

# Uncertainty Reduction vs Reciprocity: Understanding the Effect of a Platform-Initiated Reviewer Incentive Program on Regular Ratings

## Supplementary Online Appendices

### Appendix A. Supplemental Materials for the Main Analyses

#### A.1 Steps of Propensity Score Matching

We match each treated reviewer with a control counterpart who has a similar probability of receiving platform-initiated free samples based on variables that can affect the reviewer's (i) probability of getting invited to join the *Vine* program and (ii) willingness to be contacted. Amazon's official description of the *Vine* program specifies that the platform determines which reviewers to invite based on the overall review experience, review experience in different product categories, review helpfulness, and recent number of reviews.<sup>1</sup> We check each reviewer's profile webpage to collect data on their review experience and contact information. We assume that reviewers are more willing to be contacted if they voluntarily disclose their contact, description, and location information, so we collect data on the availability of this information. Table A.1 summarizes the matching variables.

We calculate each reviewer's propensity score based on the observations in August 2007 (when the *Vine* program was launched), and we match each treated reviewer with the control reviewer who has the most similar propensity score. Table A.2 shows that after matching, there is no significant difference between the two groups across most of the matching variables. We also compare the standardized percentage bias of the variables between the treated and control group, before and after matching (Rosenbaum and Rubin 1985). We observe that the biases between the treated and control groups are

---

<sup>1</sup> The announcement specifies that the *Vine* program invites reviewers based on "the helpfulness of their reviews as judged by all other customers and by their demonstrated interest in the types of products that are featured in the program." Moreover, it explains that the invitations are also based on reviewer rank, which is determined by "the overall helpfulness of all their reviews, factoring in the number of reviews they have written. More weight is given to recent reviews." See the detailed announcement at <https://www.amazon.com/gp/vine/help> (accessed June 7<sup>th</sup>, 2023).

greatly reduced (by 28–95%) after matching. The balance statistics indicate that the matching strategy is effective at balancing the observed characteristics of the treated and control reviewers.

**Table A.1 Variables Used to Match Reviewers in the Treatment and Control Groups**

Variable Category	Variable Description	Variables
The reviewer’s general experience	The log-transformed total number of months since the reviewer joined Amazon	Log (Review Tenure + 1)
The reviewer’s reviewed products	The proportion of the reviewer’s total reviews that are for products in the focal category since joining Amazon	Separate variables for the following categories: Books, Movies & TV, CDs & Vinyl, Electronics, Home & Kitchen, Grocery & Gourmet Food, and Health & Personal Care
The reviewer’s recent number of reviews	The reviewer’s monthly average number of reviews in the past five months; we include separate variables for all purchases versus for verified purchases	Average number of regular reviews in the past five months Average number of verified-purchase reviews in the past five months
The reviewer’s review quality	The average helpfulness ratio of the reviews produced since joining Amazon	Average helpfulness ratio
The reviewer’s willingness to be contacted	Whether the reviewer provides three types of personal information	Three binary variables for contact, description, and location information

**Table A.2 T-test of the Variables Before and After Matching**

Variable	Sample	Mean		T-test		Std. bias	
		Treated	Control	t	$p >  t $	% Std.  bias	% Reduced
Log (Reviewer Tenure + 1)	Unmatched	3.91	3.66	10.83	0.00	37.0	85.8
	Matched	3.91	3.94	-1.59	0.11	5.3	
Category: Books	Unmatched	0.31	0.28	2.76	0.00	9.4	40.8
	Matched	0.31	0.33	-1.65	0.10	5.6	
Category: Electronics	Unmatched	0.08	0.10	-2.87	0.00	9.8	32.5
	Matched	0.08	0.07	2.19	0.03	6.6	
Category: Health & Personal Care	Unmatched	0.01	0.01	-1.72	0.09	5.9	27.7
	Matched	0.01	0.01	2.55	0.01	4.3	
Category: CDs	Unmatched	0.08	0.11	-4.32	0.00	14.8	79.2
	Matched	0.08	0.08	1.09	0.27	3.1	
Category: Movies & TV	Unmatched	0.11	0.11	1.02	0.31	3.5	26.9
	Matched	0.11	0.12	-0.74	0.46	2.5	
Category: Home & Kitchen	Unmatched	0.07	0.06	1.05	0.29	3.6	82.1
	Matched	0.07	0.07	0.18	0.86	0.6	
Category: Grocery & Gourmet Food	Unmatched	0.01	0.00	4.61	0.00	15.7	81.4
	Matched	0.01	0.01	0.63	0.53	2.9	
Average Number of Regular Reviews in the Past 5 Months	Unmatched	2.61	1.48	5.31	0.00	18.0	78.6
	Matched	2.61	2.36	1.04	0.30	3.9	
Average Number of Verified-Purchase Reviews in the Past 5 Months	Unmatched	0.65	0.38	6.33	0.00	21.6	91.1
	Matched	0.65	0.68	-0.44	0.69	1.9	
Average Helpfulness Ratio	Unmatched	0.84	0.83	3.68	0.00	12.7	61.0
	Matched	0.84	0.84	1.67	0.10	4.9	
Contact Information	Unmatched	0.28	0.13	11.56	0.00	39.3	78.0
	Matched	0.28	0.25	2.24	0.03	8.7	
Description information	Unmatched	0.57	0.35	12.79	0.00	43.6	94.9
	Matched	0.57	0.58	-0.63	0.53	2.2	
Location information	Unmatched	0.90	0.88	1.98	0.05	6.8	57.6
	Matched	0.90	0.91	-0.88	0.38	2.9	

## A.2 Robustness Checks

We perform a series of robustness checks for our DID design: (i) classic DID and falsification tests, (ii) alternative matching approaches, and (iii) a relative time model to verify the parallel trends assumption.

We find that our main result is consistent in all robustness checks.

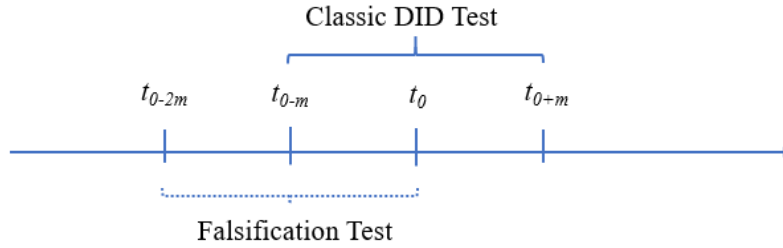
### A.2.1 Classic DID and Falsification Tests

The treated reviewers joined the *Vine* program at different times, so in our main analyses, we accommodated the staggered adoption with a two-way fixed effects DID model and PSM (e.g., Xu et al. 2016). Some existing studies also use a classic DID approach (e.g., Jung et al. 2019, Kumar et al. 2019, Pu 2022) in which researchers construct a panel dataset with a fixed number of observations (e.g.,  $m$ ) for each reviewer before and after the treatment. To check the robustness of our results, we conduct a classic DID with the same propensity score matching specification used in our model. For our model, we suppose that each treated reviewer started to receive free samples in month  $t_0$ . We include observations from  $t_{0-m}$  to  $t_{0+m}$  for the treated reviewer and matched control reviewer, and we re-estimate Equation (1).

Following Jung et al. (2019), we also conduct a falsification test to re-estimate the same DID model with all the same reviewer samples, model parameters, estimation model, propensity score matching equation, and caliper size except that we now shift  $VineMonth_{it}$ , the binary variable indicating the month in which reviewer  $i$  first received a *Vine* product, forward by  $m$  months. That is,  $VineMonth_{it}$  (described in our DID model) equals 1 from  $t_{0-m} \sim t_0$  (a period before reviewer  $i$ 's actual treatment time); Naturally, due to the time shift,  $VineMonth_{it}$  will equal 0 during  $t_{0-2m} \sim t_{0-m}$  and the total time periods for the falsification test will be from  $t_{0-2m}$  to  $t_0$ . The falsification test shifts the “treatment period” to an arbitrary time before reviewers actually joined the program.

In the falsification test,  $VineMonth_{it}$  arbitrarily points to the periods before the actual first *Vine* product receiving time, not after. The reviewers have not actually begun to receive *Vine* products yet. Therefore, the same model that we use to show the statistically significant impact should produce a non-significant estimation for the DID term because no treatment effect has actually occurred during the pre-treatment period. Figure A.1 illustrates the sampled time periods for the classic DID and falsification tests.

We check the classic DID and falsification test with  $m = 2, 3,$  and  $4,$  and the results are qualitatively consistent with one another. Table A.3 reports the results. For the classic DID test without  $VineMonth_{it}$  being shifted, the estimated treatment coefficients are positive and significant (Models 1, 3, and 5), consistent with the result in Table 2. As expected, for the falsification test where  $VineMonth_{it}$  is shifted  $m$  months before the actual program joining time, the estimated treatment coefficient is insignificant (Models 2, 4, and 6).



**Figure A.1 Time Periods for the Classic DID and Falsification Tests**

**Table A.3 Results of the Classic DID and Falsification Tests**

Specifications	Panel A ( $m = 2$ )		Panel B ( $m = 3$ )		Panel C ( $m = 4$ )	
	DID	Falsification	DID	Falsification	DID	Falsification
	(1)	(2)	(3)	(4)	(5)	(6)
$VineReviewer (\beta_1)$	-	-	-	-	-	-
$VineMonth (\beta_2)$	-0.129** (0.060)	-0.062 (0.064)	-0.048 (0.045)	0.061 (0.051)	-0.049 (0.038)	0.010 (0.043)
$VineReviewer \times$ $VineMonth (\beta_3)$	0.126** (0.051)	-0.016 (0.054)	0.094** (0.039)	-0.037 (0.042)	0.114*** (0.034)	-0.052 (0.037)
Control Variables	YES	YES	YES	YES	YES	YES
Time Fixed	YES	YES	YES	YES	YES	YES
Reviewer Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.628	0.657	0.550	0.560	0.490	0.507
Observations	6,611	5,945	9,797	8,915	12,963	11,566

Notes. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

## A.2.2 Alternative Matching Approaches

We employ alternative matching approaches and re-estimate Equation (1); the estimation results, presented in Table A.4, are consistent with those in Table 2, strengthening the robustness of our main finding.

We repeat the PSM with alternative settings: (i) In the main model, we use the matching variables in Table A.1 because they describe the official standards for program invitations; we also reconstruct the reviewer sample with more matching variables including the average regular rating as well as average rating difference (between the reviewer’s rating and the reviewed product’s average rating) in the pre-treatment period (Model 1). (ii) In the main model, we use one-to-one matching with replacement; we also verify our results with one-to-one matching without replacement and nearest-three-neighbors matching (Models 2 and 3). (iii) We employ a look-ahead matching strategy (e.g., Xu et al. 2016, Jung et al. 2019; Model 4), reconstructing the control group with only the reviewers who are treated in a future period (i.e., they will receive *Vine* products eventually but have not as of month  $t$ ). Specifically, we construct the control group with reviewers who first received *Vine* products after December 2009, and we match them with reviewers who began receiving *Vine* products before December 2009. This method can account for unobserved characteristics that affect whether reviewers get treated, so a significant result with the look-ahead sample strengthens our confidence that our effect is not attributable to selection bias arising from unobservable differences between the treated and control reviewers (Model 4).

Next, we repeat the PSM with different matching time points (Model 5). In our main analysis, we matched reviewers based on the observations in August 2007 (when *Vine* launched); now, we adopt a dynamic matching strategy in which we calculate reviewers’ propensity scores and conduct matching on a month-by-month basis. The same reviewer may match with different control reviewers in different months. With this matching strategy, the matching time point accounts for the fact that the treated reviewers start to receive *Vine* products in different months. Also, one treated reviewer can be matched with different control reviewer over time.

Finally, we validate our result with an alternative matching algorithm, Coarsened Exact Matching (CEM; Model 6). While PSM matches treated and control reviewers with the probability of getting free samples with the matching variables, CEM creates bins (i.e., ranges) for each matching variable; samples can be marked as matched if they have sufficiently close values to be assigned to the same bins (e.g., Iacus et al. 2012, Kumar et al. 2019). CEM yields a dataset of 437 reviewers, of whom 194 participated in

the *Vine* program in our study period. The multivariate imbalance tests show that the balanced level  $\mathcal{L}_1$  of the dataset improves from 0.947 before matching to 0.370 after matching.

**Table A.4 Results with Alternative Matching Approaches**

Specifications	More Matching Variables	One-to-One (w/o Replace.)	Nearest Three Neighbors	Look-Ahead	Dynamic Matching	CEM
	(1)	(2)	(3)	(4)	(5)	(6)
<i>VineReviewer</i> ( $\beta_1$ )	-	-	-	-	-	-
<i>VineMonth</i> ( $\beta_2$ )	-0.004 (0.016)	-0.014 (0.022)	-0.006 (0.011)	-0.026 (0.018)	-0.001* (0.014)	-0.063* (0.038)
<i>VineReviewer</i> × <i>VineMonth</i> ( $\beta_3$ )	0.054*** (0.018)	0.087*** (0.023)	0.080*** (0.014)	0.136*** (0.019)	0.065*** (0.020)	0.107*** (0.019)
Control Variables	YES	YES	YES	YES	YES	YES
Time Fixed	YES	YES	YES	YES	YES	YES
Reviewer Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.321	0.300	0.341	0.277	0.361	0.388
Observations	52,453	37,030	98,896	48,505	56,756	8,371

Notes. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

### A.2.3 A Relative Time Model

If the treated reviewers gradually increased their regular ratings (for unknown reasons) during the pre-treatment period, but the control reviewers do not, then the treatment effect identified in our DID estimation would be because of the differences in the pre-treatment trends rather than the *Vine* products. We found some evidence against this alternative explanation in the falsification test, and here, we adopt a relative time model to test the parallel trends assumption of the DID model.

Following Jung et al. (2019), we check for a difference in pre-treatment trends with a two-way fixed effects model that includes dummy variables for the time periods before and after treatment:

$$AvgRegRating_{it} = \alpha_0 + \sum_{j=T-5}^{T+3} \alpha_j (s_i \varphi) + X_{it} + \tau_t + \delta_i + \epsilon_{it} \quad (6)$$

where  $AvgRegRating_{it}$  is the dependent variable (reviewer  $i$ 's average regular rating in month  $t$ ),  $\tau_t$  is time fixed effects,  $\delta_i$  is reviewer fixed effects,  $s_i$  is a binary variable indicating whether reviewer  $i$  received free samples from the program in our time period, and  $\varphi$  is the relative time dummy representing the chronological distance between month  $t$  and reviewer  $i$ 's treatment time. The relative time periods range

from -5 to +3 (i.e., from five months before to three months after reviewer  $i$  received the first free sample). To avoid a dummy variable trap, we follow Greenwood and Watta (2017) and Jung et al. (2019) and omit the dummy variable for the month before reviewer  $i$ 's treatment time.

Table A.5 presents the results of the relative time analysis. None of the estimated coefficients of the pre-treatment relative time dummies are significant (i.e., for  $t - 5$ ,  $t - 4$ ,  $t - 3$ , and  $t - 2$ ), indicating pre-treatment homogeneity between the treated and control reviewers. Moreover, all of the estimated coefficients of the post-treatment relative time dummies are significantly positive, consistent with the estimations in Table 2.

**Table A.5 Relative Time Model**

Dep. Var.	<i>AvgRegRating</i>
	(1)
t - 5	-0.053 (0.041)
t - 4	-0.010 (0.041)
t - 3	-0.028 (0.037)
t - 2	-0.035 (0.039)
t - 1	Omitted
t + 0	0.096*** (0.034)
t + 1	0.089*** (0.034)
t + 2	0.059* (0.035)
t + 3	0.075** (0.036)
Control Variables	YES
Time Fixed	YES
Reviewer Fixed	YES
R-squared	0.437
Observations	16,029

Notes. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at user-month level. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

### A.2.4 Additional Control Variables that Describe the Regularly-Reviewed Products

In the main model, we included control variables related to product categories (represented by the variable  $X_{it}$ ), but other attributes of the regularly-reviewed products might affect reviewers' regular ratings. For example, perhaps reviewers tend to purchase products that are higher-quality or more popular after beginning to receive *Vine* products. We generate two sets of additional control variables and consider this alternative explanation.

We first include two control variables that can proxy product quality, namely the average current rating (i.e.,  $AvgProductRating_{it}$ ) and the log-transformed average price (i.e.,  $AvgProductPrice_{it}$ ) of  $i$ 's reviewed products in month  $t$  (De Langhe et al. 2016). Table A.6, Model 1 presents the result, and the estimation of the DID interaction term (i.e.,  $\beta_3$ ) is consistent with that in Table 2. We then use the average sales rank to proxy the popularity of the product (Hu et al. 2011) and include the log-transformed average sales rank (i.e.,  $AvgProductSalesRank_{it}$ ) of  $i$ 's reviewed products in month  $t$  in Equation (1). A higher sales rank indicates a lower level of popularity. Table A.6, Model 2 presents the result, which is consistent with that shown in Table 2.

**Table A.6 Estimations with More Control Variables**

Dep. Var.	<i>AvgRegRating</i>	
	(1)	(2)
<i>VineReviewer</i> ( $\beta_1$ )	-	-
<i>VineMonth</i> ( $\beta_2$ )	0.001 (0.015)	-0.020 (0.017)
<i>VineReviewer</i> × <i>VineMonth</i> ( $\beta_3$ )	0.050*** (0.016)	0.075*** (0.018)
<i>AvgProductRating</i>	0.683*** (0.014)	-
<i>AvgProductPrice</i>	0.017** (0.007)	-
<i>AvgProductSalesRank</i>	-	-0.006* (0.003)
Control Variables	YES	YES
Time Fixed	YES	YES
Reviewer Fixed	YES	YES
R-squared	0.468	0.343
Observations	41,044	48,698

Notes. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. There are fewer observations in this table than in Table 2 (54,668 observations) because of missing values for  $AvgProductRating_{it}$ ,  $AvgProductPrice_{it}$ , and  $AvgProductSalesRank_{it}$ . \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

## References

- De Langhe, B., Fernbach, P. M., & Lichtenstein, D. R. (2016). Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings. *Journal of Consumer Research*, 42(6), 817-833.
- Greenwood, B. N., & Wattal, S. (2017). Show Me the Way to Go Home: An Empirical Investigation of Ride-Sharing and Alcohol Related Motor Vehicle Fatalities. *MIS Quarterly*, 41(1), 163-187.
- Hu, N., Bose, I., Gao, Y., & Liu, L. (2011). Manipulation in Digital Word-of-Mouth: A Reality Check for Book Reviews. *Decision Support Systems*, 50(3), 627-635.
- Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20(1), 1-24.
- Jung, J., Bapna, R., Ramaprasad, J., & Umyarov, A. (2019). Love Unshackled: Identifying the Effect of Mobile App Adoption in Online Dating. *MIS Quarterly*, 43(1), 57-72.
- Kumar, A., Mehra, A., & Kumar, S. (2019). Why Do Stores Drive Online Sales? Evidence of Underlying Mechanisms from a Multichannel Retailer. *Information Systems Research*, 30(1), 319-338.
- Pu, J. (2022). Do Online Communities Improve Job Performance in the Geographically Dispersed Organization?. *Production and Operations Management*, 31(2), 403-421.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician*, 39(1), 33-38.
- Xu, K., Chan, J., Ghose, A., & Han, S. P. (2017). Battle of the Channels: The Impact of Tablets on Digital Commerce. *Management Science*, 63(5), 1469-1492.

## Appendix B. Supplemental Materials for the Analyses in Section 6.2

### B.1 Summary Statistics

In Table B.1, we report the summary statistics for the variables in all analyses in Section 6.2. In the first column, we indicate the corresponding analysis, models, and tables in the main text.

**Table B.1 Summary Statistics for the Variables in Section 6.2**

Analyses	Variable	Mean	Std. Dev.	N
Overall Category Analysis (Table 5, Model 1)	$VineReviewer_i$	0.500	0.500	4,809,600
	$VineMonth_{it}$	0.424	0.494	4,809,600
	$AvgRegRatingCa_{ikt}$	4.182	1.010	85,496
<i>Vine</i> Categories (Table 5, Model 2)	$VineReviewer_i$	0.500	0.500	589,392
	$VineMonth_{it}$	0.489	0.500	589,392
	$AvgRegRatingCa_{ikt}$	4.200	0.946	50,078
Non- <i>Vine</i> Categories (Table 5, Model 3)	$VineReviewer_i$	0.500	0.500	4,220,208
	$VineMonth_{it}$	0.415	0.493	4,220,208
	$AvgRegRatingCa_{ikt}$	4.156	1.094	35,418
Vine Products from Different Categories (Table 6, Model 1)	$VineReviewer_i$	0.710	0.454	34,269
	$VineMonth_{it}$	0.564	0.496	34,269
	$AvgRegRatingCa_{ikt}$	4.138	1.155	34,269
	$VineProductsFocal_{ikt}$	0.143	0.428	34,269
	$VineProductsNonFocal_{ikt}$	0.974	1.380	34,269
Vine Products from Different Subcategories (Table 7, Model 1)	$VineReviewer_i$	0.720	0.449	40,212
	$VineMonth_{it}$	0.572	0.494	40,212
	$AvgRegRatingSubCa$	4.155	1.160	40,212
	$VineProductsSubFocal$	0.050	0.227	40,212
	$VineProductsNonSubFocal$	1.028	1.415	40,212
Vine Products from Different Sub- subcategories (Table 7, Model 2)	$VineReviewer_i$	0.723	0.448	44,322
	$VineMonth_{it}$	0.576	0.494	44,322
	$AvgRegRatingSubSubCa$	4.170	1.162	44,322
	$VineProductsSubSubFocal$	0.050	0.227	44,322
	$VineProductsNonSubSubFocal$	1.046	1.426	44,322

Notes. There are fewer observations for the average rating than for other variables because not every reviewer generates a regular review in every category, every month. When conducting analyses in Tables 6 and 7, we focus on the observations where the average rating is not missing to simplify the data construction process. The subscripts for some variables in Table 7 are missing for ease of exposition.

### B.2 Comparing the Effect in *Vine* and Non-*Vine* Categories

In Section 6.2, we tested whether the effect of *Vine* products on regular ratings is more salient when the purchased product falls in the same category as *Vine* products previously received by the reviewer. In this

extension, we conduct a more formal test for the same comparison. We generate a binary variable,  $WithVineProduct_{ik}$ , that equals 1 if reviewer  $i$  previously received free sample(s) from category  $k$ , and we test the following model with a three-way interaction term. Table B.2 presents the result. The estimated coefficient of the three-way interaction term is positive and significant, confirming that the effect of *Vine* products is stronger when reviewers generate regular ratings for products in the same category as (and presumably sharing more attributes with) previously-received *Vine* products.

$$\begin{aligned}
 AvgRegRatingCa_{ikt} = & \beta_0 + \beta_1 VineReviewer_i + \beta_2 VineMonth_{it} + \beta_3 WithVineProduct_{ik} + \\
 & \beta_4 VineReviewer_i \times VineMonth_{it} + \beta_5 VineMonth_{it} \times WithVineProduct_{ik} + \\
 & \beta_6 VineReviewer_i \times WithVineProduct_{ik} + \beta_7 VineReviewer_i \times VineMonth_{it} \times \\
 & WithVineProduct_{ik} + X'_{it} + \tau_t + \delta_i + \rho_k + \epsilon_{it}
 \end{aligned}$$

**Table B.2 The Effect of *Vine* Products on Regular Ratings in *Vine* vs. Non-*Vine* Categories**

Specifications	Reviewers' <i>Vine</i> and Non- <i>Vine</i> Categories
	(1)
<i>VineReviewer</i>	-
<i>VineMonth</i>	0.016 (0.022)
<i>WithVineProduct</i>	-0.015 (0.024)
<i>VineReviewer</i> × <i>VineMonth</i>	0.040 (0.026)
<i>VineMonth</i> × <i>WithVineProduct</i>	-0.037 (0.025)
<i>VineReviewer</i> × <i>WithVineProduct</i>	-0.021 (0.027)
<i>VineReviewer</i> × <i>VineMonth</i> × <i>WithVineProduct</i>	0.052* (0.026)
Control Variables	YES
Time Fixed	YES
Reviewer Fixed	YES
R-squared	0.224
Observations	85,496

Notes. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the reviewer-category-month level. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

### B.3 Describing the Treatment at the Reviewer-Category-Month Level

In analyses at the reviewer-category-month level, we used  $VineMonth_{it}$  to represent the month in which reviewer  $i$  first received *Vine* products. In this extension, we verify our results with  $VineMonth_{ikt}$ , a binary variable representing whether reviewer  $i$  began to receive *Vine* products from category  $k$  in month  $t$ . Control reviewer  $i$  has the same value of  $VineMonth_{ikt}$  as the matched treated reviewer. We update Equation (2) as below. Table B.3 presents the results, which are consistent with those in Table 5, Models 1 and 2.

$$AvgRegRatingCa_{ikt} = \beta_0 + \beta_1 VineReviewer_i + \beta_2 VineMonth_{ikt} + \beta_3 VineReviewer_i \times VineMonth_{ikt} + X'_{it} + \tau_t + \delta_i + \rho_k + \epsilon_{it}$$

**Table B.3 Effect of *Vine* Products at the Reviewer-Category-Month Level**

Specifications	Overall	<i>Vine</i> Categories
	(1)	(2)
$VineReviewer_i$	-	-
$VineMonth_{ikt}$	-0.055*** (0.018)	-0.045** (0.019)
$VineReviewer_i \times VineMonth_{ikt}$	0.057*** (0.020)	0.058*** (0.021)
Control Variables	YES	YES
Time Fixed	YES	YES
Reviewer Fixed	YES	YES
R-squared	0.211	0.271
Observations	85,496	50,078

Notes. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the reviewer-category-month level. We cannot examine the specification in Table 5, Model 3, in this extension because  $VineMonth_{ikt}$  is always zero for reviewers' non-*Vine* categories. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

### B.4 Number of Cross-Category *Vine* Products at a Higher Level of Granularity

In the analysis at the reviewer-subcategory-month level, we can further differentiate the *Vine* products from non-focal subcategories into the ones from other subcategories and other categories. Specifically, we can differentiate *Vine* products with independent variables  $VineProductsSubFocal$ ,  $VineProductsSubOther$ ,  $VineProductsOther$ , which are the log-transformed numbers of *Vine* products

received by reviewer  $i$  in the focal subcategory, other subcategories, and other categories, respectively, before month  $t$ .

In the analysis at the reviewer–sub-subcategory–month level, we can further differentiate the *Vine* products from non-focal sub-subcategories into the ones from other sub-subcategories, other subcategories, and other categories. Specifically, we differentiate *Vine* products with independent variables  $VineProductsSubSubFocal$ ,  $VineProductsSubSubOther$ ,  $VineProductsSubOther$ ,  $VineProductsOther$ , which are the log-transformed numbers of *Vine* products received by reviewer  $i$  in the focal sub-subcategory, other sub-subcategories, other subcategories, and other categories, respectively, before month  $t$ . Table B.4 presents the results, which are consistent with those in Table 7.

**Table B.4 Number of Cross-Category *Vine* Products at a Higher Level of Granularity**

Specifications	Subcategory Level	Sub-Subcategory Level
	<i>AvgRegRatingSubCa</i>	<i>AvgRegRatingSubSubCa</i>
	(1)	(2)
<i>VineReviewer</i>	-	-
<i>VineMonth</i>	0.076*** (0.022)	0.078*** (0.022)
<i>VineProductsSubFocal</i>	0.046* (0.026)	-
<i>VineProductsSubSubFocal</i>	-	0.072** (0.033)
<i>VineProductsSubSubOther</i>	-	0.030 (0.028)
<i>VineProductsSubOther</i>	0.017 (0.017)	0.018 (0.015)
<i>VineProductsOther</i>	-0.003 (0.010)	-0.001 (0.010)
Control Variables	YES	YES
Time Fixed	YES	YES
Reviewer Fixed	YES	YES
R-squared	0.201	0.193
Observations	40,212	44,322

Notes. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the reviewer-subcategory-month level in Model 1 and at the reviewer–sub-subcategory–month level in Model 2. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .



## Appendix C. Supplemental Materials for the Randomized Experiment

### C.1 Figures of the Products and Webpages in the Randomized Experiment



Panel A. Durian Cake Sample in T1



Panel B. Toasted Edamame Sample in T2



Panel C. Strawberry Cake Sample in T3



Panel D. Original-Flavor Cookie Sample in T4

**Figure C.1 Free Samples in the Treatment Groups**



Durian Cookie  
曲格酥榴莲味

¥1.60

立即购买

Purchase Now



Sea Salt Cookie  
曲格酥海盐芝士味

¥1.60

立即购买

Purchase Now



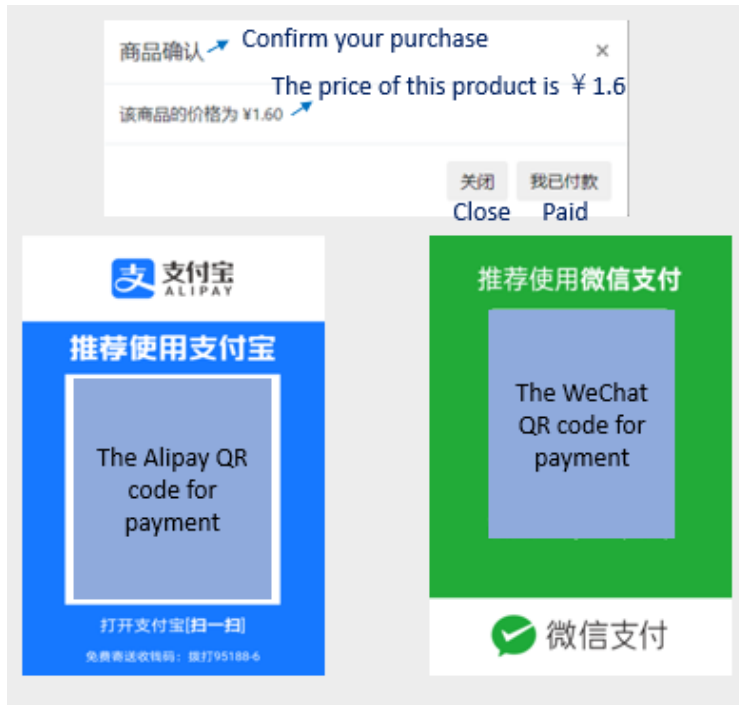
Egg Yolk Cookie  
曲格酥咸蛋黄味

¥1.60

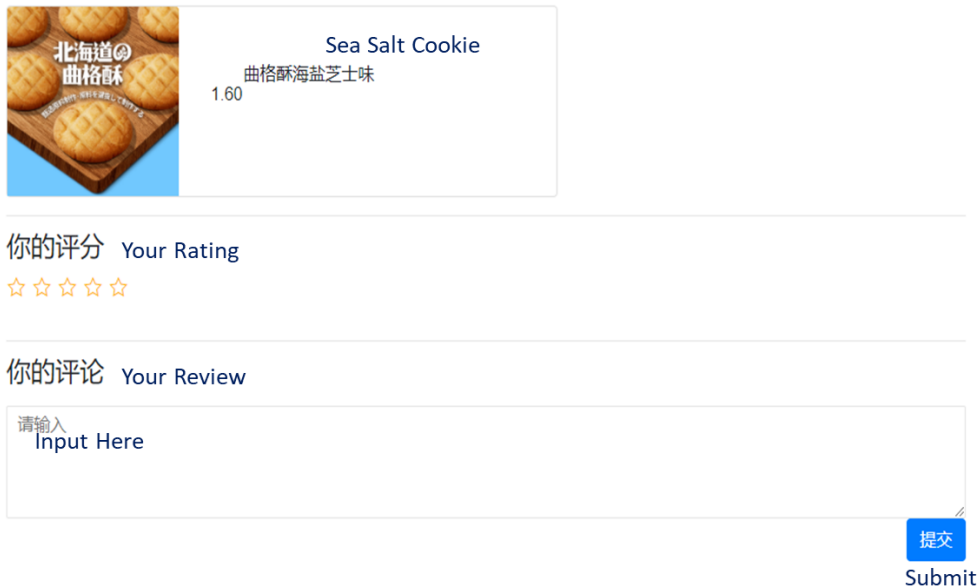
立即购买

Purchase Now

**Figure C.2 Product Options in the Purchase Stage**



**Figure C.3 Payment for the Purchased Product**



**Figure C.4 Review Page for the Purchased Product**

## C.2 Survey and Summary Statistics

**Table C.1 Survey Instrument**

Construct	Measurement Items
<b>Product Uncertainty</b>	<ul style="list-style-type: none"> <li>• I was not certain the product would match my requirements.</li> <li>• I was certain that the product would match my tastes.</li> <li>• I was sure that the product would fit my preferences.</li> <li>• I was not certain the product with these characteristics was what I was looking for.</li> <li>• I feared that the product attributes (such as texture, style, flavor, or design; whatever applies) were not what I wanted.</li> </ul>
<b>Anticipation-Based Reciprocity</b>	<ul style="list-style-type: none"> <li>• I expected that I would receive some type of return from the platform if I provided a review for the purchased product.</li> <li>• When I provided a review for the purchased product, I expected that I would be invited by the platform to a future event.</li> <li>• I believed that my review would be appreciated by the platform, and thus, I would be invited by the platform again.</li> <li>• I believed that my review of the purchased product would be reciprocated by the platform in some way.</li> </ul>
<b>Gratitude-Based Reciprocity</b>	<ul style="list-style-type: none"> <li>• I wrote a review based on my gratitude for getting invited by the platform to this event.</li> <li>• I wrote a review for the platform because I owed it for the invitation to this event.</li> <li>• The platform is receiving my review as my payback for letting me participate in this event.</li> </ul>
<b>Other Questions</b>	
Demographics	Age, birth gender, and class
Filler question 1	How many years have you been purchasing products on the internet?
Filler question 2	How many years have you been using the internet?
<b>Screening Questions</b>	
What day is it today?	
If you have read this question carefully, please choose “highly disagree.”	

Notes. The experiment platform also records the time spent on the experiment for a sanity check.

**Table C.2 Summary Statistics**

Variables	Mean	Std. Dev.	Min	Max
<i>Rating</i>	4.006	0.873	1	5
<i>UncerLevel</i>	4.575	1.200	1	7
<i>RP_Antbased</i>	3.641	1.223	1	7
<i>RP_Grabased</i>	4.121	1.339	1	7
<i>Female</i>	0.600	0.490	0	1
<i>Age</i>	22.79	2.861	18	32
<i>Selected_1</i>	0.214	0.411	0	1
<i>Selected_2</i>	0.452	0.498	0	1
<i>Selected_3</i>	0.334	0.472	0	1

Notes. N = 485.

### C.3 More Regression Analyses for the Randomized Experiment

We conduct regression analyses to compare treatment group T1 with the control group. We update the econometric equations and compare participants in T1 and the control group. The equations below are written with *UncerLevel*, but we also test *RP\_Antbased* and *RP\_Grabased*. The independent variable in the equations below is  $T1_u$ , which is a binary variable indicating whether participant  $u$  is in T1 or the control group. Table C.3 presents the results, and all results are consistent with those in Table 8.

$$Rating_u = \gamma_0 + \gamma_1 T1_u + Control_u + \epsilon_u, \quad (C.1)$$

$$UncerLevel_u = \omega_0 + \omega_1 T1_u + Control_u + \epsilon_u, \quad (C.2)$$

$$Rating_u = \mu_0 + \mu_1 T1_u + \mu_2 UncerLevel_u + Control_u + \epsilon_u, \quad (C.3)$$

**Table C.3 Regression Results of the Randomized Experiment (T1 vs. Control)**

Dep. Var.	Main Effect	Mediation Test for <i>UncerLevel</i>		Mediation Test for <i>RP_Antbased</i>		Mediation Test for <i>RP_Grabased</i>	
	<i>Rating</i>	<i>UncerLevel</i>	<i>Rating</i>	<i>RP_Antbased</i>	<i>Rating</i>	<i>RP_Grabased</i>	<i>Rating</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>T1</i>	0.238** (0.120)	-0.378** (0.175)	0.188 (0.120)	0.243 (0.186)	0.214* (0.120)	0.055 (0.200)	0.235* (0.120)
<i>UncerLevel</i>	-	-	-0.131*** (0.049)	-	-	-	-
<i>RP_Antbased</i>	-	-	-	-	0.099** (0.046)	-	-
<i>RP_Grabased</i>	-	-	-	-	-	-	0.046 (0.044)
<i>Female</i>	-0.245** (0.119)	0.270 (0.173)	-0.210* (0.117)	-0.231 (0.184)	-0.222* (0.118)	-0.128 (0.197)	-0.239** (0.119)
<i>Age</i>	0.020 (0.020)	-0.022 (0.029)	0.017 (0.020)	-0.013 (0.031)	0.021 (0.020)	0.061* (0.033)	0.017 (0.020)
<i>Selected_2</i>	0.135 (0.169)	-0.052 (0.246)	0.128 (0.166)	0.142 (0.262)	0.121 (0.168)	0.398 (0.280)	0.116 (0.170)
<i>Selected_3</i>	0.037 (0.176)	-0.057 (0.257)	0.030 (0.174)	-0.175 (0.273)	0.055 (0.175)	0.050 (0.293)	0.035 (0.176)
R-squared	0.052	0.044	0.087	0.028	0.075	0.033	0.058
Observations	196	196	196	196	196	196	196

Notes. Robust standard errors are in parentheses. The unit of analyses is at the participant level. The estimations are consistent if we remove the variable *T1* in Models 3, 5, and 7. *Selected\_1* is missing because the inclusion of *Selected\_1*, *Selected\_2*, and *Selected\_3* leads to perfect collinearity. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .