How Do Product Recommendations Help Consumers Search Products?

Evidence from a Field Experiment

Online Appendix

Xiang (Shawn) Wan,[a] Anuj Kumar,[b] Xitong Li [c]

[a] Santa Clara University; [b] University of Florida; [c] HEC Paris

xwan@scu.edu; akumar1@ufl.edu; lix@hec.fr

Appendix A. Balance Check

To validate the randomness of our assignment, we check whether visitor characteristics of treated and control visitors (purchase funnels) are similar. Specifically, we examine whether the following variables related to visitors are statistically similar for treated and control visitors (purchase funnels): (i) new (vs. repeat) visitors; (ii) frequent (vs. infrequent) visitors, indicating whether the total visitor session number is above the 75th percentile value in the distribution of the number of sessions by a visitor; (iii) the way through which a visitor comes to a session, such as paid search, and organic search, and others. Table A1 reports the results, showing that these variables are statistically indistinguishable between the treated and control visitors (purchase funnels).

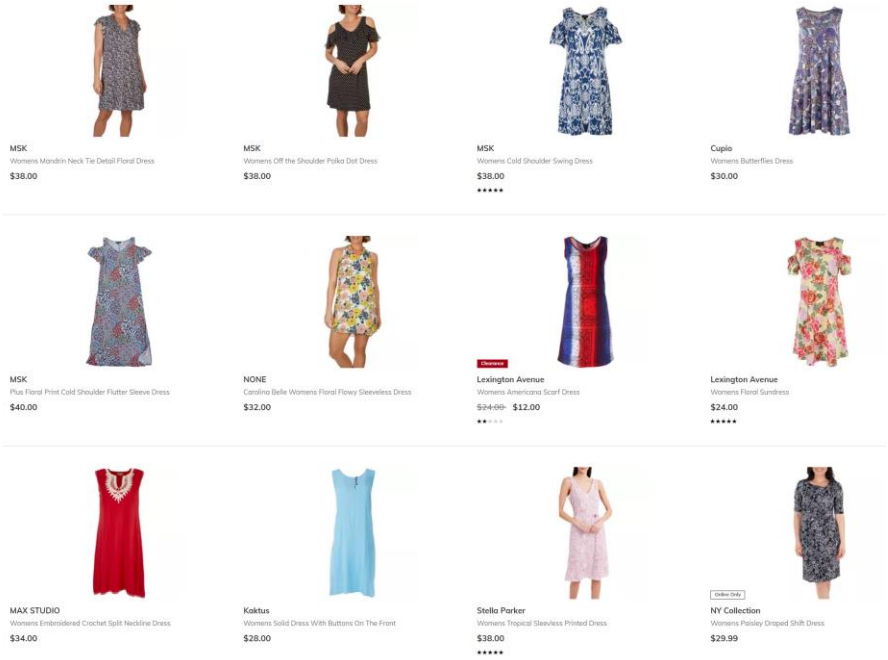
In addition, the first FP should be similar between treatment and control groups because consumers have not yet been exposed to the treatment before they view the first FP. Accordingly, (i) we compare the average and different percentile values (minimum, 25th, 50th, 75^{th,} and maximum) of price distribution of the first FP for the treated and control purchase funnels; (ii) we compute the average affinity score of the first FP with all other products in its product category for each of the treated and control purchase funnels. The average affinity score of the first FP measures its closeness with all other products (network centrality) in its product category. Similarly, we compare the average and different percentile values of the closeness for the first FPs for the two types of purchase funnels. Table A1 reports the results, showing that the first FPs viewed in treated and control purchase funnels have statistically similar distributions in terms of price and closeness.

Table A1. Balance Check at the Visitor and Purchase Funnel Level

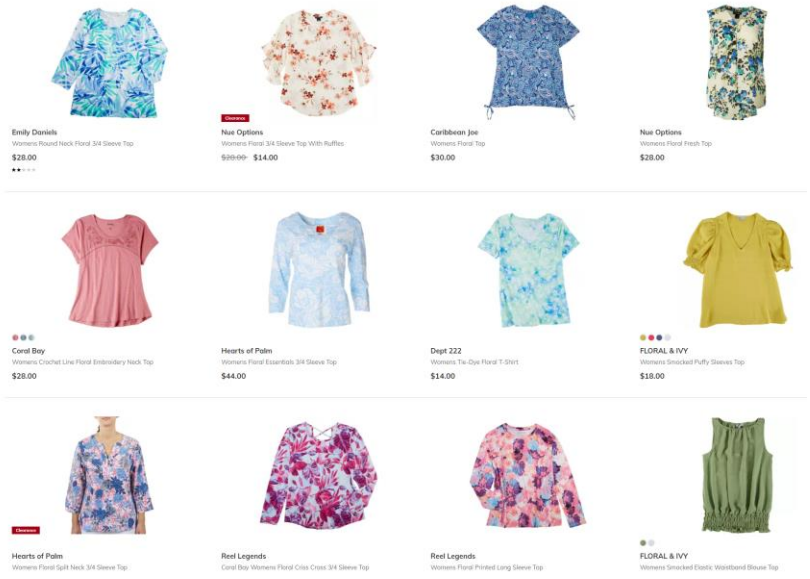
		Control	Treated	Diff. in means
		Mean (Std. Dev.)	Mean (Std. Dev.)	(*t*-stats)
Visitor Level				
New visitor		0.8779 (0.33)	0.8769 (0.33)	-0.0010 (-1.01)
Frequent visitor		0.0288 (0.17)	0.0284 (0.17)	-0.0004 (-0.70)
Loading	Paid search	0.3224 (0.47)	0.3209 (0.47)	-0.0015 (-1.05)
	Organic search	0.1110 (0.31)	0.1117 (0.32)	0.0007 (0.75)
	Others	0.5666 (0.50)	0.5673 (0.50)	0.0008 (0.51)
Purchase funnel level				
Visitor Characteristics				
New visitor		0.8452 (0.36)	0.8454 (0.36)	0.0002 (0.23)
Frequent visitor		0.0343 (0.18)	0.0348 (0.18)	0.0006 (1.18)
Loading	Paid search	0.3353 (0.51)	0.3344 (0.51)	-0.0009 (-0.66)
	Organic search	0.1214 (0.34)	0.1221 (0.34)	0.0007 (0.77)
	Others	0.5644 (0.50)	0.5647 (0.50)	0.0003 (0.24)
First FP's Characteristics				
Price	Average	31.05 (19.17)	30.97 (19.36)	-0.0821 (-0.02)
	Minimum	9.66 (4.87)	9.42 (4.65)	-0.2427 (-0.30)
	25th percentile	22.23 (11.65)	22.07 (11.86)	-0.1627 (-0.08)
	Median	28.92 (16.58)	28.76 (16.75)	-0.1602 (-0.06)
	75th percentile	38.19 (30.83)	38.29 (30.83)	0.0957 (0.02)
	Maximum	84.48 (64.45)	86.23 (64.96)	1.7466 (0.16)
Closeness	Average	3.09 (5.68)	3.07 (5.33)	-0.0362 (-0.27)
	Minimum	0.12 (0.92)	0.08 (0.59)	-0.0766 (-0.13)
	25th percentile	1.98 (3.67)	1.91 (3.14)	-0.0254 (-0.03)
	Median	3.05 (5.85)	3.02 (5.44)	-0.0756 (-0.06)
	75th percentile	4.21 (8.04)	4.14 (7.63)	-0.1195 (-0.07)
	Maximum	6.69 (10.40)	6.57 (10.21)	-0.0362 (-0.27)

Notes: Diff. in means = Treated - Control. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix B. Women's Dress Category Main Page and the Search Result Page



(a) Women's Dress Category Main Page



(b) Search Result Page for "Floral Women's Top"

Figure B1. Examples of Product Category Page and Search Result Page

Appendix C. Robustness Checks

C1. Affinity Scores Computed Based on Data with and without Recommendations

We re-estimate the effect of recommendations based on data from days 31-63 of the experiment. Unlike the first 30 days, Day 31 to the last day (Day 63) of the experiment, the affinity scores were computed based on the browsing/purchasing data from both treated and control consumers. In this setting, half of the consumers could see recommendations, and the other half could not. The possible "inflation" in the affinity scores should be less of a concern compared to a setting where all consumers see recommendations.

We re-estimate the effect of recommendations on the net value and horizontal fit with the same specifications. Panel A of Table C1 reports the effects of recommendations on product views, and Panel B reports the results on product purchase (conditional on funnels with purchased products). The results in Table C1 are qualitatively similar (in sign, significance, and magnitude) to our main results. Therefore, our main findings are robust in a setting where affinity scores are computed from data where half of the consumers could see recommendations and the other half could not.

Table C1. Estimation Results Based on Data from Day 31 to Day 63

Panel A. Recommendations and Consumers' Search Behavior

	(1) Avg. Affinity Score	(2) Avg. Horizontal Fit
Rec. Indicator	0.414***	0.269***
	(0.023)	(0.014)
R^2	0.0961	0.0092
Obs.	285,201	285,201

Notes: + p < .10, * p < .05, ** p < .01, *** p < .001.
Standard errors cluster corrected at product subcategory level in parentheses.

Panel B. Effects of Recommendations on Purchased Products

	(1) Purchase Affinity Score	(2) Purchase Horizontal Fit
Rec. Indicator	0.558***	0.451***
	(0.074)	(0.059)
R^2	0.0361	0.0149
Obs.	17,824	17,824

Notes: + p < .10, * p < .05, ** p < .01, *** p < .001.
Standard errors cluster corrected at product subcategory level in parentheses.

C2. Affinity Scores Computed Based on Data without Recommendations

Similar to the analysis in section C1, we re-estimate the effect of recommendations based on data from days 31-63 of the experiment. However, in this analysis, we recompute the affinity scores between products using the viewing and purchasing behaviors from only control visitors (i.e., those who do not see recommendations) during our experiment period. In other words, we recompute the affinity score for each pair of products from Day 31 to the last day (Day 63) by applying the formula (described in Section 3.2) to only control consumers browsing data. We report the effect of recommendations on the net value and horizontal fit based on these newly computed affinity scores in Table C2. We find qualitatively similar results (in sign, significance, and magnitude) to our main results.

Table C2. Estimation Results Based on Data from Day 31 to Day 63 without Recommendations

Panel A. Recommendations and Consumers' Search Behavior

	(1) Avg. Affinity Score	(2) Avg. Horizontal Fit
Rec. Indicator	0.295***	0.173***
	(0.026)	(0.013)
R^2	0.0868	0.0061
Obs.	285,201	285,201

Notes: + p < .10, * p < .05, ** p < .01, *** p < .001.
Standard errors cluster corrected at product subcategory level in parentheses.

Panel B. Effects of Recommendations on Purchased Products

	(1) Purchase Affinity Score	(2) Purchase Horizontal Fit
Rec. Indicator	0.325**	0.284***
	(0.104)	(0.075)
R^2	0.0582	0.0120
Obs.	17,824	17,824

Notes: + p < .10, * p < .05, ** p < .01, *** p < .001.
Standard errors cluster corrected at product subcategory level in parentheses.