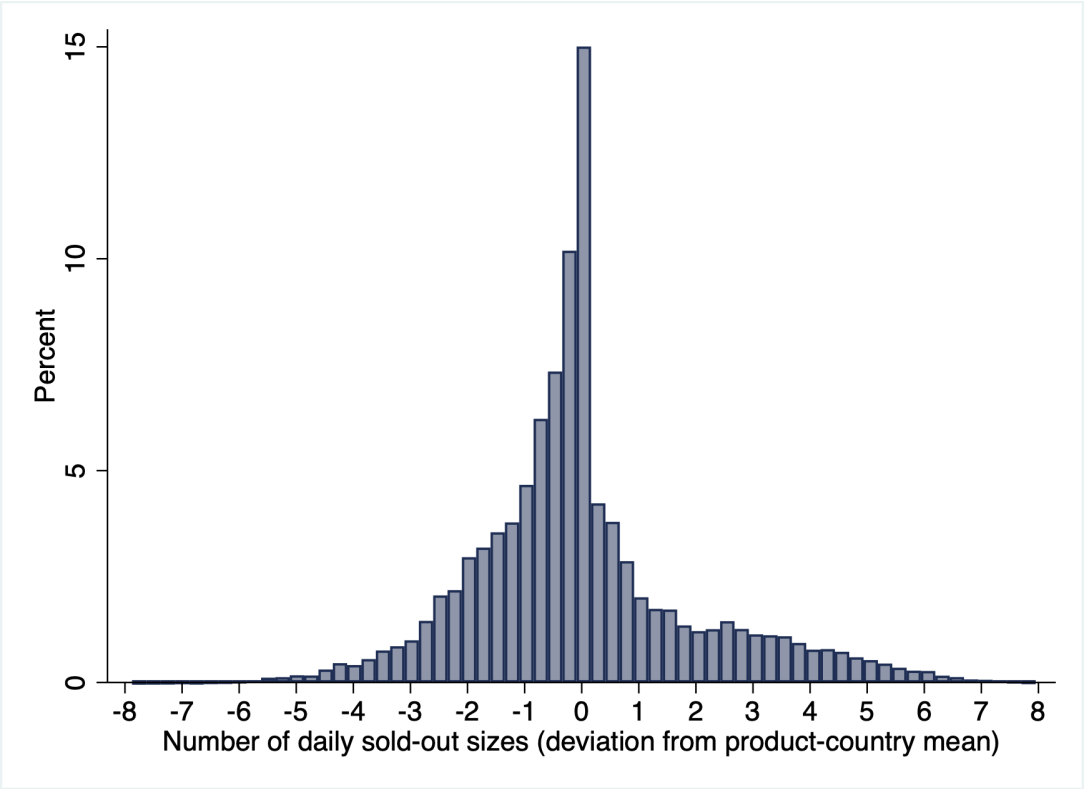
Supplementary Online Materials

Do Consumers Respond to Social Movements? Evidence from Gender-Stereotypical Purchases After #MeToo

Clément Bellet, David Dubois, and Frédéric Godart

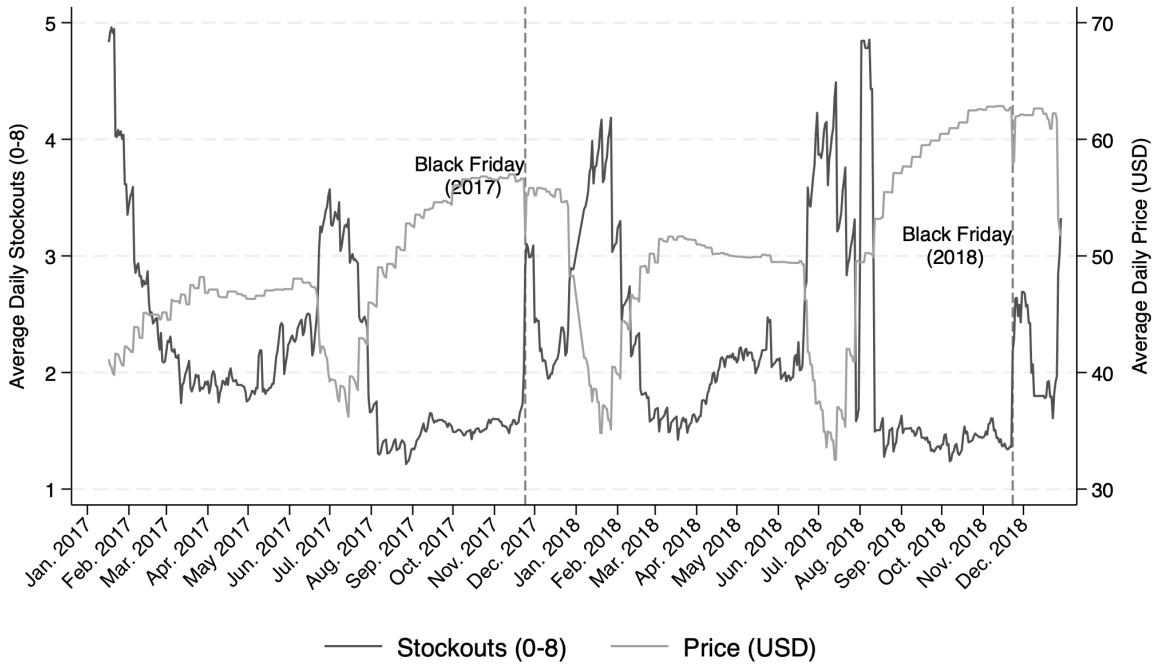
A Stockout and Price Data

Figure S1: Average Daily Number of Stock-Out Sizes (Within Product-Country)



Notes: The figure shows the extent to which the average daily number of stock-out sizes varies within product-country cells over time in 2017. Each woman shoe product has 8 possible sizes. The graph shows the residuals from an OLS regression of stock-out sizes on product-country fixed effects. An observation is a product-country-day residual from each regression.

Figure S2: Average Daily Number of Stockout Sizes and Price per Shoe (2017–2018)

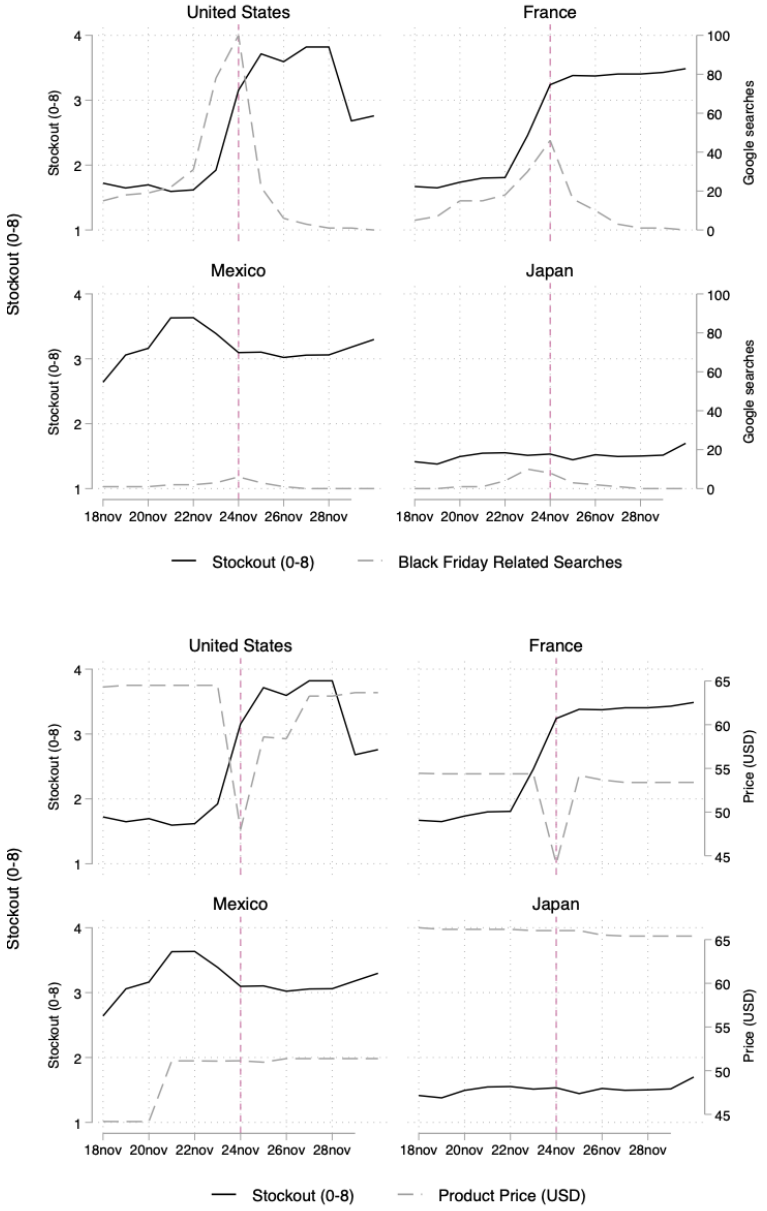


Notes: The figure plots the daily trend in the average number of stockout sizes per shoe (dark line) and average shoe price in USD (grey line) between 2017 and 2018. Each pair of shoes has 8 possible sizes, so the stockout measure fluctuates between 0 and 8.

Table S1: Summary Statistics (Product-Level Sample: 2017-2018)

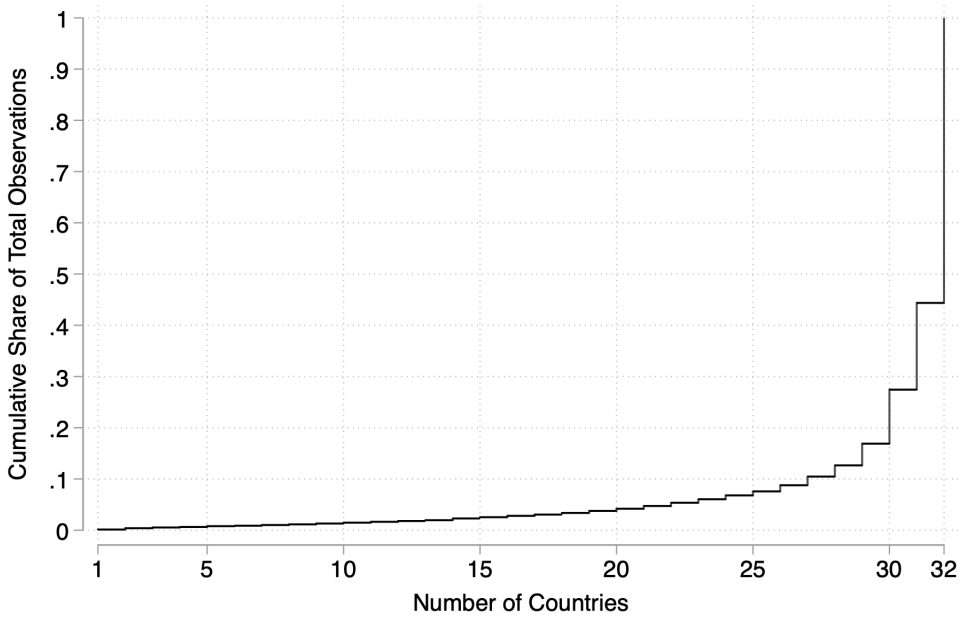
Variable	Obs	Mean	Std. Dev.	Min	Max
Daily stockouts	4989992	2.11	2.71	0	8
Daily price (U.S. $)	4973934	51.66	24.51	4.21	229
Heels and pumps	1718	.3	.46	0	1
Boots	1718	.29	.45	0	1
Lace-up shoes	1718	.01	.1	0	1
Platforms and wedges	1718	.07	.25	0	1
Sandals	1718	.12	.32	0	1
Sneakers, clogs and loafers	1718	.22	.41	0	1
Pink	1718	.05	.23	0	1
Red	1718	.03	.18	0	1
Black	1718	.44	.5	0	1
Blue	1718	.07	.26	0	1
Brown	1718	.07	.26	0	1
Green	1718	.04	.2	0	1
Neutral	1718	.24	.43	0	1
Yellow	1718	.05	.22	0	1
Heel height (inches)	1718	1.84	1.3	0	5
Heel height $\geq$ 1 inch	1718	.74	.44	0	1
Heel height $\geq$ 3 inch	1718	.27	.44	0	1

Figure S3: Average Country-Level Daily Stock-Out Rate vs. "Black Friday" Daily Searches: Illustrative Example Within Four OECD Countries (18-29 November 2017)



Notes: The solid dark line plots daily stockout rates for women shoes across four countries over the 18th-29th of November 2017. The dashed grey line plots the corresponding country-level daily search rates on the "Black Friday" topic from Google Trends (upper panels) and product price (lower panel) over the same period. The data is normalized by Google with 100 (USA, 24th of November) corresponding to the highest search volume measured over the period considered and for the four countries considered. The volume of "Black Friday" searches in the United States and in France around November 24th 2017 translates into rise in the shoe stock-out rate. In Mexico, the stockout rate peaked earlier, as "El Buen Fin" (its "Black Friday" equivalent) also occurs a week before "Black Friday." In Japan, which does not have a "Black Friday" for fashion goods, the stock-out rate remained flat around that entire period.

Figure S4: Cumulative Share of Observations by Number of Countries



Notes: Cumulative share of total product-week observations six weeks before (versus six weeks after) #MeToo as a function of country-level availability. For instance, in 85% of cases, the same product in a given week is available in at least 30 different OECD countries.

Table S2: Number of Observation per Country (12-Weeks Panel)

	Freq.	Percent
Austria	3851	3.200
Belgium	3861	3.209
Canada	3547	2.948
Czech Republic	3840	3.191
Denmark	3888	3.231
Estonia	3870	3.216
Finland	3791	3.150
France	3730	3.100
Germany	3830	3.183
Greece	3805	3.162
Hungary	3830	3.183
Iceland	3715	3.087
Ireland	3783	3.144
Israel	3772	3.135
Italy	3857	3.205
Japan	3080	2.560
Latvia	3886	3.229
Lithuania	3892	3.234
Mexico	2984	2.480
Netherlands	3904	3.244
Norway	3886	3.229
Poland	3893	3.235
Portugal	3872	3.218
Slovakia	3870	3.216
Slovenia	3913	3.252
South Korea	3856	3.204
Spain	3860	3.208
Sweden	3934	3.269
Switzerland	3799	3.157
Turkey	3138	2.608
UK	3841	3.192
USA	3758	3.123
Total	120336	100.00

Notes: Observations are at the product-week-country level over 12 weeks (6 weeks before and 6 weeks after #MeToo)

B International Survey of Product Ratings

We survey an international sample of 1,200 respondents from six OECD countries (Italy, France, Portugal, USA, UK and Mexico). Respondents were recruited on Prolific (`https://www.prolific.co/`) between February and April 2023 using gender quotas, from a set of OECD countries with a sufficiently large number of respondents registered as citizens of those countries. A total of 200 respondents per country was surveyed. Respondents were paid 0.60 euro to complete the survey, for an average completion time of 6.5 minutes per respondent.

Respondents were shown product pictures similar to the ones displayed in Figure 1 and asked to rate the style of these products on a gendered scale by answering the following question: *How would you describe this style of shoes? Very masculine (1), Masculine (2), Somewhat masculine (3), Neutral (4), Somewhat feminine (5), Feminine (6), Very feminine (7).* Each respondent rated a random sample of 30 product pictures among a pool of potential products. For a quarter of respondent, we use a different scale to rate products, namely: *This style of shoe would catch the attention of men if it were worn by a woman. Fully disagree (1), Disagree (2), Somewhat disagree (3), Neutral (4), Somewhat agree (5), Agree (6), Fully agree (7).* This gives us a total of 27,000 product ratings for the stereotypical scale and 9,000 ratings for the male attention scale.

The set of potential products was determined through a fractional factorial design covering all feasible combination of markers, based on the available attribute levels listed in Table S1 for type of shoes (six levels), shoe colors (8 levels) and heel height (3 levels: 0-1 inch, 1-3 inches, more than 3 inches).[50] Product pictures were taken directly from the women's wear section of the retailer, or if needed, on a competing website with similar pictures. The pictures did not display any brand logo, and respondents were not informed about the gender target (e.g. men or women) or the retailer or brand from which those pictures were taken.

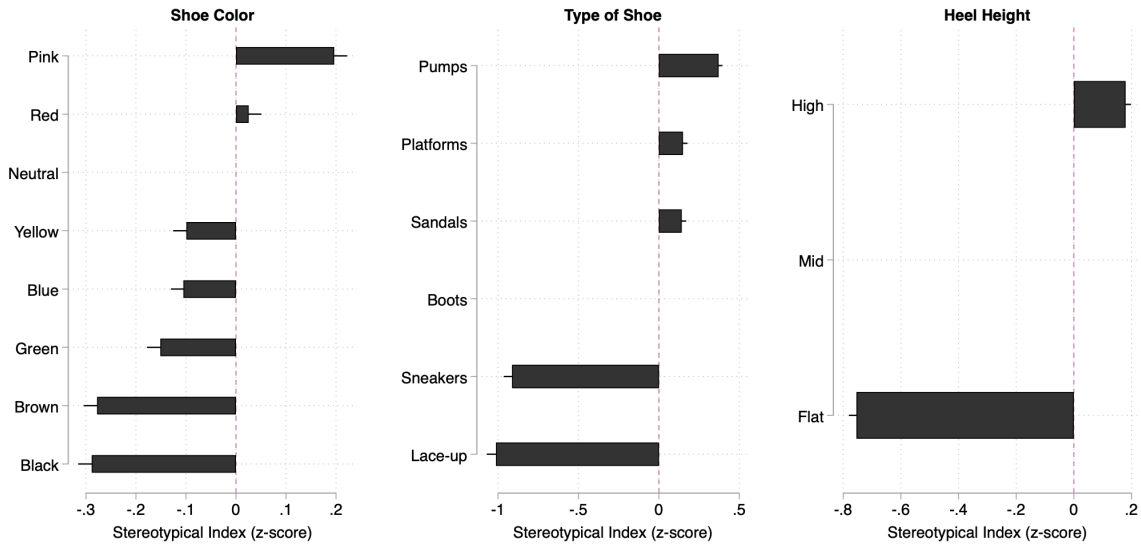
We then estimate the following regression:

$$SI_{i,j} = \alpha_j Color_j + \beta_j Type_j + \gamma_j Heel_j + X_j + u_i + \epsilon_{i,j} \tag{4}$$

where $SI_{i,j}$ captures the (stereotypical) gender association of product j by respondent i (on a scale from 1 to 7); $Color_j$ is a vector of seven color dummies ("neutral" is the excluded attribute level); $Type_j$ a vector of five product type dummies ("boots" is the excluded attribute level) and $Heel_j$ a vector of two heel height dummies ("mid-heels" is the excluded attribute level). We also include a set of product-level controls (log of product price and retailer dummy) and respondent fixed effects u_i. Robust standard errors are clustered at the respondent level.

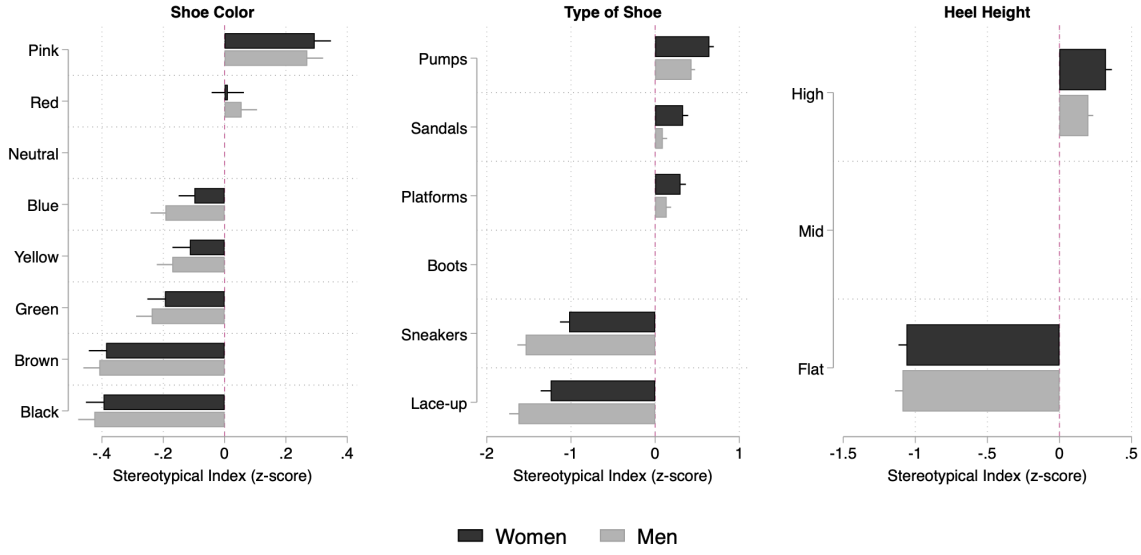
[50] Given that sneakers and lace-ups do not generally have heels higher than one inch, and that platforms and wedges must have heels, the complete set of product pictures amounts to 104 (i.e. $6 \times 8 \times 3$ - $2 \times 8 \times 2$ - $8 = 104$).

Figure S5: Decomposition of Stereotypical Index



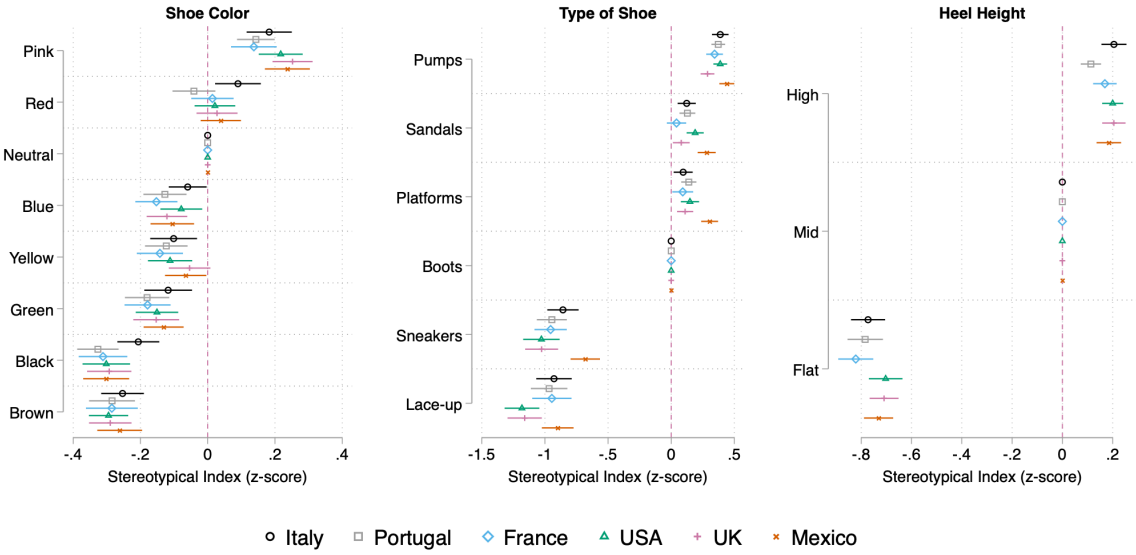
We report the estimated parameters on product marker dummies using OLS regression model (4), where the gender scale is used as the dependent variable. We use 27,000 ratings of product pictures from respondents' answers to the question: "How would you describe this style of shoes? Very masculine (1), Masculine (2), Somewhat masculine (3), Neutral (4), Somewhat feminine (5), Feminine (6), Very feminine (7)."

Figure S6: Decomposition of Stereotypical Index (Heterogeneity by Gender)



We report the estimated parameters on product marker dummies using OLS regression model (4) estimated by respondents' gender, where the gender scale is used as the dependent variable. We use 27,000 ratings of product pictures from respondents' answers to the question: "How would you describe this style of shoes? Very masculine (1), Masculine (2), Somewhat masculine (3), Neutral (4), Somewhat feminine (5), Feminine (6), Very feminine (7)."

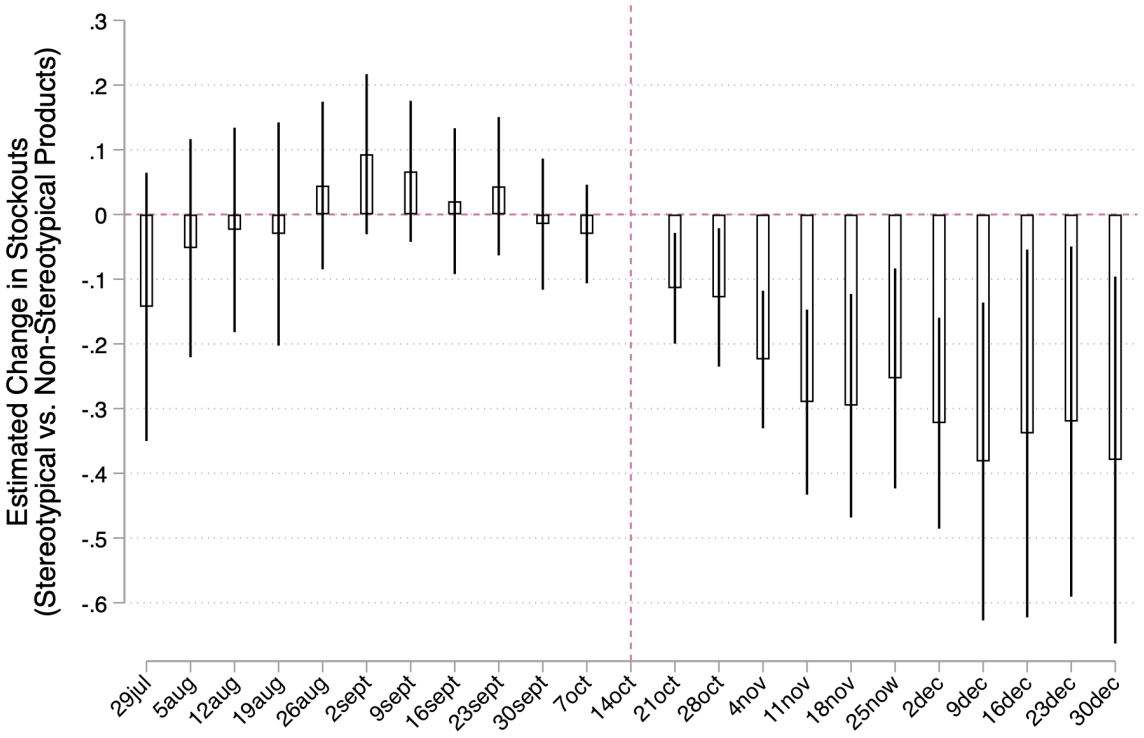
Figure S7: Decomposition of Stereotypical Index (Country-Level Heterogeneity)



We report the estimated parameters on product marker dummies using OLS regression model (4) estimated by respondents' country of citizenship, where the gender scale is used as the dependent variable. We use 27,000 ratings of product pictures from respondents' answers to the question: "How would you describe this style of shoes? Very masculine (1), Masculine (2), Somewhat masculine (3), Neutral (4), Somewhat feminine (5), Feminine (6), Very feminine (7)."

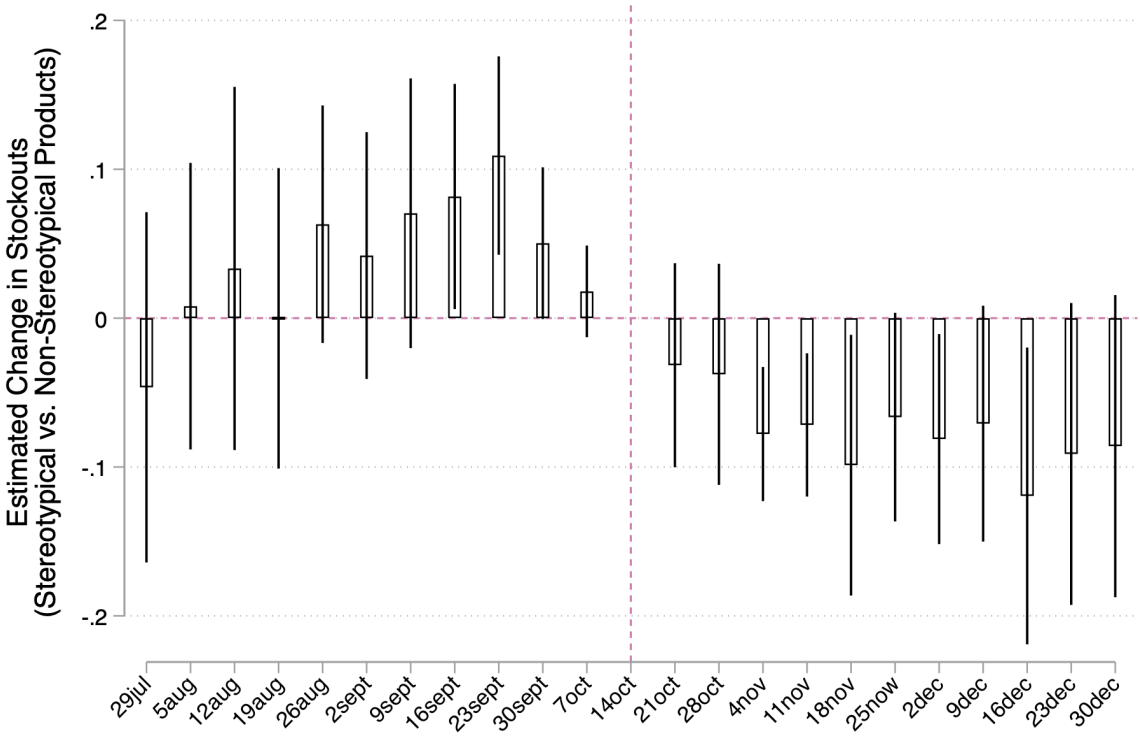
C Other Robustness Tests

Figure S8: Change in Weekly Stockouts of Gender-Stereotypical (vs. Non-Stereotypical) Products Before (vs. After) 15 October 2017 in Exposed (vs. Unexposed) Countries



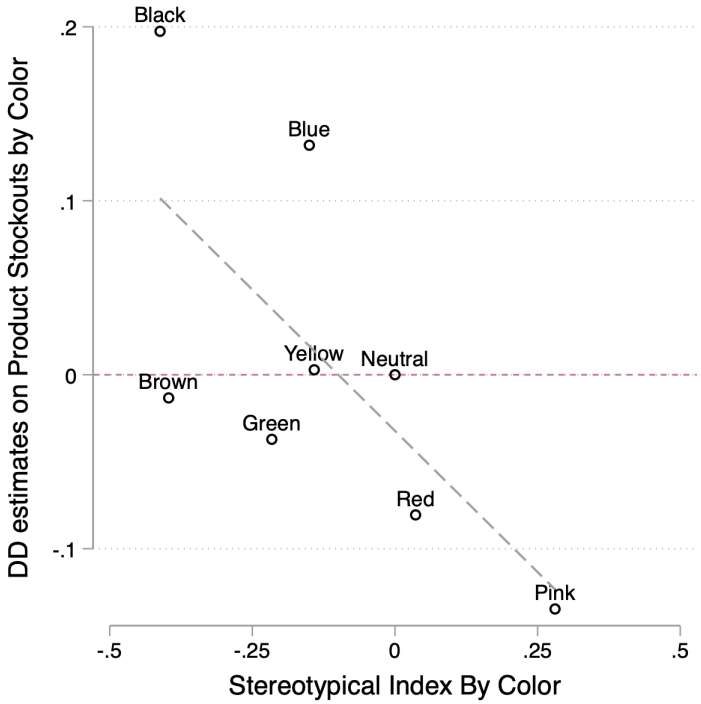
Notes: Weekly β_τ coefficients estimated using our main specification from Poisson-FE model (2), reported with 95% confidence intervals. Robust standard errors clustered at the country level are in parentheses. The model is estimated over a 24-weeks period around 15 October 2017, with the week before 15 October 2017 ($\tau = 0$) used as the reference week.

Figure S9: Change in Weekly Stockouts of Pink/Red (vs. Other) Products Before (vs. After) 15 October 2017 in Exposed (vs. Unexposed) Countries



Notes: Weekly β_τ coefficients estimated using our main specification from Poisson-FE model (2), reported with 95% confidence intervals. The data comprise all footwear, lingerie, dresses and bags (1793 products). Color pink or red is used to operationalize stereotypical products. Robust standard errors clustered at the country level are in parentheses. The model is estimated over a 24-weeks period around 15 October 2017, with the week before 15 October 2017 ($\tau = 0$) used as the reference week.

Figure S10: Impact of #MeToo vs. Stereotypical Index by Color (DD coefficients)



Notes: Figure plots difference-in-differences estimates on stockouts by shoe color using our main specification against their respective stereotypical index (SI) estimated from the international survey of product ratings.

Table S3: Alternative Clustering of Standard Errors

	# Stockouts					
	(1)	(2)	(3)	(4)	(5)	(6)
Post × Stereotypical	0.015	0.015	0.015	0.015	0.015	0.014
	(0.047)	(0.112)	(0.027)	(0.114)	(0.113)	(0.049)
Post × Stereotypical × #MeToo	-0.255***	-0.255***	-0.255***	-0.255***	-0.255***	-0.253***
	(0.055)	(0.048)	(0.034)	(0.042)	(0.051)	(0.055)
Observations	107,912	107,912	107,912	107,912	107,912	757,953
Fixed Effects:						
Product x Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Country x Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustering of Standard Errors:						
Country	✓	—	—	✓	—	✓
Product	—	✓	—	✓	✓	—
Country-Week	—	—	✓	—	✓	—
Day	—	—	—	—	—	✓

Notes: Within-product DDD estimator, estimated using Poisson quasi-maximum likelihood on weekly product-level stockouts for 411 products sold in 32 OECD countries 6 weeks before and 6 weeks after 15 October 2017. All models control for the log of product price and include product-country and country-week fixed effects. The Table reports results clustering standard errors on countries (column 1), products (column 2), or country-week (column 3) and using two-way clustering on products and countries (column 4), or products and country-week (column 5). Column 6 relies on the daily stockout sample and clusters standard errors on countries and days. Cluster robust standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table S4: Time Window Around #MeToo Movement

	(1) +/- 6 weeks	(2) +/-3 weeks	(3) +/- 9 weeks	(4) +/-12 weeks	(5) +/-15 weeks
Post × Stereotypical	0.015	0.062	-0.127*	-0.166**	-0.109
	(0.047)	(0.038)	(0.066)	(0.075)	(0.082)
Post × #MeToo × Stereotypical	-0.255***	-0.153***	-0.304***	-0.297***	-0.327***
	(0.055)	(0.043)	(0.078)	(0.090)	(0.100)
Observations	107,912	42,584	160,008	200,011	249,015
Product x Country FEs	Yes	Yes	Yes	Yes	Yes
Country x Week FEs	Yes	Yes	Yes	Yes	Yes

Notes: Within-product DDD estimator, estimated using Poisson quasi-maximum likelihood on weekly product-level stockouts. All models control for the log of product price and include product-country and country-week fixed effects. We vary the time window used to estimate the effect from 6 weeks before and after to 3 weeks, 9 weeks, 12 weeks and 15 weeks before and after, keeping the true starting date constant. Robust standard errors clustered at the country level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table S5: OLS-FE Estimates

	Poisson-FE	OLS-FE			
	(1) # Stockouts	(2) ihs(Stockouts)	(3) Frac(Stockouts)	(4) AnyStockout=1	(5) Stockout=1
Post × Stereotypical	0.015	-0.002	0.006	0.000	0.007
	(0.047)	(0.035)	(0.011)	(0.015)	(0.011)
Post × Stereotypical × #MeToo	-0.255***	-0.167***	-0.051***	-0.080***	-0.052***
	(0.055)	(0.040)	(0.014)	(0.017)	(0.014)
Observations	107,912	120,149	120,149	120,149	6,834,056
Product x Country FEs	Yes	Yes	Yes	Yes	—
Country x Week FEs	Yes	Yes	Yes	Yes	Yes
Product x Size x Country FEs	—	—	—	—	Yes
Time Unit	Weekly	Weekly	Weekly	Weekly	Daily

Notes: Poisson-FE and OLS-FE estimates of the influence of #MeToo exposure on stockouts of stereotypical products. We focus on the six weeks following October 15, 2017, compared to the six weeks before. The data comprise 411 products available in 32 OECD countries. Column 1 estimates a Poisson-FE model on the weekly number of out-of-stock sizes. Column 2-5 estimate OLS-FE models. Column 2 uses the inverse hyperbolic sine transformation (IHS) of the stockout variable as dependent variable, column 3 the fraction of out-of-stock sizes, and column 4 a binary dummy variable indicating whether a particular product experienced at least one out-of-stock size in a given week. Column 5 estimates a linear probability model of out-of-stock size at the country-product-size level. All models control for the log of product price. Robust standard errors clustered at the country level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table S6: #MeToo Effect (Daily Data)

	(1)	(2)	(3)	(4)	(5)	(6)
Ln(product price)				-2.594***	-2.480***	-2.284***
				(0.180)	(0.177)	(0.163)
Post	0.186***	0.242***	0.182***	0.130***	0.050	
	(0.040)	(0.023)	(0.042)	(0.036)	(0.044)	
Post × #MeToo	0.055		0.090*	0.093**	0.061	
	(0.045)		(0.049)	(0.042)	(0.049)	
Post × Stereotypical		-0.144***	0.032	0.035	0.024	0.014
		(0.035)	(0.048)	(0.048)	(0.045)	(0.047)
Post × Stereotypical × #MeToo			-0.257***	-0.257***	-0.258***	-0.253***
			(0.057)	(0.057)	(0.054)	(0.055)
Observations	757,960	757,960	757,960	757,953	757,953	757,953
Product x Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Country x Month FEs	—	—	—	—	Yes	—
Country x Week FEs	—	—	—	—	—	Yes

Notes: Poisson-FE estimates of the influence of #MeToo exposure on daily stockouts of stereotypical products. We focus on the six weeks following October 15, 2017, compared to the six weeks before. The data comprise 411 products available in 32 OECD countries. All models control for product-country fixed effects. Column 5 includes country-month fixed effects. Column 6 adds country-week fixed effects. Unexposed countries belong to the bottom tier countries in terms of #MeToo exposure index. Stereotypical products are defined as having at least two stereotypical markers of femininity. Robust standard errors clustered at the country level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

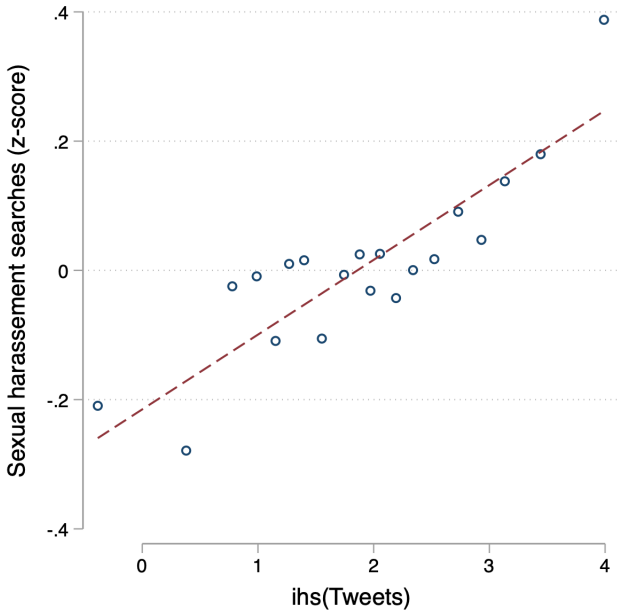
Table S7: Seasonality Tests

	Men Sample (2017)	Women Sample (2018)	Women Sample (2017)	
	(1)	(2)	(3)	(4)
Post x Stereotypical Index (z-score)	-0.325***	-0.076***	-0.038*	-0.045***
	(0.038)	(0.021)	(0.019)	(0.017)
Post x #MeToo x Stereotypical Index (z-score)	0.016	0.006	-0.085***	-0.064**
	(0.046)	(0.036)	(0.029)	(0.026)
Observations	49,092	83,167	107,912	102,630
Product x Country FEs	Yes	Yes	Yes	Yes
Country x Week FEs	Yes	Yes	Yes	Yes
Product x Country Lin. Trends	—	—	No	Yes

Notes: Poisson-FE estimates of the influence of #MeToo exposure on weekly stockouts of stereotypical products. All models control for the log of product price and include product-country and country-week fixed effects. Column 1 replicates our main specification within the sample of men's shoes, using 252 men's products available in the six weeks following October 15, 2017, compared to the six weeks before. Column 2 replicates our main specification on women's shoes but we focus on the six weeks following October 15, 2018, compared to the six weeks before. Column 3 is our main specification. Column 4 adds product-country linear time trends based on the pre-period. Robust standard errors clustered at the country level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

D Potential Channels and Behavioral Mechanisms

Figure S11: Within-Country Correlation Between Daily #MeToo Tweets and Web Searches on "Sexual Harassment" Topic



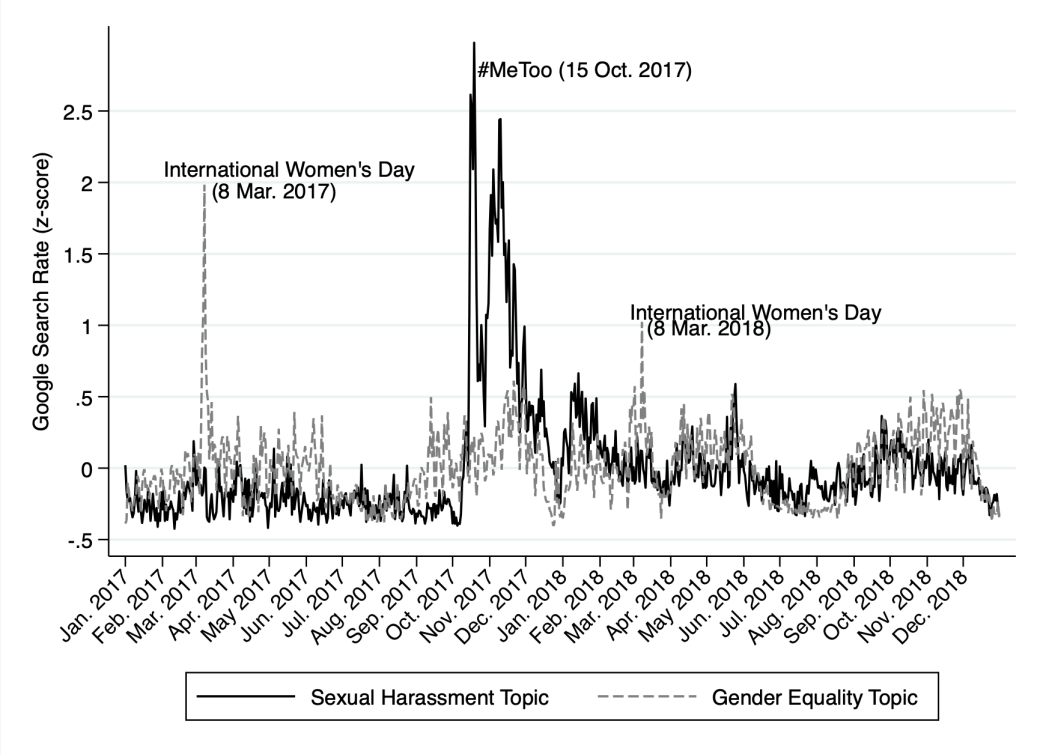
Notes: Figure shows binned scatter plots of the relationship between daily #MeToo tweets and sexual harassment related searches on Google, adjusting for country and date fixed effects.

Table S8: Heterogeneity by Product-Country Availability

	All Products	Never Removed	Never Added	Never Changed
	(1)	(2)	(3)	(4)
Post × Stereotypical	0.015	-0.006	-0.003	-0.026
	(0.047)	(0.046)	(0.047)	(0.047)
Post × Stereotypical × #MeToo	-0.255***	-0.236***	-0.228***	-0.208***
	(0.055)	(0.053)	(0.052)	(0.052)
Observations	107,912	104,878	84,798	81,972
Product x Country FEs	Yes	Yes	Yes	Yes
Country x Week FEs	Yes	Yes	Yes	Yes

Notes: Poisson-FE estimates of the influence of #MeToo exposure on daily stockouts of stereotypical products. We focus on the six weeks following October 15, 2017, compared to the six weeks before. Column 1 comprises 411 products available in 32 OECD countries. Column 2 only keeps products that were never removed over the period. Column 3 only keeps products that were never added. Column 4 only keeps products that were neither removed nor added over the period. All models control for the log of product price and include product-country and country-week fixed effects. Robust standard errors clustered at the country level are in parentheses $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure S12: Daily Web Searches, "Sexual Harassment" and "Gender Equality" Topics (Average Across 32 OECD Countries, 2017–2018)



Notes: The figure plots daily search volumes on sexual harassment and gender equality topics for the average OECD country in the dataset. Each time series is normalized using z-scores. Data are collected from Google Trends for each of the 32 countries between 1 January 2017 and 31 December 2018.

Table S9: Full vs. Partial OECD Availability

	Baseline (All Observations)	Full Availability (All 32 Countries)	Partial Availability (>26 Countries)
	(1)	(2)	(3)
Post × Stereotypical	0.015	0.016	-0.007
	(0.047)	(0.036)	(0.044)
Post × #MeToo × Stereotypical	-0.255***	-0.310***	-0.269***
	(0.055)	(0.046)	(0.056)
Observations	107,912	51,042	102,223
Product x Country FEs	Yes	Yes	Yes
Country x W FEs	Yes	Yes	Yes

Notes: Poisson-FE estimates of the influence of #MeToo exposure on daily stockouts of stereotypical products. We focus on the six weeks following October 15, 2017, compared to the six weeks before. Column 1 comprises 411 products available in 32 OECD countries. Column 2 only keeps products that were available in all 32 OECD countries over the period. Column 3 only keeps products that were available in more than 26 OECD countries over the period. All models control for the log of product price and include product-country and country-week fixed effects. Robust standard errors clustered at the country level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table S10: Gender-Stereotypical vs. Male Attention Scales (Robustness)

	(1)	(2)
Post × #MeToo × Stereotypical Index (z-score)	-0.085***	
	(0.029)	
Post × #MeToo × MenAttention Index (z-score)		-0.031**
		(0.013)
Observations	107,912	107,912
Product x Country FEs	Yes	Yes
Country x Week FEs	Yes	Yes

Notes: Poisson-FE estimates of the influence of #MeToo exposure on daily stockouts of stereotypical products (column 1) and products likely to catch men's attention (column 2). Three-way coefficient from the main specification reported. We focus on the six weeks following October 15, 2017, compared to the six weeks before. The data comprise 411 products available in 32 OECD countries. All models control for the log of product price and include product-country and country-week fixed effects. Robust standard errors clustered at the country level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

E Online Survey Experiment

The online survey experiment was conducted via Qualtrics in January 2021 on about 1,000 women representative of the U.S. population. It targeted adult women and included age and location quotas (regions). Responses were anonymous and each participant was paid 4.15 euros to complete the online survey. The average completion time was 16 minutes.

The survey included three tasks, ostensibly framed as separate studies. In the first task ("Event memory"), participants were asked to recall a particular event and describe the experience as it happened to them. Participants were assigned to one of four conditions: a baseline (control) condition, a #MeToo condition, an identity threat condition, and an identity empowerment condition. In the baseline condition, participants recalled and wrote about a particular episode in which they went to the grocery store. Participants in the threat and empowerment conditions had to recall a particular episode in which they felt, respectively, threatened or empowered as women. Participants in the #MeToo condition were asked to describe how these events made them feel, what they thought about them, and to what extent they got involved. Sub-section E.1 below replicates each condition.

In the second task, framed as a study about consumer preferences, participants indicated their preferences for different shoe products. A first choice-based question asked them to imagine that they were browsing different pairs of shoes online and planning to buy a new pair of heeled shoes. They were invited to select the pair they would prefer buying from a set of five options with increasing heel height (from 2 to 6 inches). A follow-up question asked how likely they would be to buy a pair of shoes from a set of 8 colors. Lastly, we probed how much they liked to wear heels in general. Sub-section E.2 below replicates each question used in the second task.

The final task, framed as a personality study, comprised more general questions, including a set of questions aimed to test the role of factors potentially underlying the #MeToo effect, namely, sensitivity to threat and sense of power, using well-established scales composed of 8 to 10 questions each. Sub-section E.3 below replicates each question used to construct both scales.

Table S11: Average Treatment Effects (OLS)

	(1) Composite Index (z-score)	(2) Like heels (=1)	(3) Heel height (inches)	(4) Like pink (=1)	(5) Like red (=1)	(6) Threat (z-score)	(7) Power (z-score)
Treatment condition:							
MeToo	-0.126**	-0.079***	-0.065	-0.052**	-0.011	0.225**	-0.278***
	(0.057)	(0.029)	(0.091)	(0.024)	(0.035)	(0.088)	(0.087)
Threat	0.064	-0.014	0.049	0.029	0.064*	0.193**	-0.198**
	(0.066)	(0.032)	(0.097)	(0.029)	(0.037)	(0.088)	(0.089)
Empowerment	-0.022	-0.025	0.017	-0.027	0.026	0.035	0.069
	(0.064)	(0.032)	(0.095)	(0.027)	(0.037)	(0.093)	(0.094)
Observations	984	984	984	984	984	984	984

Notes: Average Treatment Effect (ATE) of each treatment condition (vs. baseline condition) on respondents' (1) preferences for stereotypical attributes (composite index), (2) preference for heels (0,1), (3) choice of heel height (inches), (4) preference for pink shoes (0,1) and (5) preference for red shoes (0,1). All dependent variables are normalized using z-scores. Columns (6) and (7) use respondents' sensitivity to threat and sense of power as dependent variables. All regressions control for the log of time spent doing the survey. Robust standard errors in parentheses $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

We focus on four main stated measures of stereotypical preferences as our dependent variables: participants' preferred choice of heel height (in inches), whether they like to wear heels, and whether they like pink or red shoes (the most stereotypical colors).[51] We normalized each measure using z-scores and also construct a composite index of those four measures using the mean of their (z-scored) values.

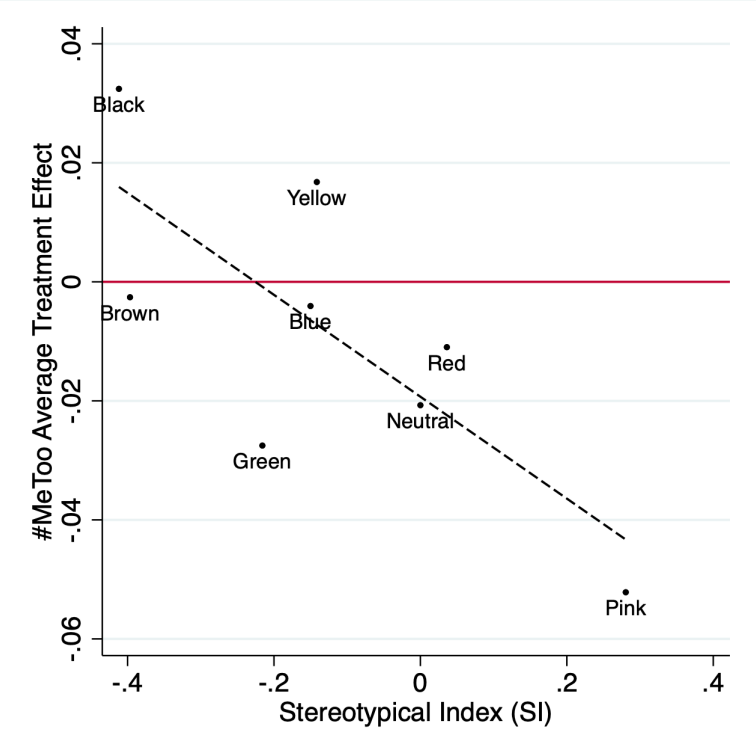
[51] Liking pink or red is equal to 1 (and 0 otherwise) if a respondent reports a 6 ("likely") or 7 ("very likely") when answering on the 1-7 Likert-scale to the question: "*How likely would you select a pair of shoes from the colors below?*". Similarly, respondents who reported a score of 6 or 7 on the heel preference scale are categorized as liking heels.

Randomization was effective as we do not find any significant difference across conditions in terms of respondent's location (region), age, or income. However, participants spent more time on the survey when allocated to the threat or empowerment conditions relative to the control condition (8% and 12% longer, respectively). We therefore control for the log of time spent doing the survey in all specifications.

Table S11 estimates average treatment effects regressing each preference measure on our treatment dummies using OLS. We find that participants reported liking heels significantly less in the #MeToo condition than in the baseline condition. The second dependent variable assessing the preference for higher heels was directionally consistent, although not significant. Similarly, participants reported significantly weaker purchase intentions for pink shoes in the #MeToo condition than in the baseline condition. In Figure S13, we report results for all colors and plot the estimated ATE against their Stereotypical Index (SI) estimated using the international survey of product ratings. Although only pink is significant at the 95% confidence interval (the estimated ATE on color red is negative, but not significant), there is a consistent relationship between a more negative ATE and shoe colors classified as more "feminine."

The two other treatment conditions (identity threat and empowerment) do not lead to any significant change in our three dependent variables. We also look at whether the treatment affected our sensitivity to threat and sense of power scales. While the empowerment treatment does not significantly affect those measures, the #MeToo treatment and threat treatment have a similar impact on both measures: they lead to higher sensitivity to threat and a lower sense of control. Given that only the #MeToo treatment has a significant effect on adherence to stereotypes, higher sensitivity to threat or lower sense of control are unlikely to drive the effect, as we would then expect it to replicate in the threat condition.

Figure S13: Average Treatment Effect vs. Stereotypical Index by Shoe Color



Notes: Figure shows estimated ATE from the #MeToo condition relative to the baseline condition on whether a respondent is "likely" or "very likely" to purchase a shoe from a particular color. Only pink is significant at the 95% confidence interval.

E.1 Study 1: Event Memory

In this first study, we are interested in the everyday language that people use to recount their experiences. Specifically, we would like you to remember a particular event and describe the experience as it happened to you, using language you would normally use. On the next screen, you will be asked to recall an event in your life. Please recall the event using words and language you would normally use.

MeToo condition. In 2017, #MeToo emerged as a major social movement on social media. The #MeToo movement was triggered by revelations about the sexual abuse of film producer Harvey Weinstein and quickly became a global phenomenon in which women shared their stories of sexual harassment. Please describe how these events made you feel, what you thought about them, to what extent you got involved in the movement, etc...

Threat condition. Please recall a particular episode in which you felt threatened as a woman. By threatened we mean a situation you went through where you were unsafe or in which you felt scared or stigmatised. Please describe this situation in which you felt threatened as a woman: what happened, how you felt, etc...

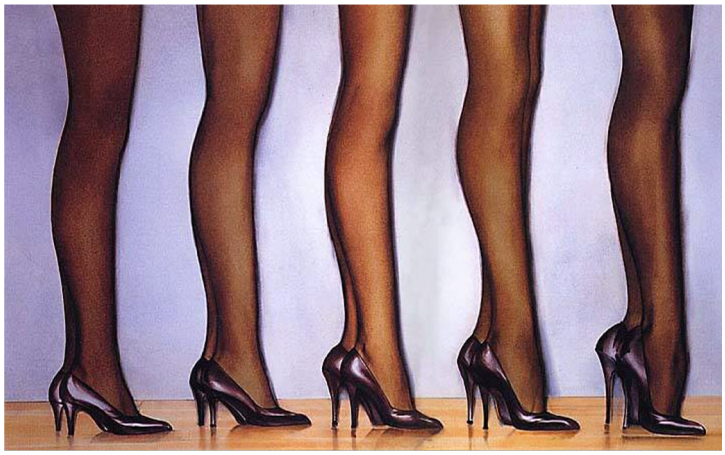
Empowerment condition. Please recall a particular episode in which you felt powerful as a woman. By powerful we mean a situation in which you felt you had the authority, the freedom, and the power to control the situation. Please describe this situation in which you felt powerful as a woman: what happened, how you felt, etc...

Baseline condition. Please remember a particular episode in which you went to the supermarket. Please describe this situation: what happened, how you felt, etc...

E.2 Study 2: Consumer Study

Welcome and thank you for participating in this consumer study. In this questionnaire, we would like to ask you about your preferences regarding different products. The product category on which we would like to collect your opinion is the category: "shoes". There are no right or wrong answers, we are interested in your opinions and experiences with these products.

Heel height choice. Imagine that you are browsing different pairs of shoes online. You are planning to buy a new pair of heeled shoes. Which pair of heeled shoes would you prefer from the options described below?



2 inches	3 inches	4 inches	5 inches	6 inches
Pair A	Pair B	Pair C	Pair D	Pair E

Shoe color preferences. Imagine that you are planning to buy a new pair of shoes with a specific color in mind. How likely would you select a pair of shoes from the colors below?

	Very Unlikely	Unlikely	Somewhat unlikely	Undecided	Somewhat likely	Likely	Very likely
Black shoes	○	○	○	○	○	○	○
Pink shoes	○	○	○	○	○	○	○
Red shoes	○	○	○	○	○	○	○
Neutral color shoes	○	○	○	○	○	○	○
Yellow shoes	○	○	○	○	○	○	○
Green shoes	○	○	○	○	○	○	○
Blue shoes	○	○	○	○	○	○	○
Brown shoes	○	○	○	○	○	○	○

Heel preference. How much do you like to wear heeled shoes?

Not at all						A lot
○	○	○	○	○	○	○

E.3 Study 3: Personality Study

Now, could you please answer the following statements using the proposed scale? Affirmations are about how you feel in general.

Sensitivity to threat scale. Please answer the following questions as precisely as possible using the following scale.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I tend to focus on the bad things that could happen to me.	○	○	○	○	○	○	○
I tend to be particularly sensitive to dangers and threats in my life.	○	○	○	○	○	○	○
I am the type of person who is very worried about a threatening situation.	○	○	○	○	○	○	○
Sometimes I feel like I am overwhelmed when I try to protect myself from all possible dangers in life.	○	○	○	○	○	○	○

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Sometimes I think I'm more at risk than I really am.	○	○	○	○	○	○	○
No matter what I do to feel more secure, I often worry about my safety.	○	○	○	○	○	○	○
Select "Somewhat agree"	○	○	○	○	○	○	○
I have the feeling that despite everything I do to avoid the dangers, it is not enough.	○	○	○	○	○	○	○

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
It's in my nature to feel like I'm more vulnerable to certain dangers, to try to overcome them, and not yet feel safe after taking certain precautions.	○	○	○	○	○	○	○
People used to describe me as an eternal worrier when it came to engaging in risky behavior.	○	○	○	○	○	○	○
When a threat to my personal safety arises, I am usually very concerned and worried beyond the level of the threat.	○	○	○	○	○	○	○

Sense of power scale. Please answer the following questions as precisely as possible using the following scale.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I can get people to listen to what I'm saying.	○	○	○	○	○	○	○
My wishes don't matter much.	○	○	○	○	○	○	○
I can get others to do whatever I want.	○	○	○	○	○	○	○
Even when I express my opinions, they have little influence.	○	○	○	○	○	○	○
I think I have a lot of power.	○	○	○	○	○	○	○
My ideas and opinions are often ignored.	○	○	○	○	○	○	○
Select "Agree"	○	○	○	○	○	○	○
Even when I try, I can't get what I want.	○	○	○	○	○	○	○
If I want to, I can decide for others.	○	○	○	○	○	○	○