Supplementary Online Appendices

Appendix A. Two Lab Experiments on the Manipulation Effectiveness

Lab Experiment I

Participants and Procedure. The experiment was conducted on *Sojump.com*, a popular online experiment platform in China. The participants were primarily students in top universities in China, including undergraduate, masters, and doctoral students. The experiment employed a between-subjects design. Participants, after indicated their gender, were randomly assigned to one of four groups (i.e., 'Control', 'Demand', 'Demand + Popularity Cue', and 'Demand + Capacity Cue') to examine information on the profile page of a user of the opposite gender. For example, a female participant views the profile page of a male user in the experiment and vice versa. We only considered the user of the opposite gender because our primary focus (as is the corporate partner) was heterosexual dating. Demand information was measured as the number of requests the user received in the past 30 days. In the experiment, we used two *vs.* twenty-four requests to dichotomize the user as a low-demand user *vs.* a high-demand user. After a few iterations, we specified the manipulations of popularity cue and capacity cue in the low-demand and high-demand group (as shown in Table A-1).

The experiment began with a brief introduction to the platform. Participants were then informed, on the same page, that the experiment was designed to improve user experience in the profile pages of the users on our corporate partner platform. To prevent them from jumping to the next session without being fully aware of the experiment, participants were mandated to spend at least fifteen seconds on this page, an estimated time to read the introduction. Next, each participant self-identified gender and was randomly assigned to one of the four treatments that included a profile page for a user of the opposite sex. In the hypothetical scenario, participants read through the information on a user profile page, indicated their agreement to our measurement questions of perceive popularity and perceived capacity in the Likert-type scales (see Table A-2), and reported their demographic information. In the experiment, we also

implemented manipulation checks to ensure the quality of the experiment responses. The first manipulation check questioned how many requests the user in the profile has received, and the second manipulation check asked how the platform describes the number of requests displayed in the user profile (e.g., 'very popular,' 'not picked by many others,' 'busy handling requests,' or 'has time to handle requests'). Among the 280 participants, 178 passed both manipulation checks, and the participants who failed in either of the manipulation checks were dropped out of the sample.

Lastly, as interviewees indicated that the they set benchmarks for how many requests made a user popular, we asked an additional question on how many requests a user should at least receive to be considered popular. The mode from the responses for this question is ten. Further, the median number of requests an active user receives in the past thirty days is nine. Based on the above, we believe using 9/10 as cut off for low *vs.* high-demand peer in the field experiment is reasonable.

Table A-1. Manipulation Information by Groups in Lab Experiment I

Group #	*Treatment*	*Manipulation Information (English)*	*Manipulation Information (Mandarin Chinese)*
1	Control	N/A	--
2	Demand	Received x requests in the past month	近一个月收到 x 份问卷
3	Demand + Popularity Cue	[High] Received x requests in the past month, this lady (or gentleman) is very popular [Low] Received x requests in the past month, this lady (or gentleman) is not picked by many others	[高]近一个月收到 x 份问卷，小姐姐（或小哥哥）很热门很火 [低]近一个月收到 x 份问卷，小姐姐（或小哥哥）没有很多人 pick
4	Demand + Capacity Cue	[High] Received x requests in the past month, this lady (or gentleman) is busy handling requests [Low] Received x requests in the past month, this lady (or gentleman) has time to handle requests	[高]近一个月收到 x 份问卷，小姐姐（或小哥哥）阅卷有点忙 [低]近一个月收到 x 份问卷，小姐姐（或小哥哥）有精力阅卷

Measurements for the Manipulations. Participants reported their agreement to questions on perceived popularity and perceived capacity on a seven-point Likert scale (i.e., 1-strongly disagree, 2-disagree, 3-somewhat disagree, 4-neither disagree nor agree, 5-somewhat agree, 6-agree, and 7-strongly agree). The

questions were designed to closely align with the definition of popularity and capacity.[1] Since the participants, i.e., potential users of the corporate partner platform, are native Chinese speakers, the experiment questions were delivered in Mandarin Chinese. We provide the English translation of the measurement questions in Table A-2, which was validated by bilingual speakers.

Table A-2. Measurement Questions in Lab Experiment I

Construct	*Measurement Question (English)*	*Measurement Question (Mandarin Chinese)*
Perceived Popularity	This user is liked by other users on the platform.	这位同学受到平台上其他同学的喜欢
Perceived Capacity	This user does not have enough time to interact with me.[2]	这位同学可能没有时间和精力与我互动交流

Results of Treatment Validation. To validate our treatment, we respectively examine the perceived popularity and perceived capacity for the low-demand users and high-demand users. The experiment results suggest that our treatment information largely leads to the intended perception of the manipulations. The participants' responses do receive proper priming of popularity and capacity for low-demand users and high-demand users.

For perceived popularity, comparing the priming for low-demand users (received two requests in the past month) amongst the three treatment stimuli ('Demand', 'Demand + Popularity Cue', 'Demand + Capacity Cue'), we find the 'Demand + Popularity Cue' group to be significantly lower in perceived popularity than the 'Demand' group (*diff* = -1.03, *s.e.* = .27, $p < .01$), and there is no significant difference between the 'Demand + Capacity Cue' group and the 'Demand' group (*diff* = -.25, *s.e.* = .30, $p > .10$). Comparing the priming for high-demand users (received 24 requests in the past month) amongst the three treatment stimuli, we find the 'Demand + Popularity Cue' group to be significantly higher in perceived popularity than the 'Demand' group (*diff* = .78, *s.e.* = .31, $p < .01$), and there is a significant but smaller

[1] We kept the items that are most relevant to the perceived popularity and perceived capacity measures for Experiment I. Given that Experiment I does not isolate the effect of demand information number from the high *vs.* low cues, we conducted Experiment II, which features a design with multiple relevant items for perceived popularity and perceived capacity measures.

[2] Note that this item was reverse coded to indicate available capacity in the analysis.

difference between the 'Demand + Capacity Cue' group and the 'Demand' group (*diff* = .62, *s.e.* = .30, $p < .05$). For perceived capacity, comparing the priming for low-demand users (received two requests in the past month) amongst the three treatment stimuli ('Demand', 'Demand + Popularity Cue', 'Demand + Capacity Cue'), we find that the 'Demand + Capacity Cue' group to be significantly higher in perceived capacity than the 'Demand' group (*diff* = .71, *s.e.* = .36, $p < .05$), and there is no significant difference between the 'Demand + Popularity Cue' group and the 'Demand' group (*diff* = .19, *s.e.* = .34, $p > .10$). Comparing the priming for high-demand users (received 24 requests in the past month) amongst the three treatment stimuli, we find the 'Demand + Capacity Cue' group to be significantly lower in perceived capacity than the 'Demand' group (*diff* = -.60, *s.e.* = .36, $p < .01$), and there is no significant difference between the 'Demand + Popularity Cue' group and the 'Demand' group (*diff* = -.30, *s.e.* = .38, $p > .10$).

Lab Experiment II to Further Validate Priming Cues

Design Overview. We were cognizant that in the first experiment, demand information (2 *vs.* 24) is not disentangled from the low *vs.* high-demand (popularity or capacity) cues. Therefore, we conduct a second lab experiment to further validate the treatments, and particularly whether the users are primed by our manipulation on the popularity and capacity, above and beyond the demand information. Notably, we seek to isolate the priming effects of the cues. First, before we launch the experiment, we conduct in-depth interviews to understand how participants interpret the manipulation. Based on the responses of participants from the interviews, we finetune our design and the measurement questions regarding perceived popularity and perceived capacity used in the first lab experiment. Second, the participants in this experiment are active users on our partner platform who have been exposed to the demand information after the feature was implemented since our randomized field experiment (for more than eight months); those users have a good understanding of the cues and demand information. Third, to isolate the effect of cues (priming) from the potential impact from the change in demand information numbers (2 *vs.* 24 in the first lab experiment), the number of matching requests (specifically, we use '9' to be consistent

with the cutoff value in our field experiment) received by the profile owner remained the same across groups. Fourth, to achieve maximum control of the potential confounding effects from user characteristics, the information on the profile page was the same except for the gender, which needs to be matched with participants' opposite gender in the heterosexual dating context.

Interviews. We conducted 33 interviews with active users in our partner platform, 18 females and 15 males. Specifically, users were invited to complete a pre-test experiment and a follow-up interview with one of the researchers on the team. The participants are Chinese students in top universities in China and the United States, such as Yale University, Carnegie Mellon University, and the University of California, Los Angeles. Each of them was randomly assigned to one of the six treatment groups (i.e., 'Control', 'Demand', 'Demand + Low Popularity Cue', Demand + High Popularity Cue', 'Demand + Low Capacity Cue', and 'Demand + High Capacity Cue'). Demand information remains the same across groups. After the experiment, all the participants accepted our invitation to a semi-structured interview and expressed how they interpreted the treatment message, if and how they felt primed by the message. For example, participants agreed that they used the popularity cues and capacity cues as salient signals. Further, our treatment messages regarding the high *vs.* low conditions played a crucial role in shaping participants' perception of popularity and capacity.

Main Experiment Participants and Procedure. The second experiment was also conducted on *Sojump.com* and adopted a between-subjects design. In the experiment, after revealing their gender information, participants were randomly assigned to one of six groups to examine information on the profile page of a user of the opposite gender. Demand information was presented as the number of matching requests the user received in the past 30 days. In the experiment, we only used 'nine' consistently across groups to present demand information since keeping the demand information constant will remove the potential confounding effect from the change of demand information. Again, we purposefully chose '9' as the demand information number because we also used 9/10 as a threshold to dichotomize the user as a low-demand user *vs.* a high-demand user in the randomized natural field

experiment. We specified the manipulations of popularity cue and the capacity cue for each group, as shown in Table A-3.

Table A-3. Manipulation Information by Groups in Lab Experiment II

Group #	*Treatment*	*Manipulation Information (English)*	*Manipulation Information (Mandarin Chinese)*
1	Control	N/A	--
2	Demand	Received nine requests in the past month	近一个月 9 人答题
3	Demand + Low Popularity Cue	Received nine requests in the past month, this lady (or gentleman) is not picked by many others	近一个月 9 人答题，小姐姐（或小哥哥）没有很多人 pick
4	Demand + High Popularity Cue	Received nine requests in the past month, this lady (or gentleman) is very popular	近一个月 9 人答题，小姐姐（或小哥哥）很热门很火
5	Demand +Low Capacity Cue	Received nine requests in the past month, this lady (or gentleman) is busy handling requests	近一个月 9 人答题，小姐姐（或小哥哥）阅卷有点忙
6	Demand + High Capacity Cue	Received nine requests in the past month, this lady (or gentleman) has time to handle requests	近一个月 9 人答题，小姐姐（或小哥哥）有精力阅卷

Similar to the procedure in the first experiment, subjects were informed that they were participating in a study in partnership with Summer Inc., aiming at optimizing user experience in the profile page. Users stayed at the introduction page for at least fifteen seconds. Each subject self-indicated gender and was assigned to one of the six treatment groups that showed a profile page for a user of the opposite sex. All the information across the six groups is the same, except we use the opposite gender, and the experimental manipulation on popularity *vs.* capacity cue when they are either high or low.

The participants indicated their agreement to measurement questions regarding perceptions of the treatments in the Likert-type scales (see Table A-4) and reported their demographics. To ensure the quality of the experiment, we conducted several standard attention checks and manipulation checks. First, the participants were asked to indicate the gender and age of the profile owner. Second, they reviewed the questions for two manipulations, the same as the manipulation checks in the first experiment. The first manipulation check questioned how many requests the user in the profile has received, and the second

manipulation check asked how the platform describes the number of requests displayed in the user profile. Among the 529 participants, 370 passed all the checks, and the participants who failed in one of the checks were removed from the sample.

Measurements. Participants reported their agreement to questions about perceived popularity and perceived capacity on a seven-point Likert-type scale. This time, improving upon the construct measurement from the first lab experiment, rather than using single item to represent the two concepts, we adopted multiple items to measure the two constructs because this approach allows us to evaluate the validity of the constructs. The items were iterated several rounds based on participants' responses in the interviews. We provide the English translation of the measurement questions in Table A-4, which was validated by bilingual speakers.

Table A-4. Measurement Questions in Lab Experiment II

Construct	*Measurement Questions (English)*	*Measurement Questions (Mandarin Chinese)*
Perceived Popularity	This user is desired by other users on the platform.	这位同学在 Summer 受到其他同学的欢迎
	This user is liked by other users on the platform.	这位同学在 Summer 受到其他同学的青睐
	This user is popular among other users on the platform.	这位同学在 Summer 上的人气高
Perceived Capacity	This user recently spends much time to communicate with other users on the platform.	这位同学近期花大量时间和 Summer 上的其他同学交友
	This user recently engages with many other users on the platform.	这位同学近期和 Summer 上的多个其他同学互动
	This user does not have enough time to communicate with me.	这位同学没有精力与我聊天

Results. The goal of this analysis is to evaluate a) the construct validity of the items; and b) whether the popularity and capacity cues contribute to the intended manipulations beyond the demand information number. We first examine the construct reliability and validity for perceived capacity and perceived popularity. For perceived popularity, the Cronbach's Alpha is higher than the standard value of 0.7 (α = .87), and composite reliability is 0.93 while the absolute value of the correlation between perceived capacity and perceived popularity (correlation = -.46) is lower than the square root of average variance

extracted (AVE) (AVE = .82). For perceived capacity, the Cronbach's Alpha is also higher than the standard value (α = .73), and composite reliability is 0.85 while the absolute value of the correlation between perceived capacity and perceived popularity is lower than the square root of AVE (AVE = .74). Since both perceived popularity and perceived capacity possess high reliability and validity, we averaged the items for the two constructs respectively and used the averaged values in later analysis.

For perceived popularity, comparing the priming for capacity cue amongst the three treatment stimuli ('Demand', 'Demand + High Popularity Cue', 'Demand + Low Popularity Cue'), we find the 'Demand + Low Popularity Cue' group to be significantly lower in perceived popularity than the 'Demand' group (*diff* = -.60, *s.e.* = .18, $p < .05$) while the 'Demand + High Popularity Cue' group to be significantly higher than the 'Demand' group (*diff* = .68, *s.e.* = .18, $p < .01$). Besides, the 'Demand + High Popularity Cue' group is significantly higher in perceived popularity than the 'Demand + Low Popularity' Cue (*diff* = 1.29, *s.e.* = .18, $p < .001$).

For perceived capacity, comparing the priming for capacity cue amongst the three treatment stimuli ('Demand', 'Demand + High Capacity Cue', 'Demand + Low Capacity Cue'), we find the 'Demand + High Capacity Cue' group to be significantly higher in perceived capacity than the 'Demand' group (*diff* = .53, *s.e.* = .19, $p < .10$) and the 'Demand + Low Popularity Cue' group to be significantly lower in perceived capacity than the 'Demand' group (*diff* =- .49, *s.e.* = .18, $p < .10$). We also find the 'Demand + High Capacity Cue' group is significantly higher in perceived capacity than the 'Demand + Low Capacity Cue' (*diff* = 1.02, *s.e.* = .18, $p < .001$).

Overall, this lab experiment and associated analyses suggest that holding the number used as the peer demand information constant, priming cues on popularity and capacity we implemented have an additional independent effect on users' perceived popularity and perceived capacity of the peer in the dating market, respectively.

Appendix B. Randomization Checks

Before data analyses, we performed randomization checks to examine the validity of our randomization procedure in the field experiment. As our experiment employs a between-subject design, the randomization and treatment manipulation are implemented on each participant, we conduct the randomization checks at the user level. Specifically, we consider the pairwise mean comparisons of the subjects' demographics, prior engagement, prior demand, date of entering the experiment, as well as distributional comparisons of users' location, school, and major. The balance checks confirm the effectiveness of the randomization procedure, and we present the details in Table B-1.

Table B-1. Randomization Checks

	Gender	*Age*	*Province*	*School*	*Major*	*Degree*	*Date*	*Log(req)*	*Log(req received)*
2 *vs.* 1	1.00	0.74	0.85	0.82	0.35	0.83	0.57	0.93	0.27
3 *vs.* 1	0.80	1.00	0.47	0.58	0.82	0.79	0.80	0.79	0.91
4 *vs.* 1	0.32	0.77	0.92	0.62	0.75	1.00	0.82	0.68	1.00
3 *vs.* 2	0.76	0.74	0.79	0.40	0.74	0.81	0.98	0.41	0.67
4 *vs.* 2	0.28	1.00	0.99	0.66	0.39	0.99	0.98	0.30	0.38
4 *vs.* 3	0.86	0.76	0.90	0.57	0.31	0.98	1.00	1.00	0.97

Notes: Group 1 is the holdout group, group 2 denotes the control group, group 3 represents the 'demand + popularity cue' group, and group 4 stands for the 'demand + capacity cue' group. The *p*-values for the pairwise *t* tests are reported for numerical variables, i.e., gender, age, log(req), log(req received); and *p*-values for non-parametric Kolmogorov-Smirnov tests for equal distributions used for categorical data, i.e., province, school, major, degree. No pairwise comparison shows significant differences at the conventional levels between any two groups in any observed aspect of user personal information or platform engagement.

Appendix C. Analyses of Profile Views

We first visualize subjects' total number of unique profile views with 90% confidence intervals, decomposed by high-demand versus low-demand user profiles, in Figure C-1. We see that subjects' number of profile views toward both high-demand and low-demand peers did not shift under the different treatments. The overlapping confidence intervals also indicates that there are no statistically significant differences between any pair of values.

We then estimate whether the treatments influenced subjects' total number of unique profile views. As specified in Equation C-1, the key indicator is $Treat_u$, which represents the experimental group of a subject, with the control group acting as the reference condition. In the equation, *u* indexes subjects and *h* indexes the level of peer demand, thus the variable $HighDemandPeer_h$ is a binary indicator of whether the observation pertains to profile clicks of high or low-demand peers, such that 0 refers to profile clicks issued towards low-demand peers (who received < 10 requests in the past 30 days) whereas 1 refers to profile clicks issued towards high-demand peers (who received >=10 requests in the past 30 days). The outcome variable, $ProfileViews_{uh}$, thus captures a subject *u*'s total number of profile clicks toward peers of type *h* during the experiment period. Finally, U_u is a user-level random effect. Subsequently, we repeat this estimation employing a log transformation to the outcome, to simultaneously account for right-skewness in the outcome and to obtain elasticity estimates.

$$ProfileViews_{uh} = \boldsymbol{Treat}_u + HighDemandPeer_h + \boldsymbol{Treat}_u * HighDemandPeer_h$$
$$U_u + \epsilon_{uh} \qquad \text{(C-1)}$$

We report the regression results in Table C-1, where we report results based on linear models with and without date fixed effects and user-level random effects. Column 1 presents the estimation coefficients of the OLS model with random effects and Column 2 presents the estimation coefficients of the Log-OLS model with random effects. We observe that the subjects' number of profile views do *not* significantly vary across the experimental groups, either for high-demand or low-demand user profiles.

The above evidence suggests that experiment manipulation was properly executed, and the effects of the treatment do not spill over to profile views.

Figure C-1. Volume of Profiles Viewed by Experiment Condition and Peer Demand Split

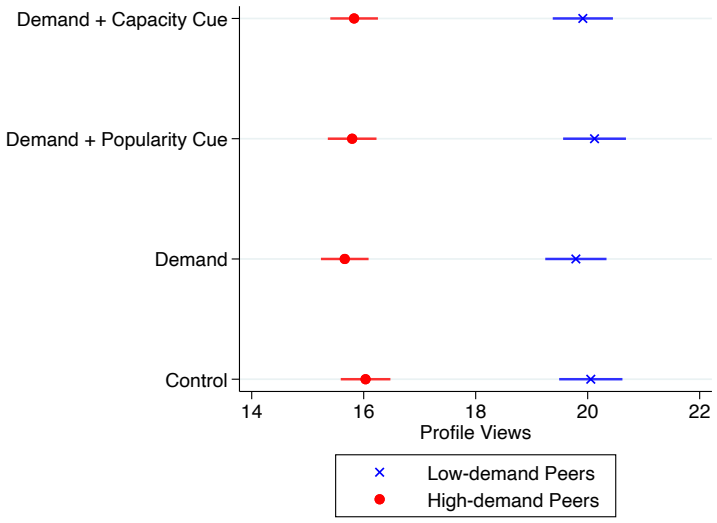

Table C-1. Regressions Analysis (Placebo Test): Profile Views

Variable	(1) OLS-RE	(2) Log-OLS-RE
Demand	0.266 (0.478)	-0.001 (0.017)
Demand+PopularityCue	0.333 (0.476)	0.011 (0.017)
Demand+CapacityCue	0.124 (0.466)	0.008 (0.017)
HighDemandPeer	**-4.127*****(0.277)**	**-0.247***** (0.011)**
Demand x HighDemandPeer	0.105 (0.391)	-0.010 (0.016)
Demand+PopularityCue x HighDemandPeer	-0.201 (0.393)	0.000 (0.016)
Demand+CapacityCue x HighDemandPeer	0.043 (0.383)	0.004 (0.016)
Constant	**19.789***** (0.332)**	**2.122***** (0.012)**
Observations	97,316	97,316
Users	48,658	48,658
Wald Chi^2	**913.94***** (7)	**1960.62***** (7)

Notes: We incorporate user random effects and clustered standard errors to address the fact that we have two observations per participant (i.e., participants' requests to high- *vs.* low-demand peers);
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix D. Assessing Selection on Treatment for Analysis of Matches per Request

In this appendix, we consider the possible issue of selection on treatment in our matching efficiency estimations. We perform balance checks on the sample of subjects who issued at least one matching request in the two-week period of observation. If our results are a product of selection on treatment, we might expect to observe systematic differences in some observable characteristics across experimental conditions. Table D-1 displays the largest observed standardized mean difference (Cohen's *d*) across all pairs of conditions in the selected sample. As can be seen, imbalance is minimal. All the maximal pairwise differences are far below 0.2 (the typical threshold for a 'small' effect in Cohen's *d*). Pairwise t-tests of significant differences similarly yield insignificance.

Table D-1 also displays the resulting maximal differences once we re-weight the sample employing covariate balancing propensity scores (Imai and Ratkovic 2014). Table D-2 presents estimation results for our matching efficiency outcomes when weighting based on the covariate balancing propensity scores. We observe consistent results; in fact, our estimates all grow larger and statistical significance also increases.

Table D-1. Adjusted vs. Unadjusted Standardized Mean Differences

Variable	Max Pairwise Standardized Mean Difference (Unadjusted)	Max Pairwise Standardized Mean Difference (Adjusted)
Female	0.0081	0.0036
Age	0.0173	0.0069
Sent Requests (30 Days)	0.0252	0.0128
Received Requests (30 Days)	0.0389	0.0115
User Experiment Entry (Days)	0.0138	0.0094
	Effective Sample Size (Unadjusted)	Effective Sample Size (Adjusted)
Control	5,338	5,337.21
Demand	5,398	5,396.37
Demand + Capacity Cue	5,357	5,355.72
Demand + Popularity Cue	5,349	5,347.15

Table D-2. Regressions Analysis: Effect on Matches per Request (IPTW Employing Covariate Balancing Propensity Scores)

Variable	(1) Initial Matches / Request	(2) Engaged Matches / Request
Demand	-0.002 (0.006)	0.004 (0.004)
Demand+PopularityCue	0.007 (0.006)	0.005 (0.004)
Demand+CapacityCue	0.021*** (0.006)	0.010** (0.004)
Constant	0.285*** (0.004)	0.117*** (0.003)
Observations	21,442	21,442
F-stat	5.619 (3, 21438)	1.693 (3, 21438)

Notes: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

References

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243-263.

Appendix E. Assessing the Potential for Interference

In the analysis that follows, we seek to mitigate the potential that our results are a product of interference or spillover, given the networked setting in which our experiment has been conducted. To achieve this, we leverage our tap-stream data, and limit our regression to the very first peer visit that was initiated by any user. This results in a much smaller sample, with the benefit that these observations occurred very soon after subjects entered the experiment, thereby mitigating the potential that their behavior would be contaminated by peers. As can be seen in Table E-1, we observe an identical pattern of attentional reallocation. Although this result cannot rule out interference completely, it suggests our results are not severely biased by such spillovers.

Table E-1. Regressions Analysis: Treatment Effect on Peer Request (Very First Peer Profile Visit)

Variable	(1) OLS (High Peer)	(2) OLS (Low Peer)
Demand	0.0003 (0.008)	0.004 (0.007)
Demand+PopularityCue	-0.015* (0.008)	0.011 (0.007)
Demand+CapacityCue	-0.021*** (0.008)	0.015** (0.007)
Constant	0.241** (0.006)	0.175*** (0.005)
Observations	22,198	21,641
F-stat	3.75 (3, 22194)	1.76 (3, 21637)

Notes: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$