Web Appendix for

Unmasking Social Compliance Behavior During the Pandemic

Table of Contents

Section S1. Data Details, Variable Measures, and Descriptive Statistics 2

S1.1. Data Construction 2

S1.2. Customer Profile Comparison 3

S1.3 Analyzing Customers in the Pre-Pandemic Period.. 3

S1.3.1. Main shopping variables in the pre-pandemic period... 3

S1.3.2. Did customers wear masks pre-pandemic?... 4

S1.4 Accuracies of Machine Learning Models for Video Analytics 5

Section S2 Contextual Details: The Store and Local Covid-19 Cases 8

Section S3. Detecting the Face Coverage Provided by the Mask 10

S3.1 Detecting Facial Landmarks and Mask-Wearing 10

S3.2 Extent of Mask Coverage 11

Section S4 Estimating In-Store Social Distancing and Crowdedness 12

S4.1 Social Distancing 12

S4.2 Store Crowdedness Measures 13

S4.2.1 Technical details on measuring store crowdedness 13

S4.2.2 Comparing store crowdedness between mask-wearing groups during the pandemic 13

Section S5 Robustness Tests.. 16

S5.1 Replicating the Main Results Using Alternative Customer Segmentation....................... 16

S5.2 Do Individual Customers Change Their Mask-Wearing Behavior Over Time? 19

References for the Web Appendix 23

Section S1. Data Details, Variable Measures, and Descriptive Statistics

The data in this study were collected from a well-known retail store chain that sells everyday fashion items at outlets in more than 50 big cities in Asia. We collaborated with the company to develop artificial intelligence (AI) technologies and implement AI-based retailing at one of the company's large stores in a major city. We chose the store for its high customer traffic (about 800 customers daily before the pandemic) and high sales volume (an average of 300 transactions daily before the pandemic). We focused on the customers who made a transaction.

To capture customer behavior, the store installed 16 high-resolution cameras inside the store. The cameras fulfilled two purposes: facial recognition (13 cameras) and wide-angle surveillance (3 cameras). The facial recognition cameras were installed outside the main entrance, at the alternative entrances, at the check-out counter, and in other locations that are not pertinent to this study. The cameras were adjusted to an optimal height and angle for recording human behaviors in detail.

S1.1. Data Construction

We collected video data from January 1 to May 31, 2020 (the store was closed from January 24–31 due to the onset of the Covid-19 pandemic). We obtained sales data for all 21,277 unique transactions that occurred during the same observation window; the data included every product name, category, price, and transaction time. Of the 21,277 unique transactions, 71% were generated by customers who had a customer member ID because they participated in the store's loyalty program.[1] We obtained the (de-identified) member IDs so that we could track repeat customers within our sample.

Next, we identify the 1,079 customers who had a member ID and who visited the store <u>both before and during the pandemic</u>. The 1,079 unique customers generated 3,479 unique transactions. (The overall store visit frequency was low compared to grocery stores.) This is our sample for our main analyses. Note that each customer's classification is based on the pandemic period exclusively, regardless of the customer's behavior in our pre-pandemic data.

[1] Joining the loyalty program was free and convenient—the customer simply scanned a QR code using their mobile phone, so most customers joined the program. Loyalty members do not receive additional promotions or discounts; rather, they accrue loyalty points from their transactions and may redeem points for a reward.

S1.2. Customer Profile Comparison

We performed a set of pairwise t-tests for the customer profiles of the two groups of mask wearers: average age and gender (proportion of female customers). In Table S1, the number in each cell indicates the pairwise difference between the two group means. As can be seen, the two groups of mask-wearers (i.e., *Fully-Compliant* and *Partially-Compliant* customers) had similar customer profiles, highlighting the difficulty of differentiating the two mask-wearing motives based on customer profiles alone.

Table S1 Comparison of the Customer Profiles of the Two Mask-Wearing Groups

Comparing ***Fully-Compliant*** Customers and ***Partially-Compliant*** Customers		
	Age	**Gender (proportion female)**
diff. = mean(*Fully-Compliant*) - mean(*Partially-Compliant*)	0.11	0.02
t value	0.27	0.99
Note. Statistics were computed based on the 582 unique *Fully-Compliant* customers and 318 unique *Partially-Compliant* customers. *p<0.05, **p<0.01, ***p<0.001.		

S1.3 Analyzing Customers in the Pre-Pandemic Period

In the main analyses, we used the extent of mask coverage to identify two segments of customers (*Partially-Compliant* and *Fully-Compliant*) who we argue had different primary motives for wearing masks during the pandemic period. Our main analyses (Sections 4.1-4.5) highlight that the two groups of mask-wearers responded to the pandemic in different ways in terms of shopping behaviors and transaction characteristics. One may wonder how the two groups compared before the pandemic.

S1.3.1. Main shopping variables in the pre-pandemic period

Table S2 presents the pairwise differences between the means of the dependent variables from Section 3.1 for *Fully-Compliant* and *Partially-Compliant* customers. Note that the dependent variables in the main analyses were logged, while Table S2 presents statistics for the unlogged measures.

As can be seen in Table S2, the two groups of mask-wearers had similar shopping behaviors and transaction characteristics in the pre-pandemic period (with two exceptions: *Item Discount* and *Shop Category Diversity*). The similarity between the two groups before the pandemic is consistent with our theorization that the subsequent differences in behaviors (in Tables 2–4) are attributable to different reactions to the pandemic.

Table S2 Statistical Test of Differences in Shopping Variables

Comparing *Fully-Compliant* Customers and *Partially-Compliant* Customers		
Variables	*diff.* = mean(Fully-Compliant) - mean(Partially-Compliant)	*t* value
Shopping Duration (seconds)	88	1.20
Shopping Duration per Item (seconds)	52.3	0.99
Purchase Quantity	-0.09	-0.36
Item Price	-5.33	-1.29
Item Discount	0.30*	2.51
Item Popularity	1.39	0.22
Shop Diversity (Category)	0.040*	2.07
Shop Diversity (Subcategory)	0.013	0.81
Shop Diversity (Color)	0.057	1.49
Social Distancing (distance to cashier)	2.66	0.31
Note. Statistics were computed based on the pre-pandemic observations involving *Fully-Compliant* customers and *Partially-Compliant* customers. *p<0.05, **p<0.01, ***p<0.001.		

S1.3.2. Did customers wear masks pre-pandemic?

We examine whether and how mask-wearing behavior changed in each customer group with the onset of the pandemic. Customers were classified based on their mask-wearing behavior *during* the pandemic, but if the classification captures a relatively stable characteristic of the customer's perceptions of and reactions to risks, then the groups should differ in pre-pandemic mask-wearing behavior as well.

We find that before the pandemic, 4.9% of the 1,079 customers wore a mask while shopping in the store; specifically, masks were worn by 7.6% of *Fully-Compliant* customers, 2.2% of *Partially-Compliant* customers, and 1.1% of *Unmasked* customers.

Note that both *Fully-Compliant* and *Partially-Compliant* customers wore a mask *during* the pandemic, but they exhibited different mask-wearing behaviors before the pandemic—*Fully-Compliant* customers were more than twice as likely as *Partially-Compliant* customers to wear a mask. It is unsurprising that mask-wearing was uncommon before the pandemic because the risk of getting sick was relatively low, though there was still some risk during flu season (which was active during our pre-pandemic data window).

S1.4 Accuracies of Machine Learning Models for Video Analytics

We employ a set of machine learning models and tools to analyze the video data. We use these models because they are trained on datasets to achieve high accuracy for performing the tasks that are needed to analyze our video data: for example, identifying human faces, predicting whether each customer was wearing a mask, computing the similarities between the faces, and measuring the customer profile (age, gender). In Table S3, for each model, we provide the task description, model accuracy on the benchmark dataset, and the specific application in our setting.

Table S3 A Summary of the Video Analytics Models

Model	Task	Accuracy
Multi-Task Cascaded Convolutional Network (MTCNN) deep learning model	Detecting and extracting human faces from a video clip	It achieved 95.04% accuracy on one face recognition benchmark dataset and has been shown to improve the accuracy of other benchmark methods by anywhere from 3.6% to 92.7%. [1] In our setting, we needed to extract only one face image of each customer from many frames of video. Provided that the cameras were installed in the optimal positions (facing the check-out queue and entries, less than 2 meters above the ground), we roughly estimate that the method can recognize and extract at least one well-qualified face image for almost 95% of the customers. After extracting the faces, we manually matched each face with a transaction by

		watching the video clip around the transaction timestamp. Hence, this process is independent of MTCNN and has almost 100% accuracy.
FaceMaskDetection model	Detecting mask-wearing	Both the precision and recall of this method for detecting mask-wearing are more than 90%. [2] We roughly estimated an accuracy of more than 95% in our setting.
Detect API (Face Model developed by Face++)	Measuring the customer profile (age, gender) from the detected face	Face++ is very accurate at recognizing gender and has achieved 100% accuracy on some datasets. Its age predictions are consistently reasonable (i.e., not far from the actual age). [3][5] We performed a quick manual check and found that the gender appeared to be 100% correct, and all ages seemed reasonable (though we do not know the real ages of the customers).
Face Similarity Generative Adversarial Networks (GAN)	Matching entry faces with payer faces	We used this method only to calculate the similarities between the payer's face and the faces on the store entry camera. Then, for each payer's face, we ranked the entry faces based on their similarities so that a human could efficiently manually identify which entry face was the payer. In other words, we used this method primarily to reduce the manual workload; the accuracy of matching depended on the human. We

		double-checked the manual labeling to ensure almost 100% accuracy. [4]

[1] Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Lett.* 23(10):1499–1503.
[2] https://github.com/AIZOOTech/FaceMaskDetection.
[3] https://www.faceplusplus.com/
[4] Wang J, Liu Y, Hu Y, Shi H, Mei T (2021) FaceX-Zoo: A PyTorch toolbox for face recognition. Preprint, submitted October 17, https://arxiv.org/abs/2101.04407.
[5] http://www.bernardjjansen.com/uploads/2/4/1/8/24188166/jansen_facial_icwsm2018.pdf

Section S2 Contextual Details: The Store and Local Covid-19 Cases

The store belongs to a national retail chain that owns about 500 stores (as of 2020), distributed across more than 50 big cities in Asia. All stores are located in urban downtown areas. Customers visit the store less frequently than a grocery store, often just once or twice per month (before the pandemic). Before the pandemic, the store in our study had a daily visitor count of about 800, which was relatively high for stores in this retail chain.

The store sells branded fashion products for daily use and offers six categories of products: (1) accessory and fashion, (2) underwear and loungewear, (3) cosmetics and skincare, (4) women's apparel, (5) men's apparel, and (6) shoes and bags. A product is categorized as "women's apparel" if it is women's clothing—outerwear, shirt, skirt, or dress. The "accessory and fashion" category includes hair accessories (e.g., hair pins, headbands, hair ties), scarfs, necklaces, earrings, sunglasses, designer mugs, home decorations, and stuffed animals.

During our pandemic sample period, an average of 7.6 customers entered the store every five minutes (computed from the camera at the store entry), and the average customer stayed in the store for 26.1 minutes, so the store was serving an average of 40 customers at any time. The area of the store is 500 square meters or 5381 square feet, so the average density of customers in the store was approximately 40/5381 = 0.0075 per square foot, and the average customer had 135 square feet to themselves. The repeat customer rate and shopping frequency were lower in our context than in many others, such as grocery shopping. The average customer with a member ID visited the store only 0.25 times per week.

The store installed 16 high-resolution cameras throughout the store. The cameras were of two models: the face recognition model is good for clearly recording human faces, while the overview coverage model captures behaviors over a large area. Of the face-recognition cameras, 1 was outside of the entry door, 3 faced the store entry, 1 faced the stairs to the second floor, 2 faced the check-out counter (to capture the customers' faces as they paid), and 6 faced the aisles in the women's clothing area. The 3 overview-coverage cameras were positioned to capture the entirety of the women's clothing area. The cameras were in the optimal position to record human behaviors.

The store appealed primarily to younger customers. Although the customers in our sample spanned ages 4–76 (as predicted by a facial recognition model), the average age in our sample was about 28. For reference, the city has an average age of 35.7, with 67% of the population in the 15–59 range.

We computed the rate of local Covid-19 cases from the public information provided by the website of the local Centers for Disease Control and Prevention (CDC). Dividing the total number of confirmed cases by the city's population, we determined that the local COVID-19 case rate was approximately 1/100,000 during our sample period. The city did not have an indoor mask mandate, and all stores and public places followed local CDC guidance to recommend (but not mandate) masks.

Section S3. Detecting the Face Coverage Provided by the Mask

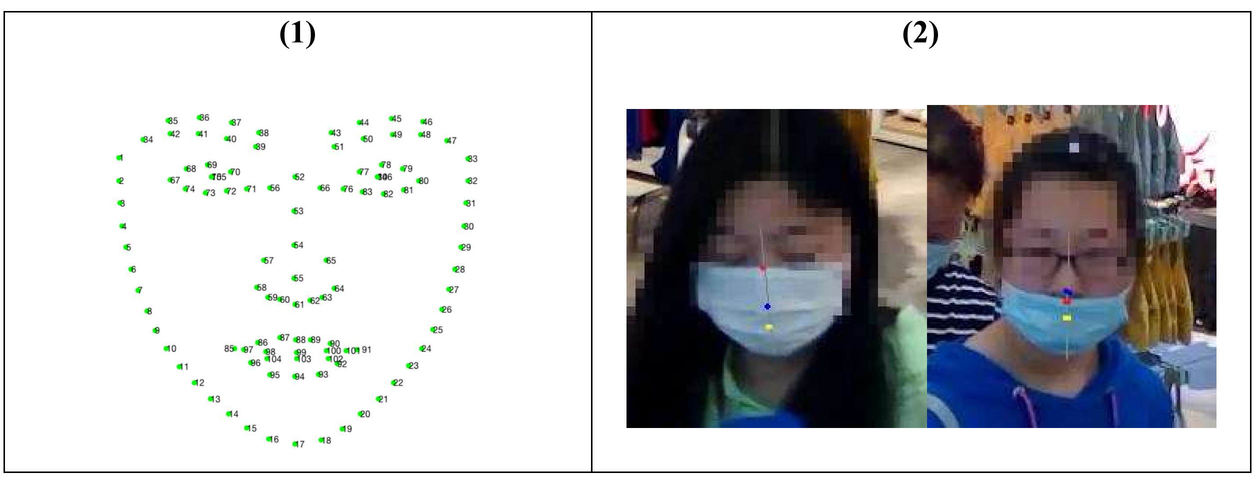
To detect whether each masked customer was wearing the mask as advised by experts, we developed a simple but effective method: detecting the mask in the discriminative color spaces along the midline of the face and calculating the height of the mask relative to the mouth and nose.

S3.1 Detecting Facial Landmarks and Mask-Wearing

For the image of each payer's face, we detected 106 landmarks (subfigure (1) in Figure S1) using the face recognition package (Wang et al. 2021). For the points along the midline of the face (points 52–55, 61, 88, 99, 103, 94, and 17), we regressed the x-values on the y-values to create the face's midline, and we checked the colors at every pixel along the midline. We considered the point to be covered by a mask if the color was blue, white, black, or red (representing 99% of masks). We adopted a combination of the HSV and YCbCr color spaces to differentiate mask colors from skin, lips, clothing, and other noise; RGB does not work well for this kind of application (Albiol et al. 2001). After detecting whether a mask was present at each point along the midline, we calculated the heights of the mask, nose (nasal tip, point 61 in subfigure (1) in Figure S1), and mouth (upper lip, point 88) relative to the chin (point 17).

Note that we devised our own procedure for detecting the midline colors rather than using existing image processing techniques to segment the whole mask area because our comprehensive experiments found the midline colors to be more robust (specifically, less disturbed by light, hair, and mask wrinkles). We found that this simple method yields surprisingly accurate predictions of mask coverage.

Figure S1 Detecting Facial Landmarks



Notes: (1) landmarks of the face; (2) two examples of detected mask coverage (gray line = midline, red point = upper mask edge, blue point = nasal tip, and yellow point = upper lip)

S3.2 Extent of Mask Coverage

For all mask-wearers, we examined the mask fit—specifically, the relationship between the nose, mouth, and the mask's upper edge (i.e., top of the mask)—to classify each mask-wearer as either *Fully-Compliant* or *Partially-Compliant*. Guidelines from the CDC state that the mask should be worn over the nose for better prevention of viral transmission via respiratory droplets.[2] Wearing a mask over the mouth only, leaving the nose exposed, is called "wearing a face mask halfway" and is dangerous because the nose is very vulnerable to the novel coronavirus.[3] At the onset of the COVID-19 pandemic, several studies as well as the WHO, CDC, and media emphasized the importance of covering the nose (Hou et al. 2020; Sungnak et al. 2020). Hence, we reason that customers who are more concerned about their own health risk should be more likely to wear the mask over the nose (and, thus, are more likely to be *Fully-Compliant* in our categorization).

In the above procedure, we determined the distance from the chin to four other points: *Nose_Height* (nasal tip), *Mouth_Height* (upper lip), *Mask_Height* (upper contour of the mask; see Figure S1), and *Eyebrow_Height* (eyebrow). Then, we calculate the *Face_Area* as *Eyebrow_Height - Mouth_Height,* and we calculate coverage of the nose, *Coverage_Nose*, as *Mask_Height - Nose_Height.* Finally, we calculate *Mask_Coverage*, the variable we use to segment the mask-wearers, as *Coverage_Nose/Face_Area.* A larger ratio indicates that more of the nose is covered by the mask, which reflects better protection.

[2] https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-to-wear-cloth-face-coverings.html

[3] https://www.discovermagazine.com/health/why-wearing-a-face-mask-halfway-can-be-dangerous

Section S4 Estimating In-Store Social Distancing and Crowdedness

S4.1 Social Distancing

We estimate social distancing from the videos at the check-out counter, where customers' faces were captured most reliably. (While customers were actively shopping, they intermittently turned away from the camera and were blocked from view.) We pre-process the video frames that contain the payers' faces. Specifically, we recognize and crop the faces in each frame using the MTCNN deep learning model (Zhang et al. 2016); for each face, the center point of the face rectangle is taken as the pixel location of the payer in the frame (see Figure S2). We calculate the distance between the focal payer and the cashier by taking the average, minimum, and maximum 2D Euclidean pixel distance in each frame between the center point of the payer's face and that of the cashier.

We use the distance to the cashier rather than to other customers because it should be easy to maintain distance from the cashier (who does not move around much). By contrast, the distance between a focal payer and another customer may be influenced by factors beyond the focal payer's control (e.g., how crowded the area is), so it may not accurately reflect the focal payer's desire to follow social-distancing guidelines.

Figure S2 Social Distancing

Notes: The face of the focal payer is highlighted with a red rectangle, and the faces of other customers are highlighted with blue rectangles. The 2D pixel distance from the focal payer to the cashier is indicated with a red dashed line.

S4.2 Store Crowdedness Measures

We developed measures of the crowdedness of several areas of the store, and we controlled for store crowdedness in our main analyses.

S4.2.1 Technical details on measuring store crowdedness

Crowdedness of the store entry. When a customer entered the store, their face was captured on one of the entry cameras. To determine the crowdedness in the entry area, first, the face in each frame of the entry videos was recognized and extracted. (To minimize noise and improve accuracy, we considered only the faces in the region of interest, or ROI: the rectangle around the entry itself, rather than the entire area captured on video.) Then, the extracted faces were clustered according to the similarities among them. (Before clustering, we dropped the noise faces, such as those with a small size or that were in profile.) Each cluster represented one customer, and we used the clusters to calculate the number of customers who entered the store in the 5 minutes prior to the focal customer's entrance.

Crowdedness of the check-out area and clothing areas. In the check-out area, women's clothing area, and underwear/men's clothing area, the position of the store cameras unfortunately prevented us from identifying specific customers; they often appeared in profile or with their back to the camera. Hence, the face-based method that we used for the store entry did not work. Instead, we adopted Facebook's object detection package, detectron2,[4] to recognize each human object in each frame. (Similar to our strategy for the store entry, we set the ROI to remove the noise human objects.) For the check-out area, we operationalized crowdedness as the number of customers who appeared on the check-out camera in the 5 minutes prior to the timestamp of the focal customer's transaction. For the clothing areas, we unfortunately could not identify whether or when the focal customer visited the area. Instead, we simply counted the number of customers who appeared on the cameras of the two clothing areas combined while the focal customer *was in the store* (i.e., between the entry timestamp and transaction timestamp).

S4.2.2 Comparing store crowdedness between mask-wearing groups during the pandemic

In the main analyses, we included the three crowdedness measures as control variables to deal with the possibility that changes in the store environment with the onset of the pandemic might affect

---

[4] https://github.com/facebookresearch/detectron2

customer behavior. Now, we compare *Fully-Compliant* customers and *Partially-Compliant* customers on the crowdedness measures during the pandemic.

We report the statistics in Table S4. Most notably, we found that the check-out area was 11% less crowded 5 minutes before the average *Fully-Compliant* customer checked out than 5 minutes before the average *Partially-Compliant* customer checked out (column 3; d = 0.24, t = 3.80). We reason that *Fully-Compliant* customers avoided checking out or lining up when the check-out area was crowded to reduce the risk of exposure to the virus. This is consistent with our finding in Section 4.1: *Fully-Compliant* customers practiced stricter social distancing than *Partially-Compliant* customers in the check-out area.

We did not find a significant difference between *Partially-Compliant* and *Fully-Compliant* customers in the crowdedness at the store entry (column 1; d = 0.011, t = 0.05), but this is unsurprising given that customers could not control the flow of traffic into the store. Finally, we also found no difference between *Partially-Compliant* and *Fully-Compliant* customers in the crowdedness of the clothing areas (column 2; d = 0.047, t = 1.01), but this measure has limited interpretability as it did not capture the choices of individual customers to visit or not visit the clothing areas (as described in Section S4.2.1).

Table S4 Store Crowdedness During the Customer's Visit During the Pandemic

Customer Group	**(1)** **Store Entry**	**(2)** **Clothing Areas**	**(3)** **Check-Out Area**
Fully-Compliant	6.10 (*S.E.* = 0.120)	1.98 (*S.E.* = 0.028)	1.91 (*S.E.* = 0.039)
Partially-Compliant	6.11(*S.E.* = 0.157)	2.02 (*S.E.* = 0.037)	2.15 (*S.E.* = 0.050)
Notes: The statistics were computed based on the *during-pandemic* crowdedness measures. For each customer *i*, we obtained the timestamp at which the customer appeared on the store entry camera and counted the number of customers who appeared on the store entry camera in the 5 minutes prior to *i*'s entry (column 1). Similarly, we operationalized the crowdedness at check-out as the number of customers who appeared on the check-out camera in the 5 minutes prior to the timestamp of *i*'s transaction (column 2). For the two clothing areas (women's and			

underwear/men’s), we simply counted the number of customers who appeared on the cameras of the two areas combined while *i was in the store* (column 3). The standard errors are in parentheses.

Section S5 Robustness Tests

S5.1 Replicating the Main Results Using Alternative Customer Segmentation

We repeat the main analyses with a less stringent definition of mask coverage—with an absolute (rather than relative) definition of fully-compliant mask coverage: whether the mask covered the nose (i.e., whether the height of the mask's upper edge exceeded the height of the nasal tip). This segmentation reflects a more restrictive view of *Partially-Compliant* customers: they were essentially *unmasked* from a medical point of view, because coverage of the nose is critical for preventing the transmission of the novel coronavirus (Hou et al. 2020; Sungnak et al. 2020). As described below, the results were unchanged.

The alternative segmentation yielded 122 *Partially-Compliant* customers, 778 *Fully-Compliant* customers, and 179 *Unmasked* customers. Note that fewer customers were labeled as *Partially-Compliant* here than in the main analysis (in which we split *Fully-* and *Partially-Compliant* customers around the average of *Mask_Coverage,* a continuous variable). Although this segmentation leads to a more conservative measure of poor mask-wearing, it still should reflect different underlying mask-wearing motives, and we expect to find different changes in shopping behavior with the onset of the pandemic between *Partially-Compliant* customers and *Fully-Compliant* customers.

First, the R^2 values reveal better model fit in the main analyses than in Tables S5–S8.

Second, all results are similar to the main results. In Table S5, *Fully-Compliant* customers shopped significantly faster during the pandemic than before the pandemic, while *Partially-Compliant* customers and *Unmasked* customers did not change their shopping speed with the onset of the pandemic. (Table S5 additionally used absolute shop duration as an alternative DV; as shown in column 1, the results are similar).

In Table S6, *Unmasked* and *Partially-Compliant* customers made purchases with similar transaction characteristics pre-pandemic and during the pandemic, with one exception: during (vs. before) the pandemic, *Fully-Compliant* customers bought significantly larger quantities of the same products and bought products with lower original prices, larger discounts, and higher popularity (in column 4, the coefficient of *Fully-Compliant X Pandemic* is positive, though it is significant at the 90% level).

In Table S7, *Partially-Compliant* customers did not change their diversity exploration behavior with the onset of the pandemic (measured three ways: number of categories, number of subcategories, and number of product colors). Meanwhile, *Fully-Compliant* customers explored significantly less diversity with the onset of the pandemic.

Finally, in Table S8, the shopping frequency decreased significantly with the onset of the pandemic for all customer groups, but more so for *Fully-Compliant* customers than for *Partially-Compliant* customers.

Table S5 Changes in the Shopping Duration by Customer Group: Robustness Test

	Estimates (Std. Err.)	
Variables	**(1) Shopping Duration**	**(2) Normalized Shopping Duration**
Pandemic (reference group: *Unmasked*)	-0.126$^{(\cdot)}$ (0.0671)	-0.121 (0.0892)
Fully-Compliant X Pandemic	-0.172* (0.0721)	-0.262** (0.0952)
Partially-Compliant X Pandemic	0.149 (0.0963)	0.0755 (0.128)
Crowdedness_Store_Entry	-0.00706 (0.00370)	-0.0104* (0.00430)
Crowdedness_Store_Checkout	0.0927*** (0.0127)	0.0363** (0.0139)
Crowdedness_Clothing_Area	0.00313 (0.0201)	0.0201 (0.0226)
Fixed Effect	Customer	Customer
Time-related Fixed Effects	Day of week, Shopping hour	Day of week, Shopping hour
Observations	3479	3479
R^2	0.492	0.448

Note. DVs are logged. In column 1, the DV is the shopping duration: the time, in seconds, between the entry and transaction timestamps. In column 2, the DV is the normalized shopping duration: the shopping duration divided by the number of purchased items. *Unmasked* was the default group. Standard errors are in parentheses. $^{(\cdot)}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

Table S6 Changes in Transaction Characteristics by Customer Group: Robustness Test

	Estimates (Std. Err.)			
Variables	**(1)**	**(2) Item Price**	**(3) Item Discount**	(4)

	Purchase Quantity			Item Popularity (total daily sales)
Pandemic (reference group:	0.0476	0.0484	-0.0120	0.055$^{(\cdot)}$
Unmasked)	(0.0519)	(0.0743)	(0.0497)	(0.031)
Fully-Compliant X Pandemic	0.134*	-0.203*	0.252***	0.258$^{(\cdot)}$
	(0.0578)	(0.0880)	(0.0537)	(0.134)
Partially-Compliant X Pandemic	-0.0239	-0.0329	-0.0622	0.321
	(0.0718)	(0.107)	(0.0721)	(0.198)
Crowdedness_Store_Entry	0.00235	-0.00827	-0.000771	0.00724
	0.00235	-0.00827	-0.000771	0.00724
Crowdedness_Store_Checkout	0.0425***	0.0375*	0.00608	0.00899
	(0.00847)	(0.0174)	(0.00905)	(0.0226)
Crowdedness_Clothing_Area	-0.0184	0.0483	0.00796	-0.0313
	(0.0133)	(0.0271)	(0.0139)	(0.0364)
Fixed Effect	Customer	Customer	Customer	Customer
Time-related Fixed Effects	Day of week, Shopping hour	Day of week, Shopping hour	Day of week, Shopping hour	Day of week, Shopping hour
Observations	3479	3479	3479	3479
R^2	0.432	0.498	0.476	0.403

Note. All DVs are logged except for in column 3, where values are reported as percentages. *Unmasked* was the default group. Standard errors are in parentheses. $^{(\cdot)}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

Table S7 Changes in Diversity Exploration by Customer Group: Robustness Test

	Estimates (Std. Err.)		
Variables	**(1) Item Diversity (Category)**	**(2) Item Diversity (Sub-Category)**	**(3) Item Diversity (Color)**
Pandemic (reference group:	0.0160	-0.0247	0.0297
Unmasked)	(0.0182)	(0.0159)	(0.0238)
Fully-Compliant X Pandemic	-0.0486*	-0.0430*	-0.0507*
	(0.0198)	(0.0173)	(0.0255)
Partially-Compliant X Pandemic	-0.00813	0.0316	0.0330
	(0.0259)	(0.0218)	(0.0384)
Crowdedness_Store_Entry	-0.000960	0.0000223	0.000228
	(0.000934)	(0.000820)	(0.00102)
Crowdedness_Store_Checkout	-0.00968**	-0.00628*	-0.00566
	(0.00319)	(0.00262)	(0.00346)
Crowdedness_Clothing_Area	0.0116*	0.00567	0.0117*
	(0.00494)	(0.00419)	(0.00483)
Fixed Effect	Customer	Customer	Customer
Time-related Fixed Effects	Day of week, Shopping hour	Day of week, Shopping hour	Day of week, Shopping hour
Observations	3479	3479	3479

R^2	0.396	0.415	0.447

Note. DVs are the log of the number of categories/sub-categories/colors divided by the number of items in the transaction. *Unmasked* was the default group.
Standard errors in parentheses. $^{(.)}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

Table S8 Changes in Shopping Frequency by Customer Group: Robustness Test

	(1) Shopping Frequency	
Variables	**Estimates**	**(Std. Err.)**
Pandemic (reference group: *Unmasked*)	-0.242***	(0.0245)
Fully-Compliant X Pandemic	-0.398***	(0.0298)
Partially-Compliant X Pandemic	0.068$^{(.)}$	(0.0350)
Crowdedness_Store_Entry	-0.00182	(0.00117)
Crowdedness_Store_Checkout	0.000144	(0.00420)
Crowdedness_Clothing_Area	0.00801	(0.00530)
Time-related Fixed Effects	Day of week, Shopping hour	
Observations	2158	
R^2	0.861	

Note. The DV is the average shopping frequency, computed as the weekly number of store visits in each period (3 pre-pandemic weeks; 18 during-pandemic weeks). *Unmasked* was the default group. Standard errors are in parentheses. $^{(.)}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

S5.2 Do Individual Customers Change Their Mask-Wearing Behavior Over Time?

A customer's mask-wearing behavior and shopping duration may depend on more than their relatively stable primary motive for mask-wearing—so mask-wearing behavior theoretically might vary within-customers across store visits. For instance, even *Unmasked* customers might wear a mask under a mask mandate policy; even *Fully-Compliant* customers might not bother wearing a mask when there is little risk of infection (such as before the pandemic).

To examine within-customer variation, we regressed the mask-wearing behavior on the customer fixed effect, week fixed effect, store crowdedness, and shopping hour (e.g., 9am or 3pm):

$$Wear\_Mask_{it} = \beta_t \cdot Week_t + \alpha_i + X_i + \varepsilon$$

where $Wear\_Mask_{it}$ is 1 (0) if customer i wore a mask during his/her store visit in week *t.* (If i did not visit the store in week *t*, then there is no observation *it*). $Week_t$ is a series of week dummies, and α_i is the customer fixed effect. The key coefficient β_t captures whether the same customer changed his/her mask-wearing choice across time and across store visits. X_i is a set of variables that may affect the mask-wearing choice: store crowdedness (customers might be more likely to wear a mask if the store is crowded) and the shopping hour (which may affect the customer's expectations about crowdedness), segmented on an hourly basis (from 9am to 10pm, using the transaction timestamp) and indexed as a 13-dimensional categorical variable.

The identification comes from the subsample of repeat customers: those who were identified by their member ID in the store's loyalty program and visited the store at least twice during the pandemic period of data collection. This subsample contained 5,351 store visits made by 2,242 unique customers. Note that, unlike the main analyses, the present analysis included customers who did not visit the store in the *pre-pandemic* period.

We report the estimation results in Table S9. Note that 3,715 store visit observations (by 1,891 unique customers) were dropped from the estimation because the outcomes were all positive (Wear_Mask = 1) or all negative (Wear_Mask = 0). Thus, 84% (= 1,891 / 2,242) of the repeat customers displayed the <u>same</u> mask-wearing behavior during every store visit during the pandemic. Other customers displayed inconsistent mask-wearing behavior, but there was little within-customer variation. All of the week dummy coefficients were insignificant, suggesting that for any given customer, the mask-wearing behavior was relatively stable across time. Also, the coefficients of the shopping hour dummies and store crowdedness variables were insignificant. We verified that there was sufficient variation in the data: the means of the crowdedness measures were approximately three times the standard deviations, and the shopping hour distribution was fairly even except for 9am and 10pm. The insignificant coefficients suggest that the mask-wearing behavior was relatively stable across variations in our context.

Nevertheless, we acknowledge that customers might adjust their mask-wearing behavior and other shopping behaviors based on the context. In this sense, under certain conditions, a customer might belong to more than one customer group. Unfortunately, our dataset lacked the granularity required to create a probability model of mask-wearing motives. Instead, we treat the customer groups as exclusively segmented (and the small within-customer variation in mask-wearing supports this decision), but we acknowledge that some overlap is possible, and we caution readers to keep this caveat in mind.

To conclude, the results suggest that mask-wearing behavior was mostly stable over time. The analysis is not a full proof for the whole customer population, as data limitations enabled us to analyze only a subsample (repeat customers with member IDs), but it helps us understand the robustness of mask-wearing decisions over time.

Table S9 Estimating Within-Customer Variation in Mask-Wearing

Wearing a Mask During the Pandemic	**ESTIMATES**	**S.E.**
$Week_w$ = 5 (reference)	-	-
$Week_w$ = 8	4.604	(1909.9)
$Week_w$ = 9	-9.377	(570.9)
$Week_w$ = 10	-8.852	(570.9)
$Week_w$ = 11	-9.558	(570.9)
$Week_w$ = 12	-9.015	(570.9)
$Week_w$ = 13	-10.37	(570.9)
$Week_w$ = 14	-9.965	(570.9)
$Week_w$ = 15	-10.57	(570.9)
$Week_w$ = 16	-10.34	(570.9)
$Week_w$ = 17	-10.77	(570.9)
$Week_w$ = 18	-11.85	(570.9)
$Week_w$ = 19	-11.35	(570.9)
$Week_w$ = 20	-12.11	(570.9)
$Week_w$ = 21	-12.83	(570.9)
$Week_w$ = 22	-12.50	(570.9)
Crowdedness_Store_Entry	0.0139	(0.0124)
Crowdedness_Clothing_Area	-0.0113	(0.127)
Crowdedness_Store_Checkout	0.0757	(0.0732)
Shopping Hour = 9am (reference)	-	-
Shopping Hour = 10am	-0.0279	(0.705)
Shopping Hour = 11am	0.700	(0.687)

Shopping Hour = 12pm	0.511	(0.650)
Shopping Hour = 1pm	0.620	(0.649)
Shopping Hour = 2pm	0.233	(0.639)
Shopping Hour = 3pm	-0.0759	(0.651)
Shopping Hour = 4pm	0.394	(0.651)
Shopping Hour = 5pm	0.466	(0.645)
Shopping Hour = 6pm	0.157	(0.643)
Shopping Hour = 7pm	0.450	(0.653)
Shopping Hour = 8pm	-0.181	(0.644)
Shopping Hour = 9pm	0.218	(0.648)
Fixed Effect	Customer ID	
Log likelihood	-475.95	
Observations	1636	

Notes: We evaluate whether mask-wearing changed over time within repeat customers. We removed the pre-pandemic data (1/1–1/23), so the week coefficients start in week 5. Then, there were almost no observations in February (the 6th and 7th weeks; only 7 average daily transactions), likely because of a strict quarantine policy. We do not include the estimated coefficient of *Shopping Hour = 10pm* because there were only 12 observations during the 10pm segment, and all were dropped from the estimation.

References for the Web Appendix

Albiol A, Torres L, Delp E (2001) Optimum color spaces for skin detection. *Proc. 2001 Internat. Conf. Image Processing* (Institute of Electrical and Electronics Engineers, Washington, DC), 122–124.

Hou YJ, Okuda K, Edwards CE, Martinez DR, Asakura T, Dinnon KH, Kato T, et al. (2020) SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* 182(2):429–446.

Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, Talavera-López C, et al. (2020) SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Natl. Medicine* (26):681–687.

Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Lett.* 23(10):1499–1503.