Is It Always "the Faster, the Better"? The Role of a Nudging-Based Function in the Relationship between Product Delivery and Online Sales

Hongpeng Wang

School of Management, Lanzhou University, Lanzhou, Gansu P. R. CHINA

Department of Business Analytics, Tippie College of Business,

University of Iowa, Iowa City, Iowa U.S.A.

whp@lzu.edu.cn

Qianzhou Du*

Faculty of Business for Science and Technology, School of Management, University of Science and Technology of China, Hefei, Anhui P. R. CHINA

qianzhoudu@ustc.edu.cn

Liangfei Qiu

Department of Information Systems and Operations Management,

Warrington College of Business, University of Florida, Gainesville, Florida U.S.A.

liangfei.qiu@warrington.ufl.edu

Weiguo Fan

Department of Business Analytics, Tippie College of Business,

University of Iowa, Iowa City, Iowa U.S.A.

weiguo-fan@uiowa.edu

* Qianzhou Du is the corresponding author.

Online Appendix A. Supplementary Results for the P_2 vs. P_3 comparison

Using the same approach as that described in Study 1 and Study 3, Table A1 shows the P_2 vs. P_3 results. We find that the sales coefficient remains positive, the price coefficient is positive, and the shipping discounts coefficient is negative. These results indicate that during P_3, slow-delivery service providers did not continue the preemptive strategies of price reductions and free shipping observed in P_2. Moreover, the coefficient on the feedback score remains positive, similar to the P_1 vs. P_2 results. This finding implies that the introduction of the function has enhanced eBay's platform regulation and encouraged sellers to exert greater efforts to improve service quality. Although sellers continuously work to improve service quality during both the announcement and enforcement periods, only after the policy is fully implemented do product sales increase significantly. Furthermore, we find that review valence is positive, while review volume and variance are negative. These findings suggest that after the toggle was implemented, the overall ratings for slow-delivery products have improved, with fewer consumer complaints, leading to a reduction in rating dispersion. Taken together, these results provide a consistent and complementary picture alongside the P_1 vs. P_2 and P_1 vs. P_3 comparisons.

Table A1. Treatment Effects for the P_2 vs. P_3 Comparison

ATT ($\hat{\imath}$)	*Staggered* DiD	*Synthetic* DiD	*Continuous* DiD	Multiequation
$log(Sales_{ijt}+1)$	.05444**	.03527***	.00593***	
	(.02951)	(.00767)	(.00068)	
$log(Price_{ijt})$	.00395	.00315	.00015	
	(.00257)	(.00333)	(.00014)	
$ShippingDiscount_{ijt}$	-.00553***	-.00654***	-.00096***	
	(.00081)	(.00101)	(.00022)	
$FeedbackScore_{ijt}$	.00108***	.00069**	.00004***	
	(.00024)	(.00028)	(.00001)	
$Valence_{ijt}$	.11275***	.05953***	.00121***	.10183***
	(.02846)	(.02012)	(.00033)	(.00493)
$log(Volume_{ijt}+1)$	-.28847***	-.15706***	-.00752***	-.23694***
	(.02566)	(.04722)	(.00134)	(.01578)
$Variance_{ijt}$	-.08113***	-.06120***	.00014	-.06435***
	(.02583)	(.02285)	(.00028)	(.00528)

Notes. Robust standard errors are clustered at the seller-product level and reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Online Appendix B. Treatment Effects with a 1:1 Nearest-Neighbor Matching Method

In our dataset, slow- delivery products constitute a substantial portion. To ensure that the imbalance between fast and slow delivery does not bias our results, we carry out a 1:1 nearest- neighbor matching exercise. Table B1 verifies that the differences in the sample sizes between the control and treatment groups do not affect our main findings.

Table B1. Treatment Effects with a 1:1 Nearest-Neighbor Matching Method

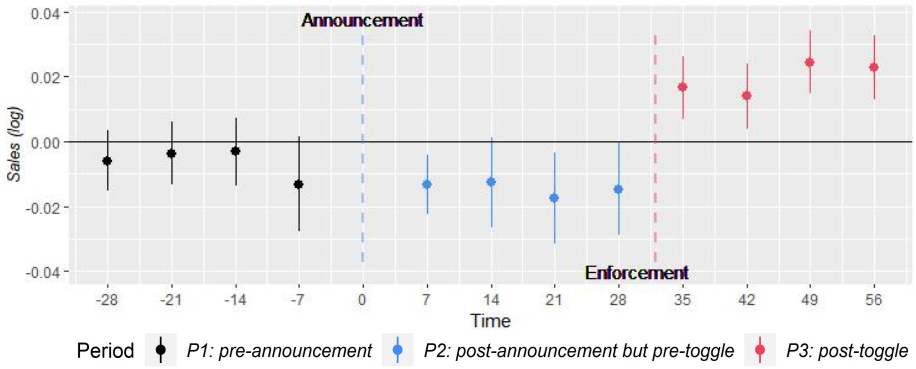
ATT ($\hat{e}$)	P_1 vs. P_2	P_1 vs. P_3
$log(Sales_{ijt}+1)$	-.00089	.00547***
	(.00059)	(.00127)
$log(Price_{ijt})$	-.00020**	-.00011
	(.00009)	(.00021)
$ShippingDiscount_{ijt}$	.00187***	.00002
	(.00019)	(.00004)
$FeedbackScore_{ijt}$	.00002***	.00010***
	(.00000)	(.00002)
$Valence_{ijt}$	.00026	.00116**
	(.00019)	(.00053)
$log(Volume_{ijt}+1)$	.00298***	-.00314***
	(.00064)	(.00100)
$Variance_{ijt}$	.00043**	.00058
	(.00018)	(.00035)

Notes. Robust standard errors are clustered at the seller-product level and reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Online Appendix C. Event-study Results for *Across-product* DiD Analyses

In Section 4.3.1, we estimate the relative-time model using the full set of observations spanning P_1, P_2, and P_3. However, the across-product DiD design focuses on short time windows around the policy announcement or enforcement. We thus re-plot the corresponding event-study figure for this primary specification. Figure C1 presents results that are qualitatively similar to those reported in Figure 4 in the main text.

Figure C1. Event-study Results

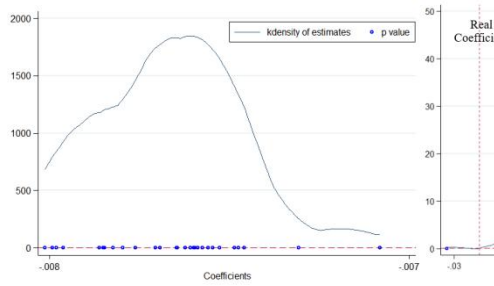


Notes. Dots represent the ATT estimates, and bars depict the corresponding 95% confidence intervals. The blue and red horizontal dashed lines denote the official announcement and enforcement of the GDT policy, respectively. Pre-announcement period: Time < 0; post-announcement and pre-toggle period: 0 ≤ Time < 30; and post-toggle period: Time ≥ 30.

Online Appendix D. Supplementary Results for the Falsification and Placebo Tests

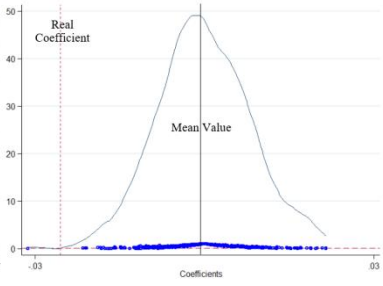
Using the same approach as that described in Section 4.3, Figure D1 visualizes the falsification and placebo tests for prices, shipping discounts, and the rate of positive feedback scores. Each of the treatment–outcome pairs shows that the coefficients are nonsignificant in most cases and that the mean value is far from the real coefficient.

Figure D1. Supplementary Results for the Falsification and Placebo Tests



Prices: DiD Analyses for "Phantom" Regressions

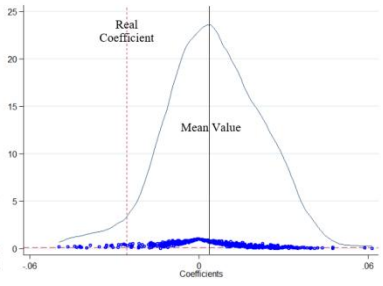
t = 60.1240, p = 0.0000



Prices: DiD Analyses for Counterfeit Groups (P_1 vs. P_2)

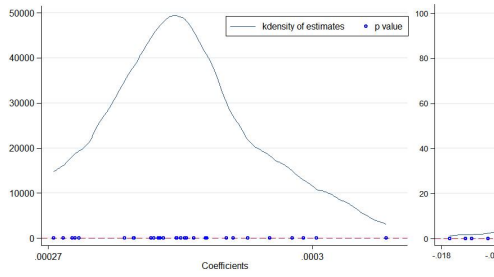
t = 2.7962, p = 0.0054

t = 39.5401, p = 0.0000

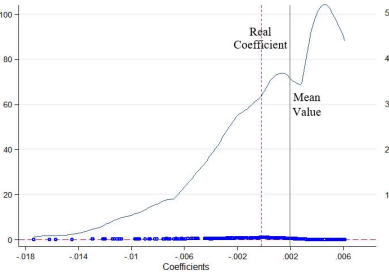


Prices: DiD Analyses for Counterfeit Groups (P_1 vs. P_3)

t = 3.1628, p = 0.0017

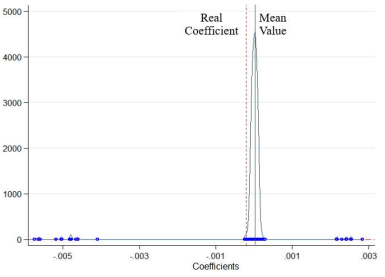


Free Shipping: DiD Analyses for "Phantom" Regressions



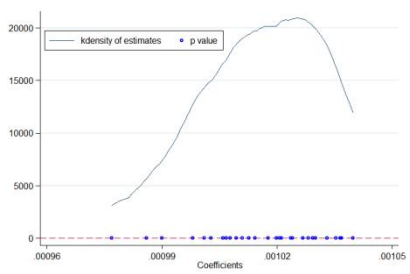
Free Shipping: DiD Analyses for Counterfeit Groups (P_1 vs. P_2)

t = -1.6e+02, p = 0.0000

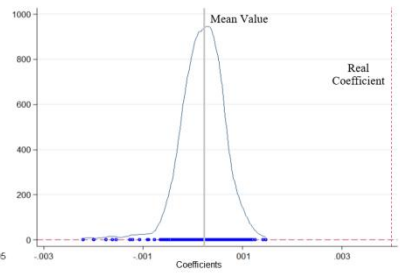


Free Shipping: DiD Analyses for Counterfeit Groups (P_1 vs. P_3)

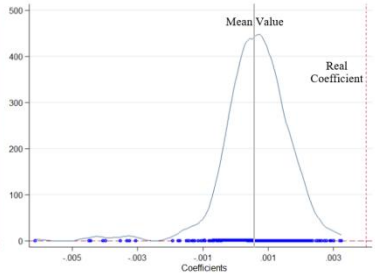
t = -71.2173, p = 0.0000



Positive Feedback: DiD Analyses for "Phantom" Regressions



Positive Feedback: DiD Analyses for Counterfeit Groups (P_1 vs. P_2)



Positive Feedback: DiD Analyses for Counterfeit Groups (P_1 vs. P_3)

Notes. Blue hollow points represent the p value corresponding to each estimate; if a point is above the red horizontal line in Figure D1, then it means the estimate is not significant at the 0.1 significance level. The red vertical line denotes the real coefficient (P_1 vs. P_3) estimated by the DiD model with TWFE.

Online Appendix E. Subsample Analyses

We conduct a robustness check by excluding the products with free returns. The results, reported in Table E1, are consistent with the findings of our full sample, indicating that a free return policy is likely not a significant confounding factor.

Table E1. Subsample Analyses Excluding Products with Free Returns

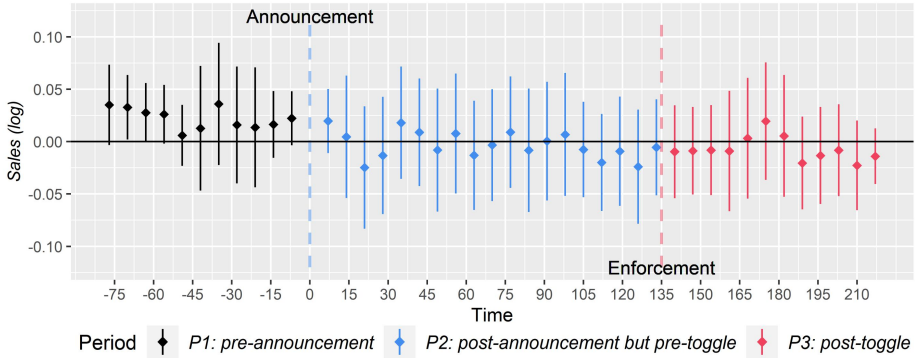
ATT ($\hat{u}$)	*Staggered* DiD		*Synthetic* DiD		*Continuous* DiD	
	P_1 vs. P_2	P_1 vs. P_3	P_1 vs. P_2	P_1 vs. P_3	P_1 vs. P_2	P_1 vs. P_3
$log(Sales_{ijt}+1)$	.03824	.09145**	-.00205	.07838**	-.00089	.00544***
	(.02666)	(.03949)	(.02207)	(.03170)	(.00059)	(.00129)

Notes. Robust standard errors are clustered at the seller-product level and reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Online Appendix F. Event Study Results for Intangible Products

The results of the event study depicted in Online Appendix F show a nonsignificant trend in nondelivery product sales, with consistent estimates over time, thus confirming that concurrent delivery-unrelated policies do not act as confounding factors.

Figure F1. Event Study Results for Intangible Products



Notes. Diamonds represent the estimated coefficients, and bars depict the corresponding 95% confidence intervals. The blue and red horizontal dashed lines denote the official announcement and enforcement of the GDT policy, respectively. Pre-announcement period: Time < 0; post-announcement and pre-toggle period: 0 ≤ Time < 135; and post-toggle period: Time ≥ 135.

Online Appendix G. External Policies of Competing Platforms

Table G1 shows the suggestive test results by regressing the total traffic from the Amazon and Walmart websites on eBay sellers' decisions to offer fast delivery. Owing to limitations in the R package "staggered", we are unable to include control variables in the *Staggered* DiD model. Thus, we incorporate total traffic as a control in *Synthetic* and *Continuous* DiD models to validate the robustness of our findings in Table G2.

Table G1. Relationship between eBay Sellers' Decisions and Competitors' Traffic

	$Slow_{ijt}$
$Traffic\_A\&W_t$	-.00004
	(.00003)
Controls	Yes
Seller-Product Fixed Effects	Yes

Notes. Robust standard errors are clustered at the seller-product level and reported in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table G2. Controlling for the Traffic of eBay's Competitors

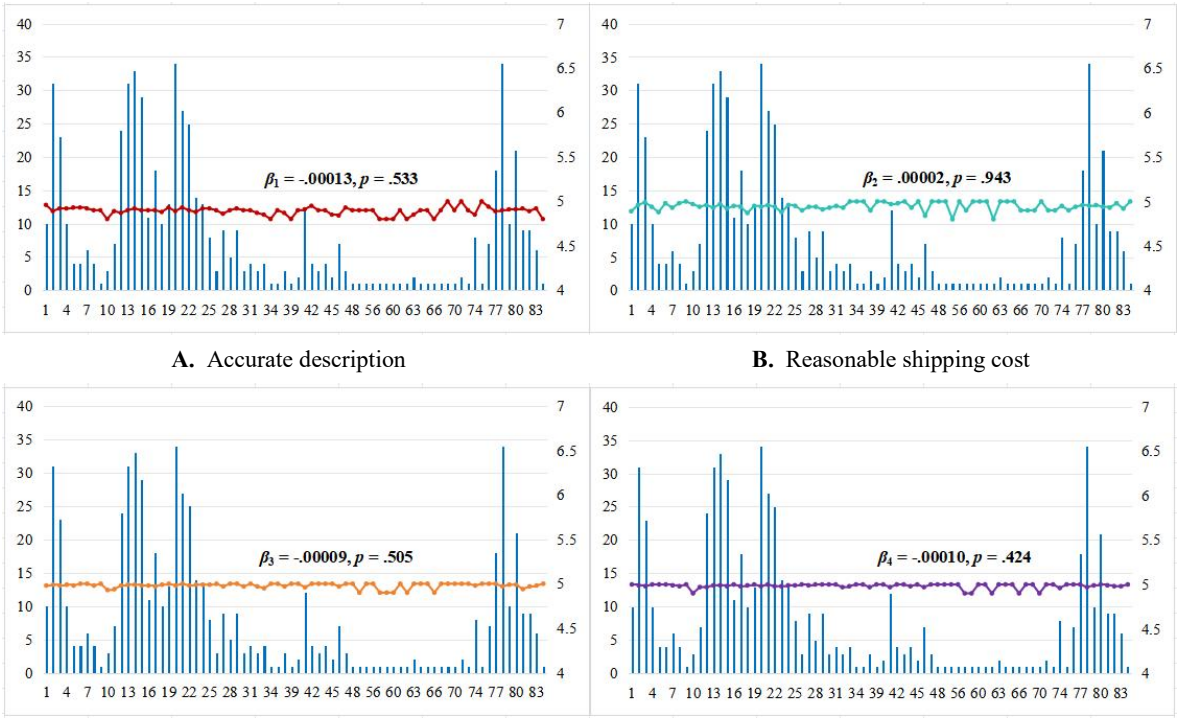
ATT ($\hat{e}$)	*Synthetic* DiD		*Continuous* DiD	
	P_1 vs. P_2	P_1 vs. P_3	P_1 vs. P_2	P_1 vs. P_3
$log(Sales_{ijt}+1)$	-.00135	.07893**	-.00089	.00544***
	(.01824)	(.03121)	(.00059)	(.00127)

Notes. Robust standard errors are clustered at the seller-product level and reported in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Online Appendix H. Alternative Explanations

This section provides detailed analyses to rule out alternative explanations for our findings. First, to address the concern that the phased rollout of GDT may have prioritized higher-quality sellers earlier, we examine whether the sequence in which sellers became eligible for GDT is systematically related to seller quality. Specifically, we track the number of newly enrolled GDT sellers from the start of GDT implementation and measure seller quality using four distinct seller rating scores: accurate description, reasonable shipping cost, shipping speed, and communication (Hui et al., 2025; Kricheli-Katz and Regev, 2016). The regression results in Figure H1 show no statistically significant relationships, indicating that the observed relative sales growth of slow-delivery products is not attributable to a decline in seller quality among newly enrolled GDT sellers.

Figure H1. Relationship between the Sequence of Eligibility and Seller Rating Scores



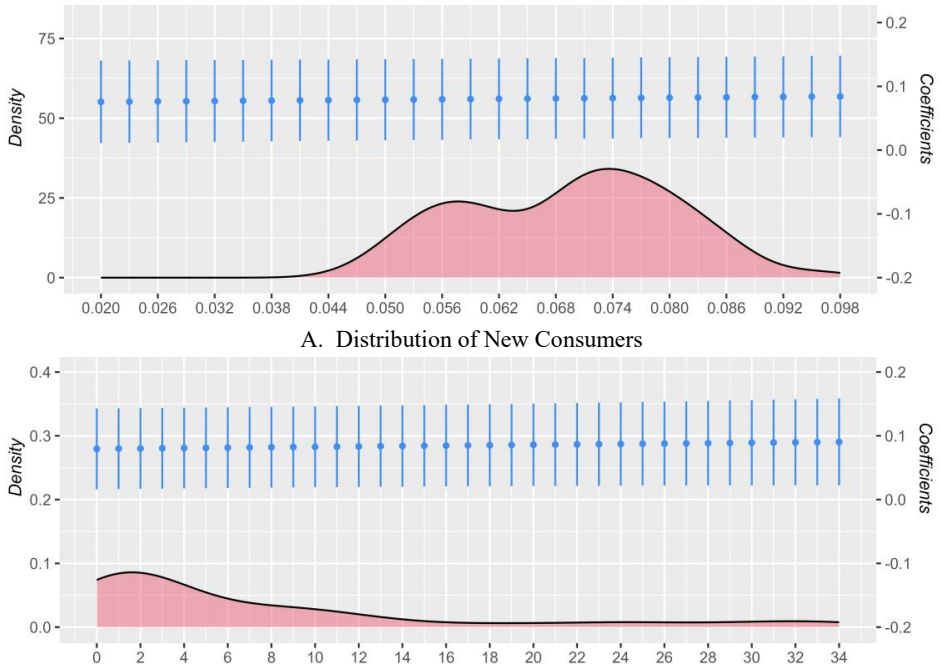
A. Accurate description **B.** Reasonable shipping cost

C. Shipping speed **D.** Communication

Notes. The x-axis represents the time sellers joined after GDT policy implementation, and the y-axis represents the number of sellers entering the GDT policy. The red, green, orange, and blue dots reflect average seller ratings for accurate description, reasonable shipping cost, shipping speed, and communication, respectively.

Second, to address the possibility that changes in user composition drive our results, the Johnson–Neyman technique is used to test whether and how the treatment effect of the GDT toggle varies with user composition (i.e., the proportion of new consumers and the number of participating sellers). This method identifies the ranges of moderator values over which the treatment effect is statistically significant (Spiller et al., 2013). As shown in the two panels of Figure H2, the treatment effect remains statistically significant across the full observed range of both moderators, suggesting that user composition does not moderate the toggle's impact on sales.

Figure H2. Johnson–Neyman Graph



A. Distribution of New Consumers

B. Distribution of Sellers Entering the GDT Policy

Notes. The x-axes represent the values of the moderator variables: i) the proportion of new consumers and ii) the number of GDT-participating sellers. The left y-axes plot the estimated treatment effect coefficients (i.e., the effect of the toggle function on product sales) along with their 95% confidence intervals, where dots indicate the point estimates and bars depict the confidence intervals. The right y-axes display the density distribution of the moderator variable, with the red shaded area highlighting the distribution for either the proportion of new consumers or the number of sellers joining the policy.

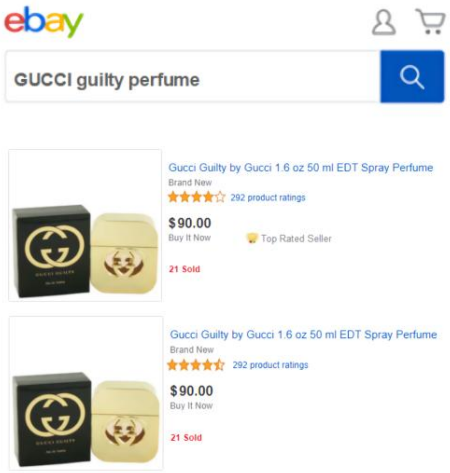
Online Appendix I. Experimental Materials

Panel I1. Questionnaire of the Prestudy Test

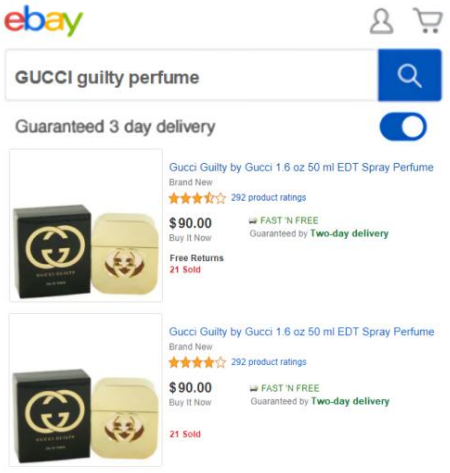
What is your gender?	☐ Male ☐ Female
Your age in years:	[-*please enter*-]
How comfortable are you with computers?	Completely uncomfortable ○ ○ ○ ○ ○ ○ ○ Completely comfortable
How frequently do you shop online?	Never ○ ○ ○ ○ ○ ○ ○ Always

Panel I2. Experimental Screen Shots

Simultaneous Attribute Choices



Sequential Attribute Choices



Panel I3. Survey of Follow-up Task

Why do you purchase this product?

☐ High product reviews
☐ Low price
☐ Fast delivery
☐ Top-rated seller
☐ Free shipping

Please consider the new product:

Are you willing to purchase the new product over your previous choice?	Extremely unwilling ○ ○ ○ ○ ○ ○ ○ Extremely willing
How satisfied are you with your previous choice?	Extremely dissatisfied ○ ○ ○ ○ ○ ○ ○ Extremely satisfied
Are you interested in the -*particular item*-?	Extremely uninterested ○ ○ ○ ○ ○ ○ ○ Extremely interested
How familiar are you with the -*particular item*-?	Extremely unfamiliar ○ ○ ○ ○ ○ ○ ○ Extremely familiar

Notes. For online Experiments 1b and 2, we add additional questions, e.g., education, income, internet usage, and eBay experiences, to the prestudy test. "*Particular item*" refers to the perfume or instant camera in different experiments.

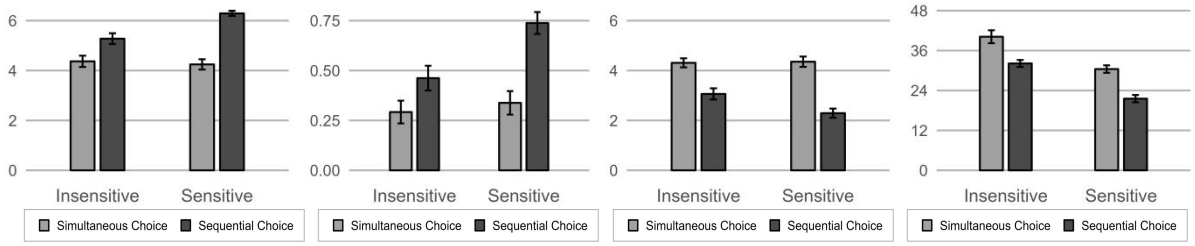
Online Appendix J. Supplementary Experiment Results

Table J1. MANOVA and Cross-Tab Results (Experiment 1a)

	Delivery-insensitive		Delivery-sensitive		Significant effects	F values or χ^2 values
	Simultaneous attribute choice	Sequential attribute choice	Simultaneous attribute choice	Sequential attribute choice		
Purchase satisfaction	4.369 (0.226)	5.277 (0.218)	4.246 (0.206)	6.292 (0.102)	DP	5.260**
					CA	57.645***
					DP × CA	8.563***
Preference consistency	0.292 (0.057)	0.462 (0.062)	0.338 (0.059)	0.738 (0.055)	DP	6.834***
					CA	21.231***
					DP × CA	27.525***
Willingness to change	4.308 (0.181)	3.062 (0.225)	4.354 (0.207)	2.292 (0.183)	DP	3.276*
					CA	68.554***
					DP × CA	4.166**
Search time	40.157 (1.937)	32.132 (1.040)	30.442 (1.129)	21.547 (1.105)	DP	56.224***
					CA	39.058***
					DP × CA	.103

Notes: DP = delivery preference, CA = choice architecture, and DP × CA = interactions between DP and CA.

Figure J1. Comparisons between CA and DP (Experiment 1a)



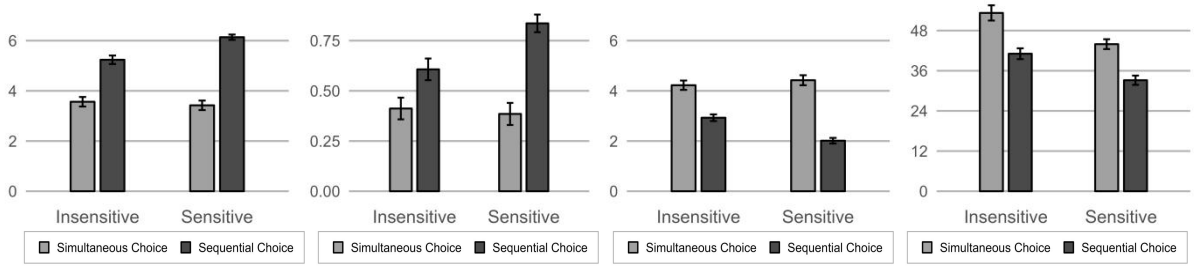
A. Purchase satisfaction **B.** Preference consistency **C.** Willingness to change **D.** Search time

Table J2. MANOVA and Cross-Tab Results (Experiment 1b)

	Delivery-insensitive		Delivery-sensitive		Significant effects	F values or χ^2 values
	Simultaneous attribute choice	Sequential attribute choice	Simultaneous attribute choice	Sequential attribute choice		
Purchase satisfaction	3.565 (0.190)	5.238 (0.169)	3.423 (0.190)	6.137 (0.106)	DP	4.920**
					CA	165.157***
					DP × CA	9.290***
Preference consistency	0.412 (0.054)	0.607 (0.054)	0.385 (0.055)	0.836 (0.044)	DP	2.837*
					CA	32.022***
					DP × CA	30.534***
Willingness to change	4.224 (0.186)	2.929 (0.133)	4.423 (0.199)	2.014 (0.111)	DP	4.778**
					CA	128.120***
					DP × CA	11.596***
Search time	53.280 (2.277)	41.087 (1.619)	43.951 (1.470)	33.174 (1.383)	DP	23.897***
					CA	42.410***
					DP × CA	.161

Notes: DP = delivery preference, CA = choice architecture, and DP × CA = interactions between DP and CA.

Figure J2. Comparisons between CA and DP (Experiment 1b)



A. Purchase satisfaction **B.** Preference consistency **C.** Willingness to change **D.** Search time

Online Appendix K. Principal Component Analysis

Before running the seemingly unrelated regression, Table K1 tests whether the residual variation in all outcome variables, i.e., review valence, volume, and variance, is highly correlated. The correlation matrix of the two stages, i.e., P_1 vs. P_2 and P_1 vs. P_3, shows the high significance of the residuals between each other.

Table K1. Correlation Matrix of Residuals

P_1 vs. P_2	*Valence*	*Volume*	*Variance*	P_1 vs. P_3	*Valence*	*Volume*	*Variance*
Valence	1			*Valence*	1		
Volume	.0450***	1		*Volume*	.0523***	1	
Variance	-.1283***	.3704***	1	*Variance*	-.1266***	.4041***	1

In addition, we visualize how many outcomes can be predicted by considering the distribution of principal components of the residuals. As shown in Table K2, the first two components have the largest eigenvalues, i.e., more than 1.0, and already explain at least 70 percent of the variation in the data.

Table K2. Principal Components

P_1 vs. P_2	*Comp1*	*Comp2*	*Explained*	P_1 vs. P_3	*Comp1*	*Comp2*	*Explained*
Valence	-.1602	.9434	95.00%	*Valence*	-.1318	.9516	95.69%
Volume	.6792	.3189	74.12%	*Volume*	.6863	.2896	75.11%
Variance	.7162	-.0914	71.65%	*Variance*	.7152	-.1026	73.27%

Online Appendix L. Content-Level Analysis

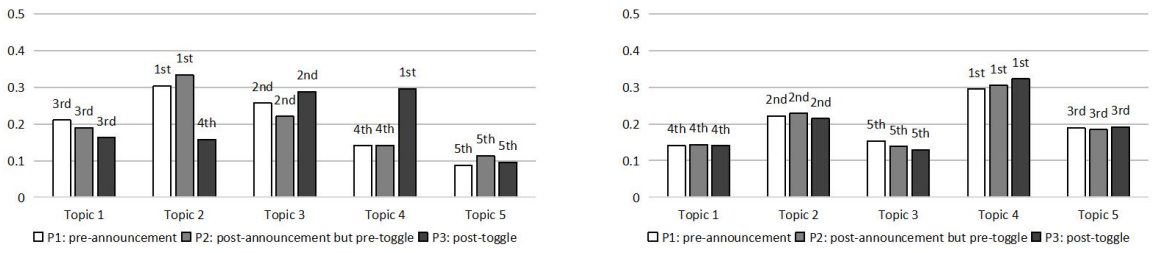
Figure L1. Model Diagnostics by Number of Topics

A. Semantic Coherence **B.** Exclusivity

Table L1. Five Topics and Their Top Words

Topic Label	*Top Words*	*Percentage*
Product functionality	'battery', 'feature', 'usb', 'camera', 'screen', 'memory', 'resolution', 'size', 'sound', 'technology'	15.0%
Product delivery	'shipping', 'delivery', 'fast', 'quick', 'ship', 'speed', 'dispatch', 'overnight', 'shipment', 'condition'	22.6%
Product service	'support', 'service', 'help', 'response', 'refund', 'return', 'warranty','question', 'issue', 'customer'	15.5%
Product experience	'shopping', 'easy', 'interface', 'apps', 'experience', 'issue', 'player', 'touch', 'video', 'music'	29.2%
Product quality	'quality', 'durable', 'reliable', 'material', 'construction', 'performance', 'build', 'defect', 'value', 'product'	17.7%

Figure L2. Differences in Topic Proportions



A. Negative reviews of slow-delivery products **B.** Positive reviews of slow-delivery products

Notes. The y-axis scale of Panels A and B represents the topic proportion. The data label on each bar indicates the ranking of the corresponding topic.