Appendices

A. Additional Information about the Pop-Up Store Event

Figure 1 Text Message in the Original Language (Chinese)

10.12-10.19 潮流牛仔节，中大银泰等你来

B. Results about Consumers Whose Alibaba Account Was Associated with an IOS device

Among iOS device users, the percentage of consumers whose presence at the pop-up store was detected increased from 0.37% in the control condition to 0.63% in the treatment condition (p-value for the two-proportion test < 0.0001). The table below presents the ITT effect of our messaging treatment and the LATE of pop-up store visits on participating and non-participating retailers among iOS device users.

Though Android penetration is generally much higher than iOS in China[1], it is not the case for Hangzhou—the city in which our experiment took place, at least for people who shop on Alibaba. According to a report generated by CBNdata (a leading business research firm in China) based on Alibaba consumers' data[2], around 50% of Alibaba consumers in Hangzhou had an iOS device in 2015 (page 10 of the report). In fact, the report indicates that Hangzhou has the third highest iOS penetration among Chinese cities (page 10 of the report). This may be because Hangzhou is one of the most developed cities in China and a larger portion of Hangzhou residents can afford iPhones than residents in an average Chinese city. Consistent with the report, among customers in our experiment whose device type was made available to us, about half of them had an iOS device (N = 202,880 as reported here) and another half had an Android device (N = 214,370 as reported in the manuscript). We unfortunately did not have information about the device type of some consumers in the experiment. That is why we report results on customers whose Alibaba accounts were known to be associated with an Android phone (in the manuscript) or an iOS device (Online Appendix B).

C. Robustness Checks and Additional Analyses

For the randomization check reported in Table 1 in the paper, we conduct an F-test on the joint hypothesis for no difference between treatment and control conditions in the means of all four variables (i.e., daily expenditure at participating retailers' Tmall stores in the pre-event period,

[1] https://www.statista.com/statistics/262176/market-share-held-by-mobile-operating-systems-in-china/

[2] https://cbndata.com/report/8

Table 1 The Intent-to-Treat Effect of the Messaging Intervention and the Local Average Treatment Effect of Visiting the Pop-Up Store on Consumers' Online Expenditure among IOS Device Users

Panel A: The Intent-to-Treat Effect of the Messaging Intervention		
Dependent Variable	Average Daily Expenditure (RMB)	
	At Participating Retailers: Model 1	At Non-participating Retailers: Model 2
Treatment	0.015**	0.432****
	(0.007)	(0.104)
Relative Effect Size	21.15%	8.23%
Observations	202,880	202,880

Panel B: The Local Average Treatment Effect of Visiting the Pop-Up Store		
Dependent Variable	Average Daily Expenditure (RMB)	
	At Participating Retailers: Model 1	At Non-participating Retailers: Model 2
Visit	5.386**	160.472****
	(2.709)	(42.969)
Observations	202,880	202,880

Note: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$, $****p < 0.001$. Robust standard errors are in parentheses.

daily expenditure at non-participating retailers' Tmall stores in the pre-event period, expenditure at participating retailers' Tmall stores on the message day, and expenditure at non-participating retailers' Tmall stores on the message day). To perform this test, we construct a dataset that has four observations per consumer, with each observation corresponding to one of the four variables. We use the following specification:

$$\begin{aligned}\text{Outcome Variable}_{ij} =& \delta_1 Treatment_i * \text{Participating retailer pre-event indicator}_{ij} + \\ & \delta_2 Treatment_i * \text{Non-participating retailer pre-event indicator}_{ij} + \\ & \delta_3 Treatment_i * \text{Participating retailer message day indicator}_{ij} + \\ & \delta_4 Treatment_i * \text{Non-participating retailer message day indicator}_{ij} + V_j + \epsilon_{ij},\end{aligned}$$

where $\text{Outcome Variable}_{ij}$ refers to the value of each of the four variables j for consumer i, Treatment_i is a binary variable equaling one if consumer i received the message, and V_j represents fixed effects for variables. $\text{Participating retailer pre-event indicator}_{ij}$ is a binary variable equaling one if $\text{Outcome Variable}_{ij}$ is the value of customer i's daily expenditure at participating retailers' Tmall stores in the pre-event period. $\text{Non-participating retailer pre-event indicator}_{ij}$ is a binary variable equaling one if $\text{Outcome Variable}_{ij}$ is the value of customer i's daily expenditure at non-participating retailers' Tmall stores in the pre-event period. $\text{Participating retailer message day indicator}_{ij}$ is a binary variable equaling one if $\text{Outcome Variable}_{ij}$ is the value of customer i's daily expenditure at participating retailers' Tmall stores on the message day. $\text{Non-participating retailer message day indicator}_{ij}$ is a binary variable equaling one if $\text{Outcome Variable}_{ij}$ is the value of customer i's daily expenditure at non-participating retailers' Tmall stores on the message day. Standard errors are clustered at the customer level since each customer has four non-independent observations.

Table 2 Robustness Checks

Panel A: The Difference-in-Differences Approach		
Dependent Variable	Average Daily Expenditure (RMB)	
	At Participating Retailers: Model 1	At Non-participating Retailers: Model 2
Treatment	0.0005	0.050
	(0.002)	(0.033)
Post	0.011****	1.267***
	(0.002)	(0.058)
TreatmentXPost	0.006**	0.326****
	(0.003)	(0.066)
Relative Effect Size	36.77%	14.89%
Observations	428,740	428,740

Panel B: Excluding Average Daily Expenditure at Non-participating Retailers That Was above the 99th Percentile		
Dependent Variable	Average Daily Expenditure (RMB) at Non-participating Sellers	
	The Intent-to-Treat Effect of the Messaging Intervention: Model 1	The Local Average Treatment Effect of Visiting the Pop-Up Store: Model 2
Treatment	0.218****	
	(0.016)	
Visit		68.879****
		(8.716)
Relative Effect Size	17.71%	
Observations	210,462	210,462

Note: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$, $****p < 0.001$. For Panel A, standard errors were clustered at the consumer level. For Panel B, robust standard errors are reported.

The interested estimates are δ_1, δ_2, δ_3, and δ_4, which represent the difference between treatment and control conditions in the means of the four variables, respectively. An F-test on the joint significance of δ_1, δ_2, δ_3, and δ_4 shows that these four coefficients are jointly not significant (F = 1.27, p = 0.28). This means that the differences between treatment and control conditions in the means of all four variables in Table 1 are jointly statistically insignificant.

As one robustness check for the ITT specification reported in our paper, we conduct a difference-in-differences analysis. We calculate the average expenditure per customer per day in the three-week pre-event period vs. six-week post-event period. We perform our DID analysis with two observations per customer: one observation for the pre-event period and one observation for the post-event period. We cluster standard errors at the customer level. As another robustness check, we remove outliers by excluding customers whose average daily expenditure on non-participating retailers was above the 99th percentile.

We also conduct an exploratory analysis to see whether our observed effects were primarily driven by consumers' willingness to purchase any jeans products online or their expenditure (conditional on them already having decided to buy jeans products online). Specifically, we compare treatment and control consumers on (1) whether they purchased any jeans products on Tmall during the post-event period and (2) how much they spent on jeans products on Tmall (if they made at least one purchase). For participating retailers, we find that our messaging treatment significantly increased the likelihood of consumers purchasing anything from these retailers' Tmall stores by 70% ($p < 0.0001$) but did not significantly change how much consumers spent, conditional on them having already decided to buy ($p = 0.40$). For non-participating retailers, we find that our messaging treatment significantly increased the likelihood of consumers purchasing anything from these retailers' Tmall stores by 13% ($p < 0.0001$) and marginally significantly increased how much consumers spent by 5%, conditional on them having already decided to buy ($p = 0.08$).

D. Heterogeneous Treatment Effect

In the main body of this paper, we separately apply the ITT specification to prospective vs. existing customers of participating retailers. Here, we run regressions that combine these two types of consumers and include an interaction between customer type and our messaging treatment to predict consumers' average daily expenditure in the post-event period. *Prospective Customer* is a binary variable that equals one for customers who did not purchase from any participating retailers prior to the pop-up store event and zero otherwise. Consistent with the split-sample analysis we present in the paper, the intent-to-treat effect of our messaging treatment is only statistically significant for prospective consumers but insignificant for existing consumers, further indicating that the insignificant results for existing customers reported in the paper are not simply due to a lack of statistical power.

Also, we explore whether consumers' spending power on the platform could produce a heterogeneous treatment effect. Based on a consumer's monthly expenditure on Alibaba in the past six months, Alibaba assigns a score, ranging from one to five, to the consumer, with five indicating the highest spending power and one indicating the lowest. We run regressions that include our messaging treatment, customers' spending power, and their interaction as predictors of consumers' average daily expenditure in the post-event period. We find that our messaging treatment had a larger effect on both participating and non-participating retailers among consumers whom Alibaba classifies as having a higher spending power on the platform than among those having a lower spending power.

Table 3 Heterogeneous Treatment Effect

Panel A: Prospective vs. Existing Customers

Dependent Variable	Average Daily Expenditure (RMB)	
	At Participating Retailers: Model 1	At Non-participating Retailers: Model 2
Treatment	-0.006	0.394
	(0.027)	(0.257)
Prospective Customer	-0.104****	-3.445****
	(0.007)	(0.187)
TreatmentXProspective Customer	0.014	-0.031
	(0.010)	(0.266)
Observations	203,015	203,015

Panel B: Customers' Spending Power on Alibaba

Dependent Variable	Average Daily Expenditure (RMB)	
	At Participating Retailers: Model 1	At Non-participating Retailers: Model 2
Treatment	0.008***	0.410
	(0.003)	(0.073)
Spending Power	0.007****	0.802****
	(0.002)	(0.047)
TreatmentXSpending Power	0.006**	0.816****
	(0.003)	(0.075)
Observations	182,170	182,170

Note: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$, $****p < 0.001$. Robust standard errors are reported in parentheses. Customers with missing data about their type (N = 11,355) are excluded from the models in Panel A. Customers with missing data about their spending power (N = 32,200) are excluded from the models in Panel B.