The Polarity of Online Reviews: Prevalence, Drivers and Implications

Web Appendix

Web Appendix 1: Extreme Distribution 24 Amazon Categories

	1 star	2 stars	3 stars	4 stars	5 stars	#products	#reviews	Polarity	Positive Imbalance
Apps	11%	5%	10%	21%	53%	61,275	2,638,172	64%	82%
Automotive	9%	5%	7%	17%	62%	320,112	1,373,768	71%	85%
Baby Products	8%	6%	9%	18%	59%	64,426	915,446	67%	85%
Beauty	9%	6%	8%	15%	62%	249,274	2,023,070	71%	84%
Books	5%	4%	9%	20%	62%	2,330,066	22,507,155	67%	90%
CD and Vinyl	4%	4%	7%	18%	67%	486,360	3,749,004	71%	91%
Cell Phones	15%	7%	10%	18%	50%	319,678	3,447,249	65%	76%
Clothing	7%	6%	10%	19%	58%	1,136,004	5,748,920	65%	86%
Digital Music	4%	2%	4%	15%	75%	266,414	836,006	79%	94%
Electronics	12%	6%	8%	18%	56%	476,002	7,824,482	68%	80%
Grocery and Gourmet Food	8%	5%	7%	14%	66%	166,049	1,297,156	74%	86%
Health	10%	6%	8%	16%	60%	252,331	2,982,326	70%	83%
Home & Kitchen	10%	6%	8%	17%	59%	410,243	4,253,926	69%	83%
Instant Video	6%	4%	7%	17%	66%	23,965	58,933	72%	89%
Kindle Stores	5%	5%	10%	22%	58%	430,530	3,205,467	63%	89%
Movies	7%	5%	9%	19%	60%	200,941	4,607,047	67%	87%
Musical Instruments	7%	5%	7%	19%	62%	83,046	500,176	69%	87%
Office	13%	6%	8%	17%	56%	130,006	1,243,186	69%	79%
Patio	12%	6%	8%	17%	57%	105,984	993,490	69%	80%
Pet	10%	6%	9%	15%	60%	103,288	1,235,316	70%	82%
Sports	8%	5%	8%	20%	59%	478,898	3,268,695	67%	86%
Tools	9%	5%	8%	18%	60%	260,659	1,926,047	69%	85%
Toys	9%	5%	9%	18%	59%	327,698	2,252,771	68%	85%
Video Games	11%	6%	9%	20%	54%	50,210	1,324,753	65%	81%

Web Appendix 2: Scale Transformation to Calculate Polarity and Positive Imbalance for Platforms with Non-5-Point Scales

Platform	Scale	Scale Transformation
Fandango	5 (but 0.5 intervals)	0.5 &1=1 1.5 & 2 = 2 2.5 & 3 = 3 3.5 & 4 = 4 4.5 & 5 = 5
Booking.com	10	1 & 2 =1 3 & 4 =2 5 & 6=3 7 & 8=4 9&10=5
Yahoo! Movies†	13	1 & 2 = 1 3 & 4 & 5 = 2 6 & 7 & 8 = 3 9 & 10 & 11 = 4 12 & 13 = 5
Metacritic	11	0 & 1 = 1 2 & 3 = 2 4 & 5 & 6 = 3 7 & 8 = 4 9 & 10 = 5
Rotten Tomatoes	5 (but 0.5 intervals)	0.5 &1=1 1.5 & 2 = 2 2.5 & 3 = 3 3.5 & 4 = 4 4.5 & 5 = 5
IMDb	10	1 & 2 = 1 3 & 4 = 2 5 & 6 = 3 7 & 8 = 4 9 & 10 = 5
Twitter Movies	10	1 & 2 = 1 3 & 4 = 2 5 & 6 = 3 7 & 8 = 4 9 & 10 = 5

MovieLens	5 (but 0.5 intervals)	0.5 &1=1 1.5 & 2 = 2 2.5 & 3 = 3 3.5 & 4 = 4 4.5 & 5 = 5
RateBeer	20	1-4=1 5-8=2 9-12=3 13-16=4 17-20=5

Web Appendix 3: External Data Sources and Link to Data Repository

Platform	Link
YourXpert	***Proprietary Data***
Frag-Mutti	***Proprietary Data***
Trust Pilot	https://osf.io/6n2kt/
BlaBlaCar	https://osf.io/6n2kt/
Airbnb	***Proprietary Data***
Amazon$^{\Delta}$	http://jmcauley.ucsd.edu/data/amazon/links.html
Google Restaurants	https://osf.io/6n2kt/
Fandango	https://osf.io/6n2kt/
Online Retailer	***Proprietary Data***
Edmunds	https://osf.io/6n2kt/
Booking.com	https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe#Hotel_Reviews.csv
Epinions	https://osf.io/6n2kt/
Yahoo! Songs (Launchcast)	https://webscope.sandbox.yahoo.com/catalog.php?datatype=r
Expedia	***Proprietary Data***
Yelp Restaurants	https://www.yelp.com/dataset
Yahoo! Movies†	https://webscope.sandbox.yahoo.com/catalog.php?datatype=r
TripAdvisor	https://www.kaggle.com/c/CL2019/data
Metacritic	https://osf.io/6n2kt/
Rotten Tomatoes	https://osf.io/6n2kt/
Goodreads	https://osf.io/6n2kt/
IMDb$^{\Omega}$	https://cran.r-project.org/web/packages/ggplot2movies/index.html
Netflix	*http://www.netflixprize.com* https://www.kaggle.com/netflix-inc/netflix-prize-data
Twitter Movies‡	https://github.com/sidooms/MovieTweetings
MovieLens	https://grouplens.org/datasets/movielens/
RateBeer	https://snap.stanford.edu/data/index.html

Web Appendix 4: Regression of Cross-Platform Polarity and Positive Imbalance on Platform Characteristics Controlling for the Average Number of Reviews per Reviewer

	Polarity	**R^2**	**Positive Imbalance**	**R^2**
Intercept	0.572 (0.050)***	0.393	0.654 (0.022)***	0.305
log (Average # Reviews per Reviewer)	-0.05 (0.014)***		-0.018 (0.006)**	
Intercept	0.637 (0.088)***	0.418	0.684 (0.039)***	0.336
Age of Platform	-0.005 (0.005)		-0.002 (0.002)	
log (Average # Reviews per Reviewer)	-0.045 (0.015)***		-0.016 (0.007)**	
Business Model (Reference: Selling products or services)				
Intercept	0.535 (0.079)***	0.440	0.614 (0.032)***	0.455
Transaction Fee	0.081 (0.088)		0.073 (0.036)*	
Information Platform	0.004 (0.078)		0.024 (0.032)	
log (Average # Reviews per Reviewer)	-0.045 (0.015)***		-0.015 (0.006)**	
Product Category (Reference: Products/Services)				
Intercept	0.551 (0.051)***	0.491	0.643 (0.022)***	0.442
Travel/Restaurants	0.043 (0.062)		0.024 (0.026)	
Entertainment	-0.083 (0.068)		-0.037 (0.029)	
log (Average # Reviews per Reviewer)	-0.038 (0.016)**		-0.013 (0.007)*	
Intercept	0.58 (0.043)***	0.570	0.654 (0.022)***	0.305
Scale Points	-0.137 (0.049)**		-0.003 (0.026)	
log (Average # Reviews per Reviewer)	-0.036 (0.013)**		-0.018 (0.007)**	
Intercept	0.598 (0.051)***	0.463	0.653 (0.024)***	0.305
Network among Reviewers	-0.080 (0.051)		0.002 (0.024)	

log (Average # Reviews per Reviewer)	-0.049 (0.013)***		-0.018 (0.006)***	
Intercept	0.565 (0.051)***	0.407	0.651 (0.023)***	0.316
Reviewers Recognition	0.036 (0.055)		0.014 (0.024)	
log (Average # Reviews per Reviewer)	-0.053 (0.015)***		-0.019 (0.007)***	
Intercept	0.516 (0.066)***	0.440	0.628 (0.029)***	0.365
Verified Reviews	0.078 (0.062)		0.036 (0.027)	
log (Average # Reviews per Reviewer)	-0.040 (0.016)**		-0.013 (0.007)*	
Intercept	0.583 (0.078)***	0.394	0.661 (0.034)***	0.308
Popularity Ranking	-0.001 (0.008)		-0.001 (0.003)	
log (Average # Reviews per Reviewer)	-0.050 (0.014)***		-0.018 (0.006)***	
Intercept	0.36 (0.072)***	0.634	0.613 (0.039)***	0.358
Sellers Ability to Respond to Reviews	0.192 (0.054)***		0.037 (0.029)	
log (Average # Reviews per Reviewer)	-0.015 (0.015)		-0.011 (0.008)	

Note: Std. Error in parentheses. * p <0.1, ** p <0.05, *** p <0.01.

Web Appendix 5: Overview #Reviewers per Platform and Comparison of Randomly Selected Reviewers vs. Full Sample of Unique Reviewers

This appendix provides information about the number of reviewers sampled from each platform and a comparison of key variables regarding the subsampled datasets.

Platform	#Reviewers Sampled	Number Full Sample (#unique reviewers)	Average Rating	# of Reviews of Reviewers	Average Number of Reviews of Products	Mean Age of Reviews
			(t-test all unique reviewers vs. random sample (n=1,000))			
Frag-Mutti	1,000 (subsample)	10,810	4.15 vs. 4.13 (p = 0.456)	20.73 vs. 16.18 (p = 0.031)	20.51 vs. 20.69 (p=0.794)	327.32 vs. 305.73 (p = 0.167)
Trust Pilot	500		-	-	-	-
BlaBlaCar	300		-	-	-	-
Amazon	1,000 (subsample)	21,176,522	4.16 vs. 4.17 (p=0.777)	3.90 vs. 3.5 (p=0.511)	646.42 vs. 612.53 (p=0.612)	1033.44 vs. 1020.89 (p=0.760)
Google Restaurants	1,000 (subsample)	172,343	4.35 vs. 4.35 (p = 0.329)	38.29 vs. 39.04 (p = 0.785)	530.92 vs. 512.02 (p = 0.032)	391.69 vs. 375.72 (p = 0.249)
Online Retailer	1,000 (subsample)	99,133	4.29 vs. 4.27 (p = 0.294)	1.61 vs. 1.66 (p=0.702)	236.29 vs. 232.99 (p = 0.677)	726.07 vs. 750.83 (p = 0.222)
Edmunds	1,000 (subsample)	143,139	4.44 vs. 4.43 (p= 0.396)	1.31 vs. 1.21 (p=0.061)	118.62 vs. 121.92 (p=0.480)	780.29 vs. 776.32 (p=0.982)
Booking.com	1,000 (subsample)	71,970[1]	4.40 vs. 4.41 (p= 0.632)	7.17 vs. 7.40 (p= 0.585)	909.54 vs. 851.85 (p = 0.034)	355.55 vs. 342.71 (p = 0.050)
Epinions	1,000 (subsample)	91,026	3.68 vs. 3.71 (p= 0.503)	11.96 vs. 12.39 (p= 0.855)	15.37 vs. 15.87 (p= 0.565)	447.12 vs. 472.63 (p= 0.235)
Yelp Restaurants	1,000 (subsample)	1,220,601	3.75 vs. 3.75 (p= 0.986)	26.99 vs. 27.67 (p= 0.780)	599.47 vs. 588.94 (p=0.757)	1610.50 vs. 1630.30 (p=0.611)
Yahoo! Movies	1,000 (subsample)	7,642	3.92 vs. 3.92 (p= 0.897)	28.97 vs. 29.73 (p= 0.712)	783.37 vs. 823.72 (p= 0.712)	No information on date
TripAdvisor	1,000 (subsample)	66,647	4.21 vs. 4.21 (p = 0.884)	45.68 vs. 44.75 (p=0.740)	2624.14 vs. 2615.93 (p= 0.969)	No information on date of first hotel review

[1] Approximated because we have no unique customer id in our data (#Number of Reviews/Average Number of Reviews of Reviewers).

Metacritic	1,000 (subsample)	3,650	3.27 vs. 3.22 (p = 0.073)	30.93 vs. 34.38 (p=0.407)	1315.17 vs. 1380.42 (p=0.367)	No information on date of first movie review
Rotten Tomatoes	100		-	-	-	-
Goodreads	200		-	-	-	-
IMDb	100		-	-	-	-
Netflix	1,000 (subsample)	480,189	3.64 vs. 3.65 (p=0.395)	209.25 vs. 206.21 (p=0.746)	62402.97 vs. 63756.71 (p=0.369)	851.25 vs. 865.47 (p=0.494)
MovieLens	1,000 (subsample)	259,138	3.80 vs. 3.78 (p=0.075)	94.17 vs. 87.48 (p=0.241)	31062.99 vs. 29392.15 (p=0.027)	2854.02 vs. 2829.12 (p=0.722)
RateBeer	1,000 (subsample)	23,423	3.63 vs. 3.64 (p= 0.443)	68.74 vs. 71.16 (p= 0.795)	857.41 vs. 919.18 (p=0.032)	1605.86 vs. 1658.49 (p= 0.167)
YourXpert	1,000 (subsample)	3,412	4.65 vs. 4.64. (p= 0.765)	1 vs. 1	4.82 vs. 4.14 (p=0.224)	46.94 vs. 37.18 (p=0.072)
Twitter Movies	1,000 (subsample)	45,882	4.05 vs. 4.04 (p=0.324)	11.98 vs. 11.33 (p=0.585)	583.47 vs. 577.70 (p=0.804)	No information on date of first movie review

Web Appendix 6: Robustness Random Effect Hierarchical Model

Extreme Scale Points as Polarity

In this analysis, we re-run the analysis in Table 3 in the main manuscript where we define polarity for platforms with scales longer than 5-point scales as only the extreme scale points (e.g., 1 and 10 for a 10-point scale). The results are consistent with those of Table 3 in the main manuscript.

DV: Polarity (1=extreme scale points, 0=otherwise)	
Polarity Self-Selection: Log(Number of Reviews per Reviewers)	-0.144 (0.014)***
Log(Number of Reviews per Product)	-0.018 (0.011)
Age of Platform	-0.088 (0.039)**
Business Model (Reference: Selling products or services)	
Transaction Fee	-1.953 (0.995)*
Information Platform	-0.909 (0.638)
Product Category (Reference: Products/Services)	
Travel/Restaurants	-0.536 (0.744)
Entertainment	1.040 (0.763)
Scale Points	-1.917 (0.570)***
Network among Reviewers	-0.060 (0.531)
Reviewers Recognition	0.154 (0.395)
Verified Reviews	0.568 (0.624)
Popularity Ranking	-0.019 (0.056)
Sellers Ability to Respond to Reviews	0.491 (0.747)
Constant	3.293 (1.430)**
N	17,200

Note: Std. Error in parentheses. * p <0.1, ** p <0.05, *** p <0.01.

Alternative Operationalization Metacritic and Yahoo! Movies

In this analysis, we re-run the analysis in Table 3 in the main manuscript but using an alternative operationalization for Metacritic and Yahoo! Movies as the original scales are not divisible by 5 (see Page 11). The results are consistent with those of Table 3 in the main manuscript.

DV: Polarity (1=extreme scale points, 0=otherwise)	
Polarity Self-Selection: Log(Number of Reviews per Reviewers)	-0.169 (0.013)***
Log(Number of Reviews per Product)	-0.011 (0.011)
Age of Platform	-0.063 (0.029)**
Business Model (Reference: Selling products or services)	
Transaction Fee	-2.162 (0.758)***
Information Platform	-0.832 (0.485)*
Product Category (Reference: Products/Services)	
Travel/Restaurants	-0.946 (0.564)*
Entertainment	0.648 (0.577)
Scale Points	-1.002 (0.434)**
Network among Reviewers	-0.158 (0.403)
Reviewers Recognition	0.012 (0.300)
Verified Reviews	0.540 (0.475)
Popularity Ranking	-0.029 (0.043)
Sellers Ability to Respond to Reviews	0.560 (0.568)
Constant	3.354 (1.090)***
N	17,200

Note: Std. Error in parentheses. * p <0.1, ** p <0.05, *** p <0.01.

Only Platforms with at Least 1,000 Reviews

To keep the number of reviews per platforms the same across platforms, in this analysis, we re-run the analysis in Table 3 in the main manuscript only for the 16 platforms for which we had a full sample of 1,000 reviews. The results are consistent with those of Table 3 in the main manuscript.

DV: Polarity	
Polarity Self-Selection: Log(Number of Reviews per Reviewers)	-0.157 (0.014)***
Log(Number of Reviews per Product)	-0.013 (0.012)
Age of Platform	-0.079 (0.031)**
Business Model (Reference: Selling products or services)	
Transaction Fee	-0.689 (0.876)
Information Platform	-0.069 (0.479)
Product Category (Reference: Products/Services)	
Travel/Restaurants	-0.841 (0.544)
Entertainment	0.072 (0.490)
Scale Points	-0.266 (0.448)
Network among Reviewers	0.464 (0.376)
Reviewers Recognition	-0.411 (0.307)
Verified Reviews	-0.386 (0.456)
Popularity Ranking	-0.068 (0.050)
Sellers Ability to Respond to Reviews	1.188 (0.554)**
Constant	2.807 (1.469)*
N	16,000

Note: Std. Error in parentheses. * p <0.1, ** p <0.05, *** p <0.01.

Subsample of 100 Reviews for All Platforms

To keep the number of reviews per platforms the same across platforms, in this analysis, we re-run the analysis in Table 3 in the main manuscript only with 100 reviews per platform for all 21 platforms. The results are consistent with those of Table 3 in the main manuscript.

DV Polarity	
Polarity Self-Selection: Log(Number of Reviews per Reviewers)	-0.19 (0.031)***
Log(Number of Reviews per Product)	-0.047 (0.025)*
Age of Platform	-0.048 (0.024)**
Business Model (Reference: Selling products or services)	
Transaction Fee	-2.209 (0.641)***
Information Platform	-1.41 (0.404)***
Product Category (Reference: Products/Services)	
Travel/Restaurants	-1.128 (0.46)**
Entertainment	0.631 (0.473)
Scale Points	-1.088 (0.366)***
Network among Reviewers	-0.262 (0.335)
Reviewers Recognition	0.394 (0.24)
Verified Reviews	0.36 (0.402)
Popularity Ranking	-0.038 (0.035)
Sellers Ability to Respond to Reviews	0.588 (0.473)
Constant	3.65 (0.894)***
N	2,100

Note: Std. Error in parentheses. * p <0.1, ** p <0.05, *** p <0.01.

Web Appendix 7: Analysis of Polarity Self-selection with Additional Proxies for Polarity Self-Selection Based on the Yelp Reviewers Survey

		Model 1	Model 2	Model 3	Model 4	Model 5
(1) Polarity Self-selection Proxy: Log(Reviews per Month)	Polarity	-0.041 (0.022)*		-0.059 (0.015)***		
Polarity Self-selection Proxy: Median Time Between Reviews (in 1,000s days)	Polarity	0.172 (0.119)			0.331 (0.094)***	
Polarity Self-selection Proxy: Standard Deviation Time Between Reviews (in 1,000s days)	Polarity	0.001 (0.06)				0.121 (0.050)*
Polarity Self-selection Proxy: Log(Reviews per Month)	Positive Imbalance	-0.038 (0.021)*		-0.055 (0.014)***		
Polarity Self-selection Proxy: Median Time Between Reviews (in 1,000s days)	Positive Imbalance	0.134 (0.113)			0.294 (0.090)***	
Polarity Self-selection Proxy: Standard Deviation Time Between Reviews (in 1,000s days)	Positive Imbalance	0.021 (0.058)				0.124 (0.047)**
(2) Composite Measure	Polarity		-0.060 (0.016)***			
	Positive Imbalance		-0.056 (0.015)***			
Intercept	Polarity	0.300 (0.034)***	0.291 (0.034)***	0.29 (0.033)***	0.359 (0.021)***	0.355 (0.026)***
	Positive Imbalance	0.351 (0.033)***	0.343 (0.032)***	0.343 (0.032)***	0.409 (0.020)***	0.399 (0.025)***
Number of Friends (in 1,000s)	Polarity	0.062 (0.211)	0.016 (0.207)	0.021 (0.206)	0.122 (0.209)	0.133 (0.216)
	Positive Imbalance	0.158 (0.201)	0.108 (0.197)	0.112 (0.196)	0.205 (0.200)	0.222 (0.204)
Number of Followers (in 1,000s)	Polarity	0.197 (4.910)	0.331 (4.926)	0.376 (4.906)	-1.205 (4.935)	-1.880 (5.090)
	Positive Imbalance	0.198 (4.687)	0.395 (4.686)	0.428 (4.67)	-1.094 (4.721)	-1.616 (4.803)
Number of Years Reviewer has received an Elite Badge	Polarity	-0.006 (0.013)	-0.001 (0.013)	-0.002 (0.013)	-0.019 (0.012)	-0.019 (0.012)
	Positive Imbalance	-0.008 (0.013)	-0.004 (0.013)	-0.004 (0.012)	-0.021 (0.011)*	-0.019 (0.012)
R^2 Polarity		0.236	0.211	0.218	0.195	0.141
R^2 Positive Imbalance		0.230	0.211	0.216	0.185	0.154
N		95	95	95	95	95

Note: Std. Error in parentheses. * p <0.1, ** p <0.05, *** p <0.01.

When including not only the number of reviews of reviewers as a proxy for self-selection but also the median inter-review time as well as its standard deviation we find that only the number of reviews has a (marginal) significant negative effect on polarity and positive imbalance. Moreover, adding median time between reviews and the standard deviation of time between reviews do not meaningfully increase the variance explained and even decrease the adjusted R^2 relative to the model with only number of reviews per reviewers as a proxy for self-selection (Polarity: R^2=0.218 vs. R^2=0.236, Positive Imbalance: R^2=0.216 vs. R^2= 0.230). Additionally, when using median and standard deviation of review time separately as proxies for self-selection, we find that all three are significant, however, the number of reviews explain the highest degree of variance in polarity and positive imbalance (Polarity: $R^2_{\text{number reviews}}$= 0.218, $R^2_{\text{median time}}$= 0.195, $R^2_{\text{std time}}$= 0.141; Positive imbalance: $R^2_{\text{number reviews}}$= 0.216, $R^2_{\text{median time}}$= 0.185, $R^2_{\text{std time}}$= 0.154). Finally, we build a composite measure by using the unstandardized regression coefficients from the regression of the three proxies on the self-selection measure from our Yelp survey (see Model 4 in Table 5 in main manuscript) as weights. In line with the results from the other models we do not find a substantial increase in the variance explained.

Web Appendix 8: Analysis of Polarity and Positive Imbalance with Additional Proxies for Polarity Self-Selection for the Yelp Dataset

Median Time Between Reviews as Proxy for Self-selection		
	Polarity	*Positive Imbalance*
Intercept	0.460 (0.001)***	0.544 (0.000)***
Median Time Between Reviews of Reviewers (in 1,000s days)	0.125 (0.002)***	-0.123 (0.002)***
Number of Years Elite Batch	-0.035 (0.000)***	0.011 (0.000)***
Number of followers (multiplied by 1,000)	-0.251 (0.018)***	-0.243 (0.015)***
Number of friends (multiplied by 1,000)	0.037 (0.002)***	0.057 (0.002)***
R^2	0.012	0.024
N	355,878	355,589
Standard Deviation Time Between Reviews as Proxy for Self-selection		
	Polarity	*Positive Imbalance*
Intercept	0.509 (0.001)***	0.524 (0.001)***
Standard Deviation Time Between Reviews of Reviewers (in 1,000s days)	-0.104 (0.001)***	0.017 (0.001)***
Number of Years Elite Batch	-0.035 (0.000)***	0.012 (0.000)***
Number of followers (multiplied by 1,000)	-0.217 (0.018)***	-0.250 (0.016)***
Number of friends (multiplied by 1,000)	0.027 (0.002)***	0.060 (0.002)***
R^2	0.012	0.028
N	355,878	355,589^

Note: Std. Error in parentheses. *** p <0.01. ^For reviewers that only wrote 3-star reviews, positive imbalance cannot be calculated.

Web Appendix 9: Stimuli and Study Instructions - Self-selection Experiment

Please take a minute and think about a **leisure book** that you have read, i.e., a book that you did not read for study purposes, and that you have written an online review for.

If you have never written an online review for any leisure book in the past, please think about the leisure book you would be **most likely to write a review** for.

[In last book (forced condition) condition replaced by:
Please take a minute and think about the last leisure book that you have read, i.e., a book that you did not read for study purposes.]

What is the name of the book?

_____

Please indicate the approximate date when you have read the book.

< 1 week	≥ 1 week and <1 month	≥ 1 month and <6 months	≥ 6 months and <1 year	≥ 1 year and <2 years	≥ 2 years
☐	☐	☐	☐	☐	☐

Now please imagine that you were to write a review about this book on an online website such as Amazon.com.

How would you rate the book?

I hate it	I don't like it	Its ok	I like it	I love it
				
☐	☐	☐	☐	☐

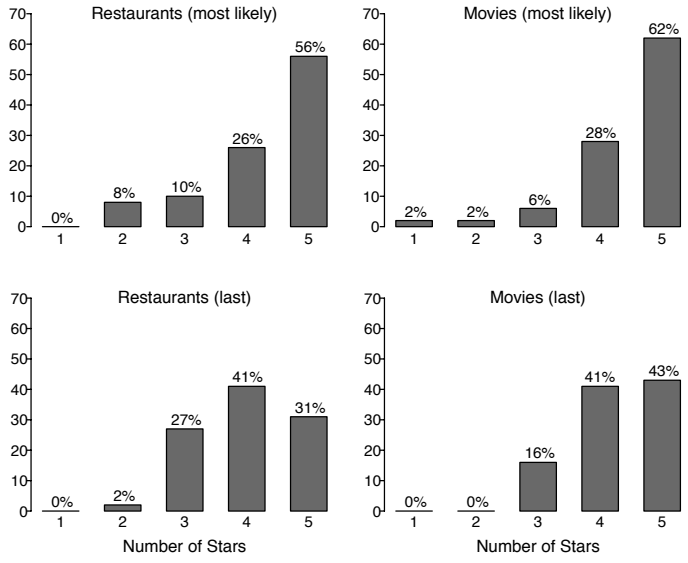
Randomization Check

Variable	Most Likely Condition	Last Condition	Difference
Gender	Proportion Female$_{Restaurants}$: 61.18% Proportion Female$_{Books}$: 57.50%	Proportion Female$_{Restaurants}$: 60.94% Proportion Female$_{Books}$: 55.13%	$\chi^2_{restaurants}$(1, N = 149) = 0.001, p = 0.976); χ^2_{books} (1, N = 158) = 0.090, p =0.764)
Age	Restaurants: M = 22.91, SD = 2.85 Books: M = 22.62, SD = 2.52	Restaurants: M = 22.72, SD = 2.56 Books: M = 22.34, SD = 2.24	Restaurants: t(147) = 0.422, p = 0.6736 Books: : t(156) = 0.751, p = 0.454
# Restaurants [Books] Reviewed in Past 12 Months	Restaurants: M = 0.52, SD = 2.56 Books: M = 0.21, SD = 0.55	Restaurants: M = 0.20, SD = 0.67 Books: M = 0.06, SD = 0.24	Restaurants: t(147) = 0.961, p = 0.338 Books: t(156) = 2.175, p = 0.031
# Restaurants [Books] visited [read] in Past Month	Restaurants: M = 3.94, SD = 2.96 Books: M = 0.95, SD = 0.82	Restaurants: M = 3.57, SD = 2.28 Books: M = 1.27, SD = 1.59	Restaurants: t(147) = 0.820, p = 0.414 Books: t(156) = -1.514, p = 0.132

Web Appendix 10: Replication of the In-Class Self-selection Experiment with MTurk Workers

We replicate the in-class experiment with an online experiment using MTurk workers. In this experiment we replaced books by movies but kept restaurants in order to compare products versus services (n = 100; 61% females; $n_{self\text{-}selection} = 50$; $n_{random} = 50$). We find no significant order effect.[2] The results are consistent with the results from the in-class experiment in the main manuscript (see Figure 4). The distribution between the two conditions confirms that the restaurants [movies] in self-selection ("most likely") condition exhibit an extreme distribution, while the forced ("last") restaurants [movies] condition exhibit a far less extreme distribution of reviews. This is confirmed when (1) comparing the proportion of 5 and 1 star reviews between the two conditions (Restaurant: χ^2 (1, N = 100)= 6.8951, p = 0.0086; Movies: χ^2 (1, N = 100) = 4.8575, p = 0.0275) and also (2) when comparing the two overall distributions (Restaurants: Fisher exact test p = 0.0135, Movies: Fisher exact test p = 0.0611). For positive imbalance we find no significant difference between the two conditions (Restaurant: $\chi^2(1, N = 100)$ = 1.2224, p = 0.2689; Movies: χ^2 = 1.7853 (1, N = 100), p = 0.1815).

Review Distributions Experiment MTurk



[2] Restaurant: Fisher's Exact Test, $p_{Self\text{-}selection} = 0.7124$; $p_{Random} = 0.1655$; Movie: Fisher's Exact Test, $p_{Self\text{-}selection} = 1$; $p_{Random} = 0.3344$.

Web Appendix 11: Robustness Check Experiment Self-Selection - Controlling for Time Since Purchase

One possible confound in the experimental design used in the experiment in the main text is that the two conditions imply a different time frame. Whereas the no self-selection reviews are written for a recent experience the self-selected reviews can be for an experience that occurred long-time ago. This, in and of itself, could lead to differences in the reported reviews. To mitigate that concern, we ran a follow-up experiment in which we force all participants to rate their "last experience" for a restaurant [book], but we ask participants to indicate their likelihood to write a review for the respective restaurant [book] rated. We then split participants into two groups based on whether their likelihood of rating the restaurant [book] was greater than 3 on a 5-point scale ($n_{Restaurants} = 25$; $n_{Books} = 35$) or lower or equal to 3 ($n_{Restaurants} = 37$; $n_{Books} = 41$). Thus, we hold the time of the experience constant and separate ratings based on the likelihood to review. Based on polarity self-selection, we expect that restaurants/books with higher likelihood to review (self-selection) will exhibit a more extreme distribution than those with a low likelihood to review.

Results. The differences between the two distributions are statistically significant when (1) comparing the proportion of 1 and 5 star reviews between the two conditions (Restaurant: χ^2 (1, N=62) = 19.212, $p < 0.001$; Books: χ^2 (1, N=76) = 25.621, $p < 0.001$), with the reviews with higher likelihood to review having higher proportion of 5 and 1 star review, relative to those with a low likelihood to review, and (2) when comparing the two overall distributions (Restaurants: Fisher exact test $p < 0.001$, Books: Fisher exact test $p = 0.0036$). These results further support the previous finding that polarity self-selection is the underlying driver of the extreme distribution, ruling out potential confounds from the first experiment (1) time since consumption as well as (2) differences in the mental process between the conditions.