	272	272	272	272
R^2	0.01	0.004	0.01	0.10				
Adjusted R^2	0.01	−0.003	−0.004	0.03				
Log Likelihood					−154.41	−155.59	−152.83	−139.27
Akaike Inf. Crit.					312.83	317.18	317.66	320.53
Residual Std. Error	100.53 (df = 270)	101.06 (df = 269)	101.07 (df = 266)	99.34 (df = 251)				
F Statistic	2.93* (df = 1; 270)	0.55 (df = 2; 269)	0.81 (df = 5; 266)	1.43 (df = 20; 251)				

Note: *p<0.1; **p<0.05; ***p<0.01

Columns 1–4: OLS on purchase amount (in currency units). Columns 5–8: Logistic regression on any purchase (1=yes, 0=no).

3.1.3.3 Regression models for both outcomes Table 8 includes the comparison of both outcomes: purchase amount and the binary purchase outcome.

Our regression analysis does not provide statistically significant evidence to reject the null hypotheses for H1, H2, or H3; that is, neither identity-based personalization, product bundling, nor their interaction had a detectable effect on total purchase amounts in our study

3.1.4 Heterogeneous Treatment Effect

To assess whether the personalization treatment effect varies by participants' subjective experience, we conducted an exploratory heterogeneity analysis using perceived personalization ("Definitely," "A little," "Not at all")—a post-treatment variable—as a moderator in regression models. While this approach cannot isolate the causal effect due to potential post-treatment bias, it provides diagnostic insight into variation in treatment impact. Results (table "Model regression result") show that in the linear model for purchase amount (Y1), "Definitely" personalized customers spent about 34 more units on average ($p < 0.05$), whereas "Not at all" personalized customers spent about 54 fewer units ($p < 0.001$). In the logistic model for purchase likelihood (Y2), "Definitely" personalization increased purchase probability (log-odds $+ 1.093$, $p < 0.05$), while "Not at all" personalization decreased it (log-odds $- 1.527$, $p < 0.001$). These results suggest meaningful heterogeneity in treatment effects linked to participants' perceptions of personalization, offering diagnostic evidence that the treatment's impact depends on how personalized the intervention is perceived to be. Table 10 compares the heterogeneous treatment effects on purchase amount and on the binary purchase outcome, using perceived personalization as the diagnostic.

Moderator	Diagnostic model	Outcome
personalization (= perceived personalization)	`lm(purchase_amount (Y1) ~ bundle * Personalization_Treatment + personalization)`	`Y1 = Purchase amount (\\$)`
personalization (= perceived personalization)	`glm(if_purchase (Y2) ~ bundle * Personalization_Treatment)`	`Y2 = if_purchase (0,1)`

3.1.5 Limitations

Our linear regression model used the original outcome variable, purchase amount, which consists of bounded discrete values (\$30 ~ \$300). The residual vs. fitted value plot reveals vertical lines at specific points, reflecting this discrete and bounded nature. This violates the assumption of normally distributed residuals and suggests the model may not fully capture the underlying data structure. The QQ plot further indicates deviation from normality, with residuals showing right skewness, implying that the outcome distribution is not symmetric. These issues can lead to biased standard errors and unreliable coefficient estimates, reducing the model's validity and interpretability. Although we attempted to address these issues using robust standard errors, the fundamental mismatch between the data distribution and model assumptions remains a limitation.

Table 10: HTE Linear and Logistic Regression Result with Perceived Personalization Diagnostic

	Dependent variable:	
	purchase_amount	if_purchase
	OLS	*logistic*
	(1)	(2)
bundle1	-21.865	-0.524
	(21.504)	(0.532)
Personalization_TreatmentAccurate	-23.715	0.313
	(18.284)	(0.615)
Personalization_TreatmentInaccurate	-14.085	-0.326
	(20.363)	(0.542)
personalizationDefinitely	33.845^{**}	1.093^{**}
	(14.195)	(0.459)
personalizationNot at all	-53.887^{***}	-1.527^{***}
	(14.158)	(0.345)
bundle1:Personalization_TreatmentAccurate	6.044	-0.406
	(27.953)	(0.786)
bundle1:Personalization_TreatmentInaccurate	7.662	0.375
	(30.670)	(0.761)
Constant	133.074^{***}	1.545^{***}
	(15.412)	(0.436)
Observations	272	272
R^2	0.110	
Adjusted R^2	0.086	
Log Likelihood		-132.404
Akaike Inf. Crit.		280.809
Residual Std. Error	96.431 (df = 264)	
F Statistic	4.662^{***} (df = 7; 264)	
Note:		*p<0.1; **p<0.05; ***p<0.01

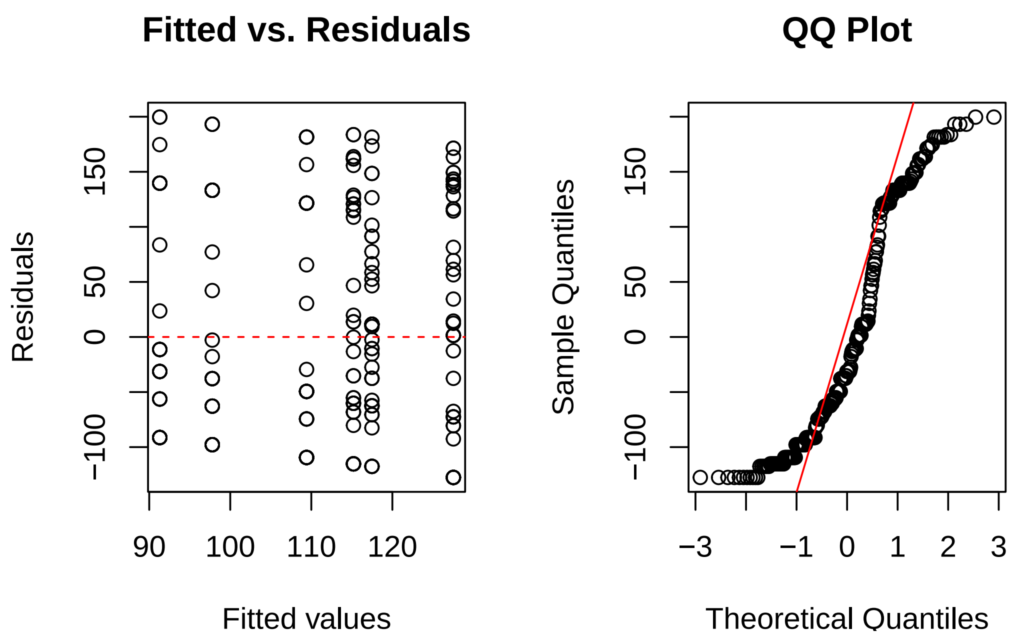


Figure 10: Linear Regression Model Diagnostic Plots

3.2 Discussion

This study provides important insights into the interaction between personalization and product bundling, though several limitations and design lessons emerged that should inform future work.

Our sample size of 272 participants, while sufficient to detect relatively large effects (~15%) according to our power analysis, may not have been large enough to uncover more modest but practically meaningful differences. In a multifactorial design with six experimental groups, each condition had an average of ~45 participants, limiting statistical power to detect small interaction effects at the individual lifestyle identity level. A larger-scale study, particularly with stratified personalization sampling, would allow for a more granular exploration of heterogeneous effects across demographics or consumer profiles.

While the experiment was constructed to simulate a realistic online shopping experience, including product images, profile-based recommendations, and a fixed budget constraint, it remains a hypothetical purchasing environment. Even with the inclusion of an incentive (i.e., one randomly selected participant receiving either their chosen items or the remaining budget in cash-equivalent), this setup cannot fully replicate the psychological and contextual variables involved in real-world purchase behavior. However, the effort to ground the study in real trade-offs and recognizable consumer products likely improved external validity compared to purely attitudinal or preference-based measures.

A key design improvement for future iterations would be to offer participants in the bundling condition the option to purchase bundle items individually. In our design, the bundle was presented as a fixed set of three items, while the remaining and different three items were shown separately. In real-world e-commerce platforms, users often have the flexibility to purchase items either within or outside a bundle. Allowing participants to disaggregate the bundle could more accurately reflect consumer autonomy and potentially increase purchases by lowering the commitment to an all-or-nothing set. Additionally, it may be beneficial to measure secondary outcomes, such as time spent viewing product descriptions, order of selection, or engagement with messaging, to better understand how and when personalization or bundling exerts influence. Even if purchase totals remain unaffected, these intermediate metrics may reveal subtle behavioral shifts that could compound over time in real commercial settings.

4 References

Arora, N., Ensslen, D., Fiedler, L., Liu, W. W., Robinson, K., Stein, E., & Schüler, G. (2021, November 12). The value of getting personalization right—or wrong—is multiplying. McKinsey & Company.https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying

Rao, V.R., Russell, G.J., Bhargava, H. et al. Emerging Trends in Product Bundling: Investigating Consumer Choice and Firm Behavior. Cust. Need. and Solut. 5, 107–120 (2018). https://doi.org/10.1007/s40547-017-0075-x

XinYun and Myung Hwan Chun. The impact of personalized recommendation on purchase intention under the background of big data. https://www.aimspress.com/aimspress-data/bdia/2024/0/PDF/bdia-08-005.pdf

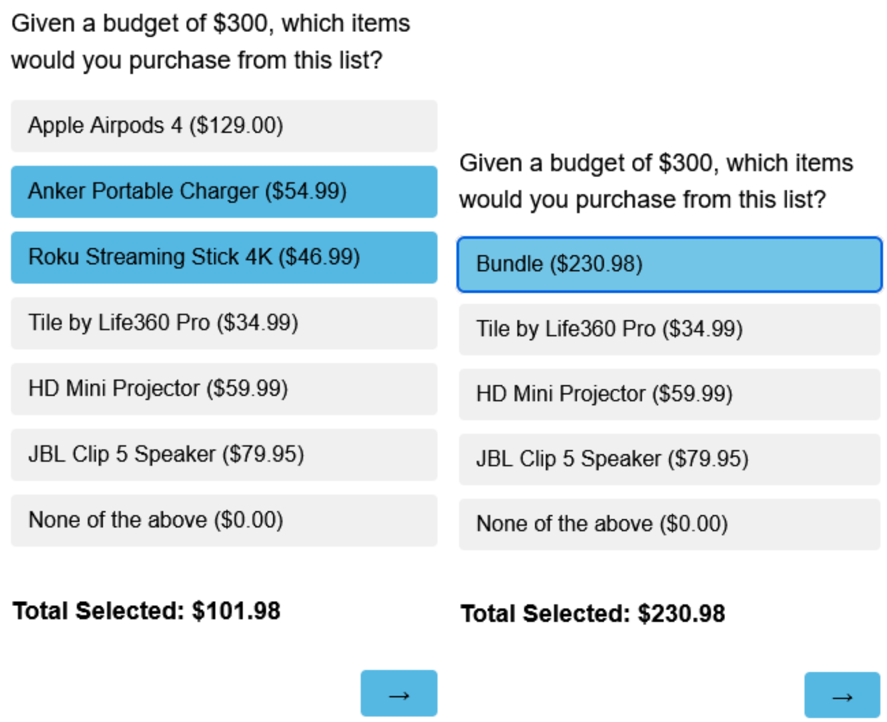
5 Appendix

5.1 Appendix A

In bundled conditions, three of the six products, AirPods, Portable Charger, and Streaming Stick, were grouped together with a promotional "bundle" label and accompanying message, while the other three were shown individually. In the non-bundled condition, all six products were presented as individual options with no grouping.

5.2 Appendix B

Participants are given a theoretical \$300 they can spend on any or all of the six tech-focused products they are shown. Depending on whether or not they fall into the bundling treatment, they are shown either a bundle option (which totals \$230.38) with 3 more options that are each \$34.99, \$59.99, or \$79.95 respectively, or 6 individual options. Both treatments have an option for the participant to spend \$0.00. The items are the same in each treatment and have the same prices.



5.3 Appendix C

Anova test comparing regression models for purchase amount outcomes. The test reveals no significant improvements in model fit across the more complex and simple models.

Analysis of Variance Table

Model 1: purchase_amount ~ bundle Model 2: purchase_amount ~ Personalization_Treatment Model 3: purchase_amount ~ bundle * Personalization_Treatment Model 4: purchase_amount ~ bundle * Personalization_Treatment + Age + Gender + Q15 Res.Df RSS Df Sum of Sq F Pr(>F)
1 270 2728765
2 269 2747256 1 -18491
3 266 2717190 3 30066 1.0156 0.38633
4 251 2476841 15 240349 1.6238 0.06786 . — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ” 0.05 ‘.’ 0.1 ’ ’ 1

5.4 Appendix D

CACE

Earlier, we described the proxy question we used to identify participants that “complied” in our experiment. “Compliance” means they fully read the descriptions, which implies they were making choices in the simulation that aligned with their real world buying choices.

To calculate the CACE, we took the subset of participants who answered this question with a confirmation, and estimated the effects of bundling and personalization on purchase amount. Overall, the point estimates for compliers were smaller than for the entire population, but none of the effects were shown to be significant.

Table 11: CACE Regression on Purchase Amount

	Dependent variable:
	purchase_amount
bundle1	−12.500
	(24.420)
Personalization_TreatmentAccurate	−3.466
	(21.773)
Personalization_TreatmentInaccurate	−6.028
	(25.016)
bundle1:Personalization_TreatmentAccurate	−19.141
	(32.604)
bundle1:Personalization_TreatmentInaccurate	−18.448
	(35.419)
Constant	129.695***
	(16.500)
Observations	224
R^2	0.021
Adjusted R^2	−0.002
Residual Std. Error	103.538 (df = 218)
F Statistic	0.929 (df = 5; 218)
Note:	*p<0.1; **p<0.05; ***p<0.01