The Effect of Increasing Model Size on Intent to Purchase Clothing

Spring 2021 - W241 Final Report
By Haley Farber, Brandon Scolieri, Candice Sener, Anjali Unnithan, Derrick Xiong

I. Abstract

The modeling industry has experienced tremendous changes within the past decade regarding more diverse representations of models based on gender, race, and body type. Our goal of this project was to determine if the inclusion and increased visibility of plus-size models would cause women to be more likely to purchase clothing. Essentially, does the exposure of plus-size models in certain clothing garments increase or decrease purchase intent of those clothing garments on average?

To help solve this question, we created a survey in which each participant was shown twenty images of models wearing particular garments and asked to rate their likeliness of purchasing the advertised item. We administered two versions of a survey on Mechanical Turk, one with plus-size models and one with model sizes most represented in the modeling industry (i.e. US womens' sizes 0-4). We were interested in determining whether or not someone would give a different purchase intent score when viewing a model that resembled the average size of a female U.S. shopper versus viewing a model that resembled the average size of a female U.S. model, which is a U.S. women's size 0-4.

As part of our experimentation we also investigated whether different levels of self-esteem have an impact on purchase intent. As is elaborated on in our results section, we found no significant difference in average purchase intent between the treatment and control groups overall. However, we did discover some differences between the control and treatment group's responses as it relates to self-esteem, but no statistically significant results or differences across various body sizes.

II. Background

The body positivity movement, officially started in 2012 to encourage individuals to love and appreciate all body types openly, has become a household name and made significant strides on a large scale, with companies like Adidas and American Eagle expanding their clothing selections to include plus-size garments. Even with the increasing awareness of the concept of "loving one's body no matter the size", we still see most online clothing models wearing much

smaller sizes than the average American woman size, which is roughly 16-18[1]. Will consumers be more inclined to express the intention of a purchase when they see the clothing fit more closely to their true body shape? In this project, we decide to conduct an experiment to find out the impact of plus-size models on the intent of purchasing clothing.

We do realize that intent to purchase doesn't necessarily guarantee that a customer would go out of their way to buy the advertised product, but gauging the intent can provide insight for creating better apparel marketing practices which can have significant financial implications for the businesses that utilize this information. In addition, we would be remiss not to mention that this has the potential to propel further conversation about the importance of inclusion and diversity in advertising and the modeling industry at large.

III. Research

A study[2] conducted by Iowa State University the impact of self-esteem and consumption values on consumers' attitude and intention to purchase luxury fashion products. The data were collected through online surveys and college students were invited to participate. A convenience sampling strategy was applied. The survey link with an invitation was sent to college students; and extra credits were provided to those who completed the survey. In total, 214 usable responses were generated. Multi-item scales were developed to measure self-esteem, social value, emotional value, quality value, price value, attitude and intention to purchase luxury fashion brands. The results suggested that social value and quality value motivate consumers to purchase luxury fashion brands directly and indirectly. According to literature in the field, body satisfaction has been found to play a large role in self-esteem and in many disorders that affect the health and well being of adolescent and young adult women[3]. The research shows that a person's attitudes towards an ideal image can be skewed by their personal body image, self-esteem, or self worth. Results from this study prompted us to consider self-esteem and their own identified body size as integral parts of determining consumer's purchasing choice and we decided to use both self-esteem and self-identified body size as causal pathways leading to intent to purchase.

Another study we found particularly interesting and served as a prime motivation behind this project, especially its findings regarding self-esteem levels, was one conducted at Kent University[4]. The experiment explored the effect of replacing size 0 with average sized models in

---

[1] Deborah A. Christel & Susan C. Dunn (2017) Average American women's clothing size: comparing National Health and Nutritional Examination Surveys (1988–2010) to ASTM International Misses & Women's Plus Size clothing, International Journal of Fashion Design, Technology and Education, 10:2, 129-136, DOI: 10.1080/17543266.2016.1214291

[2] Perrier, Angela Michelle, "Influence of model body size on consumer attitudes and purchase intention" (2008). Graduate Theses and Dissertations. 11393. https://lib.dr.iastate.edu/etd/11393

[3] Cash, T. F., & Pruzinsky, T. (Eds.). (1990). *Body images: Development, deviance, and change.* Guilford Press.

[4] Bian, X. and Wang, K.-Y. (2015), "Are size-zero female models always more effective than average-sized ones? Depends on brand and self-esteem!", European Journal of Marketing , Vol. 49 No. 7/8, pp. 1184-1206. https://doi.org/10.1108/EJM-08-2013-0414

advertising on consumer response on model attractiveness, and observed any differences amongst newer and older, more established clothing brands. The main hypotheses out of that study were that for newer brands, people, specifically females ranging from ages 18-25, are more likely to view average size models as more attractive, whereas for older brands, people are more likely to rate size 0 models higher due to smaller models historically being associated with high end fashion brands. In addition, the study cites work from previous work that claimed [5] "physical attractiveness of a model in an advertisement increases consumers' positive attitudes towards the product, their willingness to purchase, and actual purchase", implying that a consumer's perception of a model's appearance plays a role in intent to purchase.

The first hypothesis was supported by the results, where for newer brands, average model attractiveness ratings were significantly higher than size 0 model ratings, supported by a p value less than 0.01. The second hypothesis was not, as there was reported to be no significant difference between the ratings of both model types for older brands. Additionally, the Kent University study comes to the conclusion that self esteem has an impact on how attractive women views an average-sized model and a size-0 model; those with lower self esteem rated average-size models as more attractive for newer clothing brands, while those with high self esteem rated both model types as equally attractive. Through their regression analysis for a three way interaction between clothing brand, self-esteem, and model size showed statistically significant results at the 10% cutoff, while the two-way interaction between self-esteem and model body size when considering only newer brands showed statistically significant results at the 1% level, where low self-esteem survey participants rated size 0 models as significantly less attractive. For older, established brands, there was no significant two-way interaction, nor were there heterogeneous treatment effects for individuals in high or low self-esteem groups. While the Kent University study's focus was on the newness of clothing brands and the effects of consumer response to model attractiveness, we were inspired to investigate the link between how self-esteem level may impact the purchase intent for the outfit that the model advertises.

Based on these conclusions, we cannot conclude the psychological reasons behind why self-esteem would influence their particular choices, but for our project, we hope to take it a step further to determine purchase intent. Our analysis focuses on the difference in intent to purchase between treatment and control and we use self-esteem as a covariate to determine if we see differences in purchase intent amongst respondents of various self-esteem levels. A potential conditional hypothesis that reflects our study is that for individuals with high self-esteem, there is no overall difference in average purchase intent after viewing the treatment versus after viewing the control models. However, for individuals of lower self-esteem, we should expect a difference in average purchase intent between treatment and control groups.

---

[5] Ohanion, R. (1990), "Construction and validation of a scale to measure celebrity endorser's perceived expertise, trustworthiness, and attractiveness", *Journal of Advertising,* Vol. 19 No. 3, pp 39-52

IV. Hypothesis

The hypothesis we test is: “Is there a difference in average purchase intent when participants view plus size models instead of size 0-4 models?” Our null hypothesis is that there is no difference in average purchase intent after viewing size 0-4 models (control group) and after viewing plus size models (treatment group).

The Kent University paper includes another study that evaluates whether high and low self-esteem participants differ by body size. Based on those results, “the two way interaction of self-esteem and model body size was not significant”, supported with a p-value of 0.85. Based on this conclusion, we hypothesize that self-esteem and self-reported participant body size are independent. This prompted our motivation to consider participant body size as a covariate and to observe potential HTEs within various body size groups.

To summarize, we will consider the impact of self-esteem and self-identified body size, and we predict that there are heterogeneous treatment effects within various self-esteem levels and separately within various participant body sizes.

V. Experimental Design

V.i. Experimental Overview

Our experiment contains two different versions of a survey each with various demographic and shopping behavior questions and 20 images of a particular size model, either size 0-4, or plus-size. We collected two groups of images: control and treatment. The images in the control group are pictures of clothing worn by size 0-4 models and images in the treatment group are pictures of the same clothing worn by plus-size models. We randomly assigned subjects to the treatment or control group and asked them how likely they were to purchase a certain item of clothing shown in each picture on a Likert Scale with values 1-5, where 1 is “Extremely Unlikely” and 5 is “Extremely Likely”. The randomization process is explained in more detail in the Randomization section (Section VI). We then measured the difference in the average purchase intent for each set of pictures amongst those in the treatment group and those in the control group. Below are the project timeline and measurement and communications tooling.

V.ii. Project Timeline

The project timeline can be found below:

Survey design started	Initial survey created	Pilot data collection	Final Survey sent via Mechanical Turk	Analyzing data and results	Drawing conclusions and finalizing reports
Feb 28	March 9	March 10 - March 15	March 19 - March 26	March 24 - April 10	April 10 - April 12

V.iii. Measurement and Communications Tooling

We used Qualtrics to design our survey. Qualtrics is a platform that has built-in randomization and blocking features. It is a powerful online survey tool that allows one to build and distribute surveys and analyze responses. We designed the survey to be specifically restricted to women since we are only interested in analyzing womens' purchase intent by viewing various female model sizes. After the design of the survey, we deployed the survey on Slack to gather pilot data. With the pilot data, we tweaked the survey to include the suggestions provided by the information such as time per question and better demographic information. We then set up our Mechanical Turk account to deploy the finalized survey. We gathered around 100 responses from Mechanical Turk with subjects coming from a variety of backgrounds. After the initial batch of 101 subjects, we started the analysis process, but realized we needed more statistical power, and as a result added more batches and settled on a maximum sample size of 144 respondents due to resource constraints. With the additional responses, we discovered that our variance increased. As such, we performed a followup power test to estimate how many subjects we would need to reach 80% power and discovered that we would need approximately 637 participants in each group. These additional respondents could add more statistical power to our final results, although this number could change depending on how much variance new respondents inject into the sample. Lastly, it should be noted that we used the worker ID provided by Mechanical Turk and the random ID generated by the qualtrics survey to cross reference participants and ensure that there were no repeat respondents across Mechanical Turk batches. This resulted in the removal of seven responses from participants who repeated the survey twice.

VI. Randomization

We decided on a blocking approach on the participant body size to ensure that there was a relatively even split of body types among the treatment and control surveys. Two of our survey questions corresponded to clothing size on US womens' scale, where for one question, a

respondent was asked to indicate a size bucket between the choices 0-4, 6-10, 12-16, 18-24, and 26+. We then programmed the Qualtrics survey to ensure an approximately even representation of treatment and control options for each size bucket (i.e. 50% of 0-4 respondents get the treatment version of the survey, and the other 50% get the control version of the survey).

Apart from the exposure to treatment and control, we wanted identical surveys to be administered, which essentially came down to ensuring the image sizes were of a constant height and width.

VII. Data Completeness

Initially 144 participants who self-identified as female signed up to the survey, and they were either assigned to a control or treatment group. All of the 144 participants completed the survey. Diagram IX.a indicates the breakdown of the number of participants in control and treatment groups, and also it shows the number of participants who completed the survey. Diagram IX.b shows the number of participants in treatment and control groups for each size bucket. As we wanted, there is an even split of respondents between the treatment and control groups overall, and 100% survey completeness. Among individual size buckets, size 6-10 is the most represented, with a total of 63 responses. Within the buckets, there is a relatively even split amongst treatment and control responses.

Diagram IX.a

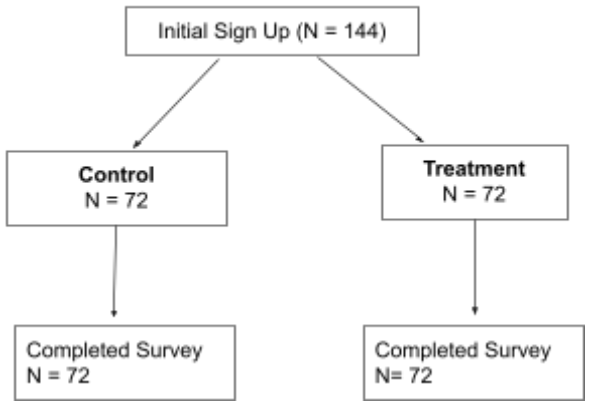
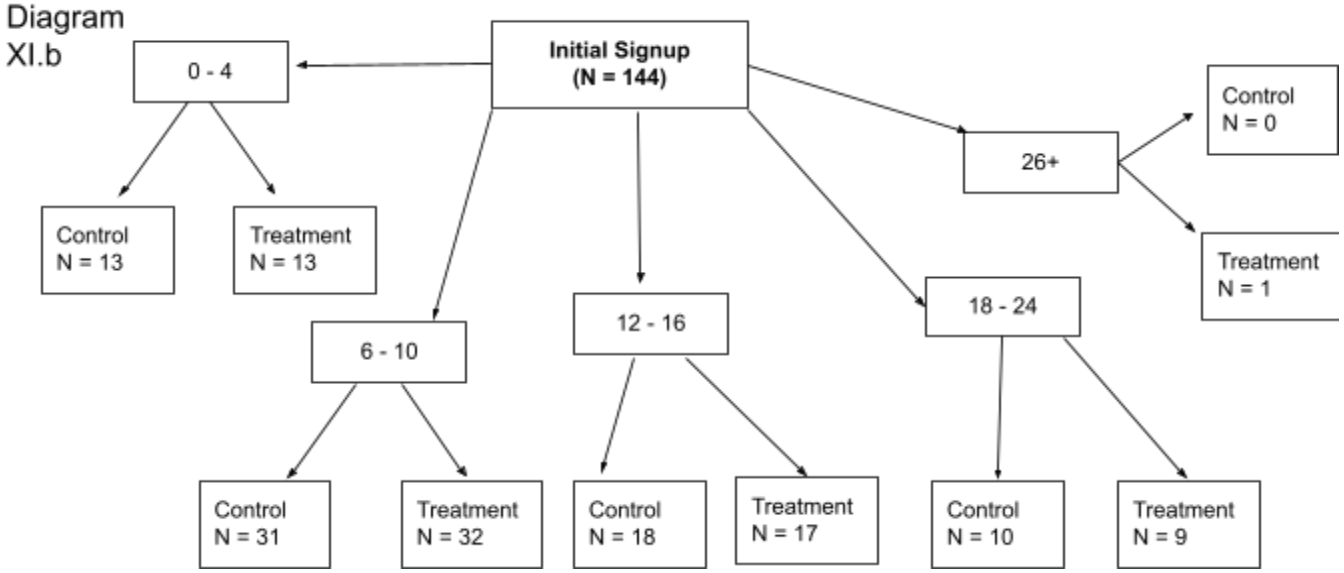


Diagram
XI.b



VIII. Treatment and Covariates

Below is a snapshot of what a participant would have seen when they are told they would view twenty images and evaluate purchase intent. The left image corresponds to the control survey, while the right corresponds to the treatment survey.

How likely are you to purchase this item?

Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
○	○	○	○	○

HOw likely are you to purchase this item?

Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
○	○	○	○	○

The remaining nineteen images and purchase intent follow-up questions maintain the same format, with an image appearing followed by “How likely are you to purchase this item?”. The

clothing order for both survey versions is the same. The only differences between the two surveys are the models' body sizes.

To see more questions included in the survey, please refer to the "Survey Questions" section under the Appendix.

Within the regression models, the main covariates we included are the survey participant's self-esteem rating and the clothing size. Due to lower representation of participants for certain responses to survey questions (i.e. one or two), we chose to pool some of the responses into a larger category. For example, the self-esteem question (i.e. "How would you rate your self-esteem?") originally had five choices for the participant to choose from: "Very low", "Somewhat low", "Neutral", "Somewhat high", "Very high". We condensed these into three levels, "High", "Medium", and "Low". For size-buckets, there was only one respondent who identified as 26+ so we condensed the size buckets into "0-4","6-10", "12-16", and "18+". We ran a regression using these size buckets. Based on the summary statistics, the most common response was size 6-10, which we called the "modal" bucket, and we renamed the three levels to "Below modal" (0-4), "At modal" (6-10), and "Above modal" (12+). Another regression model involved condensing into binary buckets, "Below or at modal" (0-10) and "Above modal" (12+).

VIII.i. Data Cleaning

We had to do several cleaning steps before the creation of our models. The first processing we did after gathering the data was translating all the Likert scale responses to numerical variables. For instance, "Extremely Unlikely" was converted to 1, and "Extremely Likely" was converted to 5. As mentioned above, we did a similar condensing process for some of the additional survey questions as we realized that some of the responses had lower representation of participants. To view the summary statistics and the numerical breakdown of the covariates and other variables, please refer to the "Summary Statistics" section in the Appendix.

Our outcome variable is average purchase intent. For each participant, we calculated the average purchase intent by averaging the Likert scale responses to all twenty model images that they received depending on treatment assignment.

IX. Observations and Outcome Measures

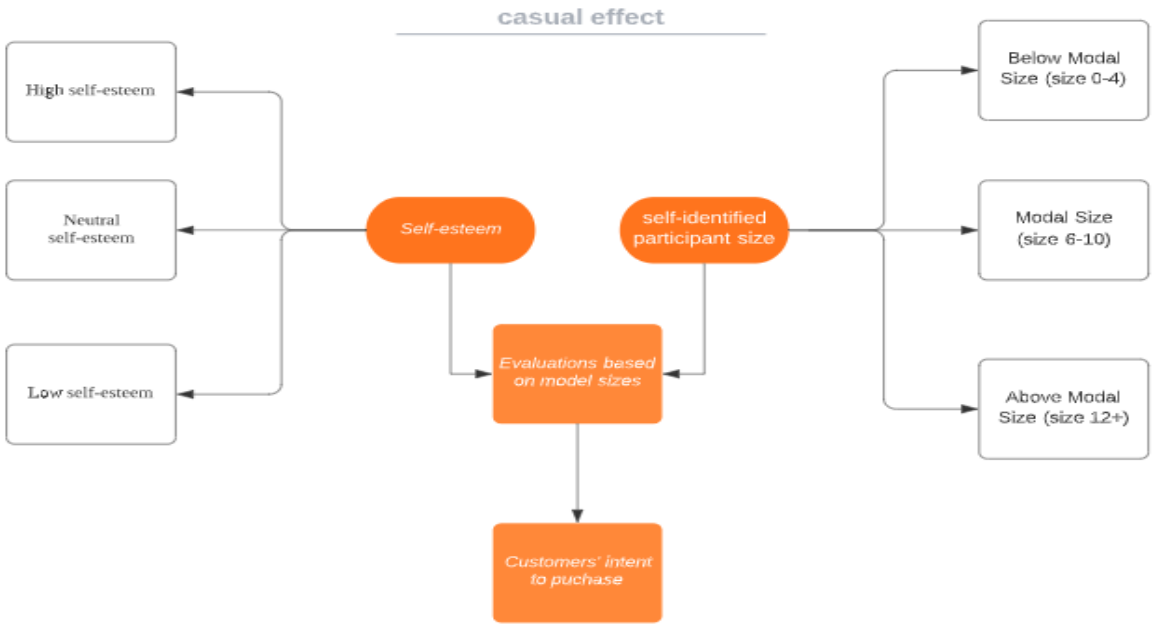
The primary outcome measure is average purchase intent. We briefly looked into secondary outcome measures such as the number of times that a survey respondent clicked on choices. Another measure of interest was the difference in time spent answering a purchase intent question for individual treatment or control situations. In the case of our click count outcome, we were attempting to measure an increase or decrease in hesitation for a participant's selection via using click count as a proxy for hesitation. With regard to our question-duration metric, we were curious to see whether or not some images caused participants in different groups to take

more or less time to answer the survey questions on average. Our results were highly insignificant, and therefore we decided not to pursue further analysis into these secondary outcomes.

As mentioned above in the abstract, we hope to look into self-esteem and participant body size in order to measure any potential heterogeneous treatment effects.

X. Heterogeneous Treatment Effect Pathway

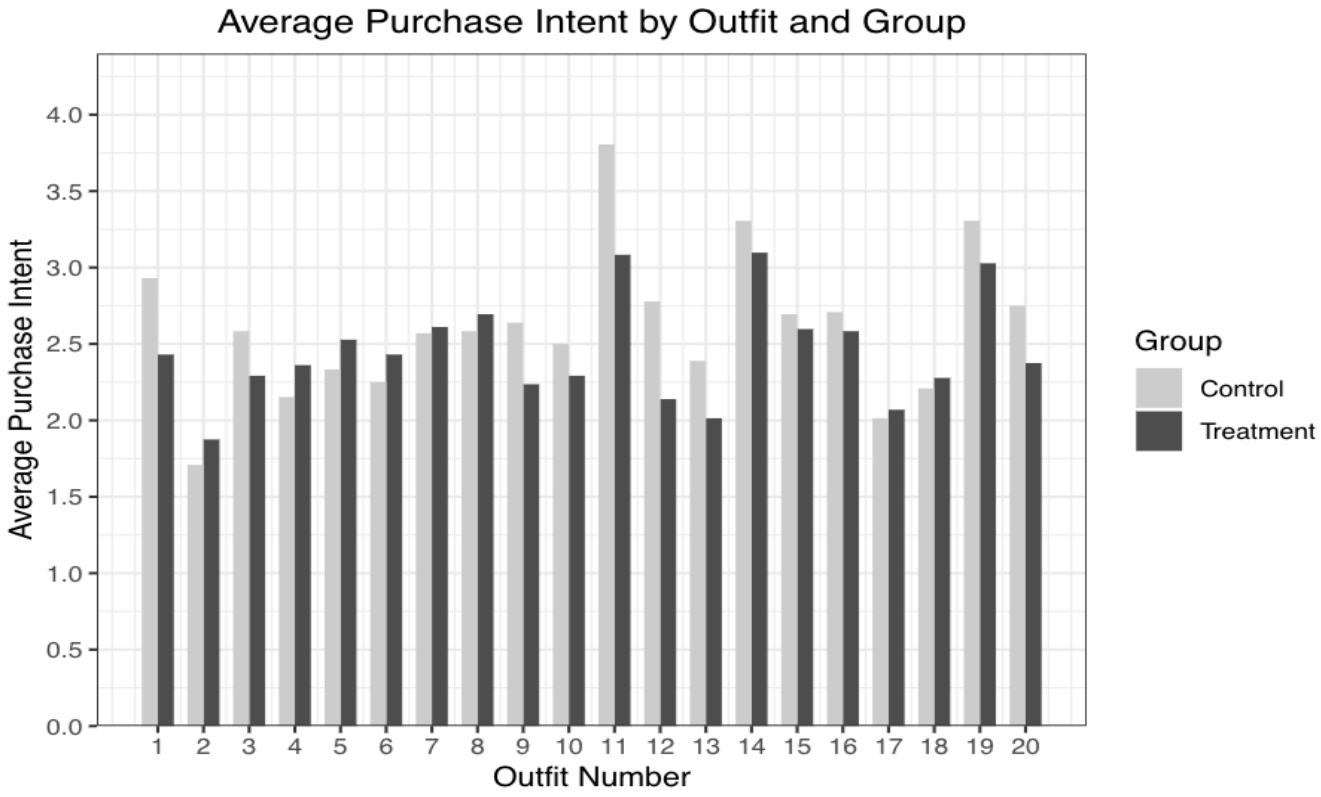
Below is a pathway model to visualize the possible covariates that would result in potential heterogeneous treatment effects amongst participants of various self-esteem and body sizes. (This diagram includes only the version of the size buckets condensed into three). The diagram here specifies the two possible heterogeneous treatment effects. On the one side we have high self-esteem, normal self-esteem and low self-esteem and on the other side we have the three buckets for self-identified participant size. We are testing whether the two factors will have an impact on how participants evaluate the clothing garments and thus affect their purchasing intent.



XI. Results

XI.i. Purchase Intent by Outfit and Treatment Assignment

Before diving into the regression analysis, we wanted to first take a look at the breakdown of average purchase intent broken down by individual outfit, rather than by the individual participant. To reiterate, this is *not* the same as the primary outcome variable for average purchase intent that was mentioned under “Observations and Outcome Measures”. This was calculated by averaging all of the Likert responses for a particular outfit in both the control group, where size 0-4 models wore the outfit, and the treatment group, where plus-size models wore the outfit.



Above is a side by side bar chart, broken down by treatment and control responses of the average purchase intent for each of the twenty outfits. (For a table view of the specific values for average purchase intent, please refer to Table A in the Appendix below.)

Based on the graph, there are noticeable differences in the overall average purchase intent for some outfits versus others. Outfits 1, 11, 14, and 19 all have higher scores compared to the norm, while outfits 2, 13, 17, 18 have lower average purchase intent scores in comparison. The outfits with the higher scores correspond to models wearing a more casual look, like jeans, simple shirts, and sundresses, while the outfits with the lower scores were niche clothing, like flashy party dresses and gym wear. A possible explanation is the universality of casual wear and its appeal to all demographics, whereas other outfits tend to appeal mainly to a younger demographic and may not be compatible with everybody's tastes.

Based on the bar chart, we see for the most part there aren't obvious differences between control and treatment for one single outfit, with the exception of outfits 1, 9, 11, 12, 13, and 20. These correspond to a black full sleeved cocktail dress, black party dress with a slit in the leg, skinny jeans with a white full sleeve shirt, gym wear, monochrome sweater and sweats, and a white dress with cherries on it. For all of those larger differences, the average purchase intent for control is higher than that of the treatment. When examining the images more closely, there are differences in race of the models and poses that the model is standing in, which could affect the angle of the shot and therefore the presentation of the clothing.

There were repeats of particular models within the showing of images, as well as variation in photo background. While the majority were taken behind a solid color background, some had other locations, like the beach for the swimsuit photo and dressing room for another one. Racially, the vast majority of the plus size models, with repeats included, were women of color, with little to no white women. However, there was a more diverse racial representation in the set of size 0-4 models.

Overall, a good majority of outfits, around twelve of the twenty, were casual, ranging from gym wear to everyday dresses. There were approximately five pieces of party wear and formal clothing, and for the most part, casual clothes were rated more highly than the party wear, and gym clothing and party wear outfits were ones with the highest differences in average purchase intent amongst treatment and control groups.

XI.ii. Regression Analysis

XI.ii.a. Regression analysis for participant body size

Below is a table of multiple models and their corresponding results with the primary covariate being participant body size. It includes the baseline regression with just the overall treatment effect, followed by the addition of the condensed and further condensed size buckets.

The first model we ran was baseline regression, linear model, which included purchase intent and treatment assignment. The purpose of this was to see whether or not the average purchase intent varied between the treatment and control groups. The coefficient indicates the average purchase intent of the participants who were given the control version of the survey. Out of a scale of 5, those in the control group gave a rating of approximately 2.610 on average, which corresponds to the “Somewhat unlikely” range. The average purchase intent of the participants who were assigned to the treatment group was 0.16 points less on average than the control group. This coefficient was accompanied by a large standard error that implies no statistical significance and therefore no noticeable treatment effect.

Table 1: All Size Buckets Regressions

	Dependent variable:		
	Purchase Intent		
	Treatment Effect	Size Buckets	Size Buckets with HTEs
Treated	−0.160	−0.160	−0.462
	(0.171)	(0.173)	(0.382)
Size 6-10		0.358	0.210
		(0.232)	(0.341)
Size 12-16		0.315	0.114
		(0.243)	(0.345)
Size 18+		0.064	−0.199
		(0.320)	(0.444)
Treated * Size 6-10			0.297
			(0.472)
Treated * Size 12-16			0.405
			(0.497)
Treated * Size 18+			0.527
			(0.661)
Intercept	2.610***	2.369***	2.519***
	(0.118)	(0.208)	(0.277)
SE Type	Robust	Robust	Robust
Observations	144	144	144
R^2	0.006	0.028	0.034
Adjusted R^2	−0.001	−0.0002	−0.016
Note:			*p<0.1; **p<0.05; ***p<0.01

The next model incorporates ‘size buckets’ on the right hand side in addition to treatment assignment. The baseline intercept coefficient, representing the average purchase intent for all respondents in the control group who indicated they fell in the size 0-4 bucket, is approximately

2.369. This corresponds to the “Somewhat unlikely” response on the original survey question. There is no statistically significant difference across responses by various size buckets, implied by the large standard errors. The treatment effect is still negative, and although not statistically significant, this suggests that on average, participants in the treatment group are more likely to give lower purchase intent ratings to plus size models than size 0-4 models overall.

The third model involves an interaction term between the treatment assigned variable and the size bucket, and based on the table results, there are no heterogeneous treatment effects amongst the different groups. In addition, our robust standard errors are also very large.

Table 2: Condensed Size Buckets Regressions

	Dependent variable:				
	Purchase Intent				
	Treatment Effect	Condensed Size Buckets	Condensed Size Buckets with HTEs	Binary Size Bucket	Binary Size Bucket with HTEs
Treated	−0.160	−0.162	−0.016	−0.160	−0.016
	(0.171)	(0.172)	(0.272)	(0.172)	(0.272)
Below Modal Size		−0.224	−0.002		
		(0.231)	(0.328)		
Modal Size		0.134	0.208		
		(0.192)	(0.266)		
Treated * Below Modal Size			−0.446		
			(0.469)		
Treated * Modal Size			−0.149		
			(0.388)		
Modal and Below Modal Size				0.030	0.146
				(0.175)	(0.238)
Modal and Below Modal Size * Treated					−0.233
					(0.352)
Intercept	2.610***	2.593***	2.521***	2.592***	2.521***
	(0.118)	(0.153)	(0.176)	(0.153)	(0.176)
SE Type	Robust	Robust	Robust	Robust	Robust
Observations	144	144	144	144	144
R^2	0.006	0.022	0.028	0.006	0.010
Adjusted R^2	−0.001	0.001	−0.007	−0.008	−0.012
Note:					*p<0.1; **p<0.05; ***p<0.01

In this second table, we condensed the size buckets based on the modal value of size 6-10. After doing so, we still observed no statistically significant difference in treatment and control groups, nor across sizes. With the binary condensed buckets variable of “Modal and Below Modal Size” and “Above Modal Size”, we still see no statistically significant results. The treatment effect is still negative across the models, and although not statistically significant, this suggests that on average, participants in the treatment group are more likely to give lower purchase intent ratings to plus size models than size 0-4 models overall.

There is no observed heterogeneous treatment effect or difference in the way that participants of various sizes respond to purchase intent.

XI.ii.b Regression analysis for participant self esteem

Table 3: Self-Esteem Regression

	Dependent variable:		
	Purchase Intent		
	Treatment Effect	Self-Esteem	Self-Esteem with HTEs
Treated	−0.160	−0.169	−0.425
	(0.171)	(0.164)	(0.271)
Low Self-Esteem		−0.527***	−0.707***
		(0.200)	(0.268)
Neutral Self-Esteem		−0.725***	−1.003***
		(0.195)	(0.256)
Low Self-Esteem * Treated			0.364
			(0.407)
Neutral Self-Esteem * Treated			0.534
			(0.388)
Intercept	2.610***	2.973***	3.105***
	(0.118)	(0.155)	(0.186)
SE Type	Robust	Robust	Robust
Observations	144	144	144
R^2	0.006	0.106	0.119
Adjusted R^2	−0.001	0.087	0.087
Note:			*p<0.1; **p<0.05; ***p<0.01

When considering self-esteem as a covariate, we do find heterogeneous treatment effects while the main treatment effect itself is not statistically significant. When adding self-esteem as an

additional covariate, we ran three different models with treatment effect, self-esteem, and self-esteem with HTE as dependent variables respectively.

Women in the treatment group have a lower purchase intent than women in the control group in regressions including self-esteem. The coefficient on “Treated” is negative throughout the regressions, suggesting that the viewing plus-sized models instead of size 0-4 models causes women to be less likely to want to purchase clothing. However, the coefficient is insignificant throughout these regressions which means we may not reject the null hypothesis that viewing plus-sized models instead of size 0-4 models has no effect on a woman’s intent to purchase clothing.

Respondents with low-self esteem in the control group on average have a lower average purchase intent than those with high self-esteem, by 0.707 points lower on the Likert scale. Respondents with neutral-self esteem in the control group on average have a lower average purchase intent than those with high self-esteem, by 1.003 points lower on the Likert scale. Both of these coefficients are highly significant at the 1% significance level suggesting that those who do not have high self-esteem are much less likely to purchase clothing than those with high self-esteem. However, the interaction terms are insignificant suggesting that the treatment effect does not vary by self-esteem. Seeing the plus-size models appears to have the same negative effect on purchase intent across all self-esteem levels.

It is worth noting that our regression results are mostly insignificant because we have a very small sample size. As the next section shows, our experiment has very little statistical power so it would be difficult for us to detect a treatment effect and determine that viewing plus-sized models impacts a woman’s purchase intent even if such a treatment effect exists. Our regressions suggest that seeing plus-sized models decreases a woman’s purchase intent; however, our standard errors are so large (sometimes even the size of the respective coefficients) that this effect could truly be the opposite if a treatment effect exists.

XI.iii. Power Test

On the left is a power test conducted on our data, and on the right is a power test to determine the minimum sample size needed for the study to have enough power. We used 0.8 as our threshold for minimum power. The power test enables us to determine whether we need more samples to confidently explain our results. The test of choice is power t-test and we pre-set the 0.05 significance level and obtained 0.1597 power, which is very low and therefore it would be difficult for us to detect a treatment effect and determine that viewing plus-sized models impacts a woman’s purchase intent even if such a treatment effect were to exist. To determine how many participants we would need to be very likely to see a treatment effect if it exists, we ran a power test to see how many participants we’d need to achieve a power of 0.8. We would need about 1,274 participants total whereas we only have 144. Due to the lack of time, we didn’t further increase our participants’ size and we did not have the funding to achieve over a thousand participants, so we denote the statistical power as a point of future endeavor where more participants are needed in this study to further improve the power of the results.

```
     Two-sample t test power calculation

              n = 72
          delta = 0.1597222
             sd = 1.016509
      sig.level = 0.05
          power = 0.154909
    alternative = two.sided

NOTE: n is number in *each* group
```

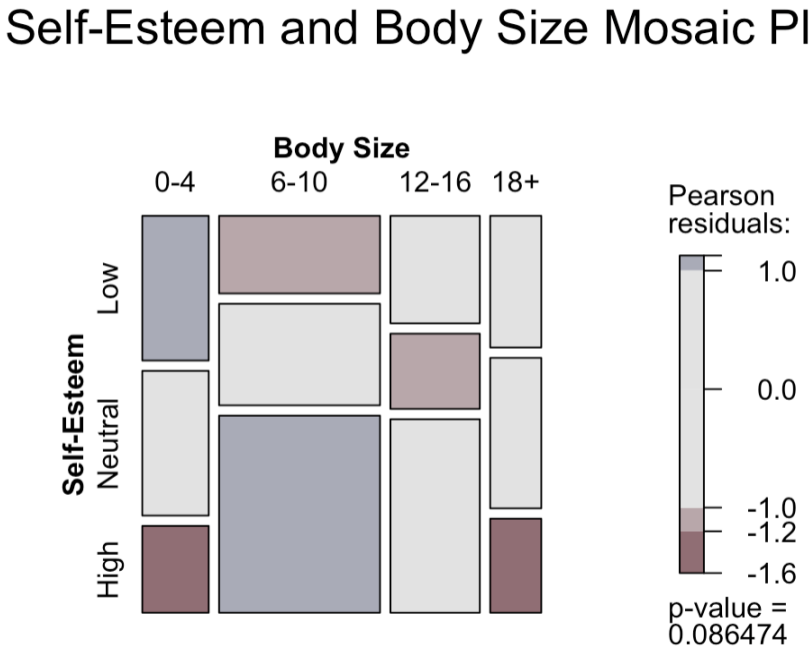
```
     Two-sample t test power calculation

              n = 636.7733
          delta = 0.1597222
             sd = 1.016509
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

XI.iv. Reported Size and Self-Esteem Analysis

Based on a study reported in the Kent University paper, there is no association between self-esteem levels and participant body size levels. To recap from our regression analysis, we ran separate models to test HTEs for body size and separate models for self-esteem, but we were curious to see if there is a relationship. If we did see a relationship, we could have potentially run more models with an interaction term between body size and self-esteem and determined the existence of HTEs for them. To confirm the result from that paper, we decided to create a mosaic plot visualization, and based on that, we do confirm independence between the two variables.

We used a chi-squared test to test independence between size and self-esteem. This is analogous to a correlation test, but since these variables are categorical, a chi-squared test is the most appropriate. We added another diagram indicating the size bucket and self-esteem independence relationship.

Instead of looking at the correlation, chi squared test was used to test the independence between the reported size bucket and self-esteem, as the variables were categorical. According to the mosaic plot below, the most common body size indicated by participants is 6-10, while neutral and low self-esteem levels were commonly reported overall. There are more observations in the cells with pearson residuals above 1.0, where the participants are in the 0-4 size bucket and have low self-esteem, and also where the participants are in the 6-10 size bucket and have high self-esteem, than would be expected. There are less observations in the cells with pearson residuals smaller than -1.2, where the participants are in the 0-4 size bucket and have high self-esteem, and also where the participants are in the 18+ size bucket and have high self-esteem, than would be expected. There is no statistically significant association between self-esteem and body size for the cells with the pearson residuals between 1.0 to -1.0. Overall, there is no statistically significant dependence between self-esteem and body-size, based on the person's residuals and p value is 0.0864, which does not pass the 5% cutoff.



XII. Conclusion

There appears to be a negative effect of plus-size models on a woman's purchase intent. All of our regressions showed that viewing plus-sized models decreased a woman's intent to purchase clothing. However, the standard errors were incredibly large so the coefficients were mostly insignificant, suggesting that there could be no effect of viewing plus-size models instead of size 0-4 models on a woman's intent to purchase clothing or there could actually be a positive effect.

We saw no evidence of heterogeneous treatment effects based on a woman's body size or self-esteem. However, for self-esteem levels, we do see some differences in baseline purchase intent. Based on statistically significant evidence at the 0.01 cutoff, our results indicate that women with low or neutral self-esteem generally have lower purchase intent than women with high self-esteem. After separate regression analyses for self-esteem and body size, we wanted to determine whether these two covariates were associated with one another. To do so, we created a mosaic plot, and found there to be no statistically significant dependence between self-esteem and body-size.

In conclusion, most of our results were insignificant and we could not confirm a treatment effect. However, our sample size was small and thus our experiment lacked power meaning we were very unlikely to detect a treatment effect even if there were one. If we had a larger sample size,

and therefore more power, we could be more certain about our results and be more likely to detect a treatment effect if it exists.

XIII. Next Steps and Future Iterations

A larger sample size beyond 144 survey respondents would yield more power for the study. As stated above in the “Power Test” section, more than 1,274 participants would be ideal. We would like to confirm the negative effect of average purchase intent for plus size models with a larger sample size. With an even larger sample size, we could possibly explore and research other covariates and its link to self-esteem or participant body type. If we can find further studies conducted on the variability of self-esteem levels or body types, we could evaluate for more heterogeneous treatment effects for other covariates.

Incorporating a more diverse set of clothing might help to increase power, as the majority of the clothing presented in the survey appeals to a younger audience. With a more diverse set of clothing, we can break different designs into categories and block on the different styles of clothing in the experiment as well. In addition, in the future research, presenting models from a broader range of age might be useful, as the majority of models were within the 20-30 years range. With a more diverse set of clothing, we can block on various types of clothing, like casual vs. formal, in order to look for possible heterogeneous treatment effects within style.

XIV. References

1. Deborah A. Christel & Susan C. Dunn (2017) Average American women's clothing size: comparing National Health and Nutritional Examination Surveys (1988–2010) to ASTM International Misses & Women's Plus Size clothing, International Journal of Fashion Design, Technology and Education, 10:2, 129-136, DOI: 10.1080/17543266.2016.1214291

2. Perrier, Angela Michelle, "Influence of model body size on consumer attitudes and purchase intention" (2008). Graduate Theses and Dissertations. 11393. https://lib.dr.iastate.edu/etd/11393

3. Cash, T. F., & Pruzinsky, T. (Eds.). (1990). *Body images: Development, deviance, and change.* Guilford Press.

4. Bian, X. and Wang, K.-Y. (2015), "Are size-zero female models always more effective than average-sized ones? Depends on brand and self-esteem!", European Journal of Marketing , Vol. 49 No. 7/8, pp. 1184-1206. https://doi.org/10.1108/EJM-08-2013-0414

5. Ohanion, R. (1990), "Construction and validation of a scale to measure celebrity endorser's perceived expertise, trustworthiness, and attractiveness", *Journal of Advertising,* Vol. 19 No. 3, pp 39-52

XV. Appendix

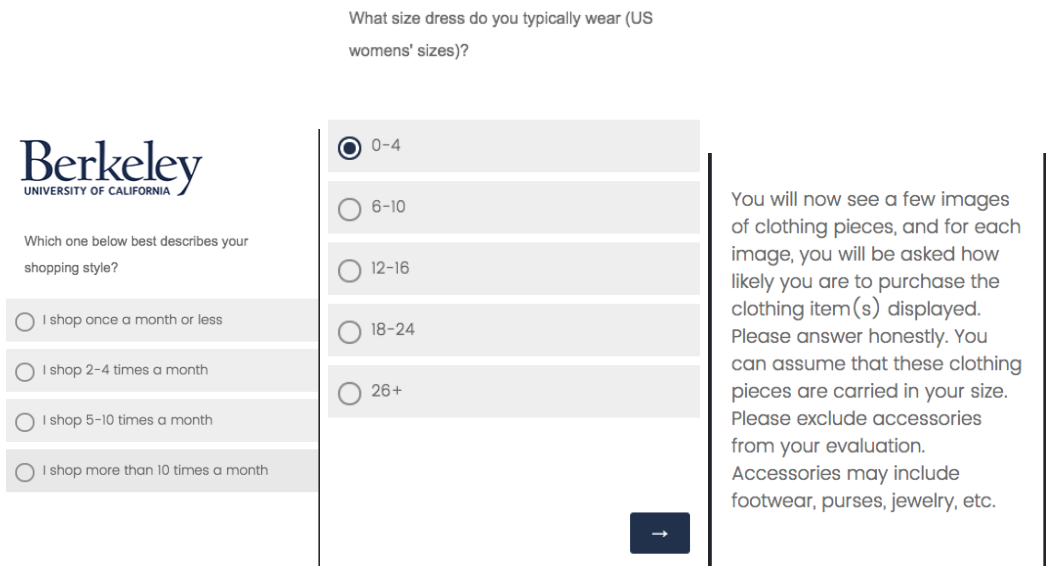
Slack Message posted in women-in-mids channel to collect pilot data:

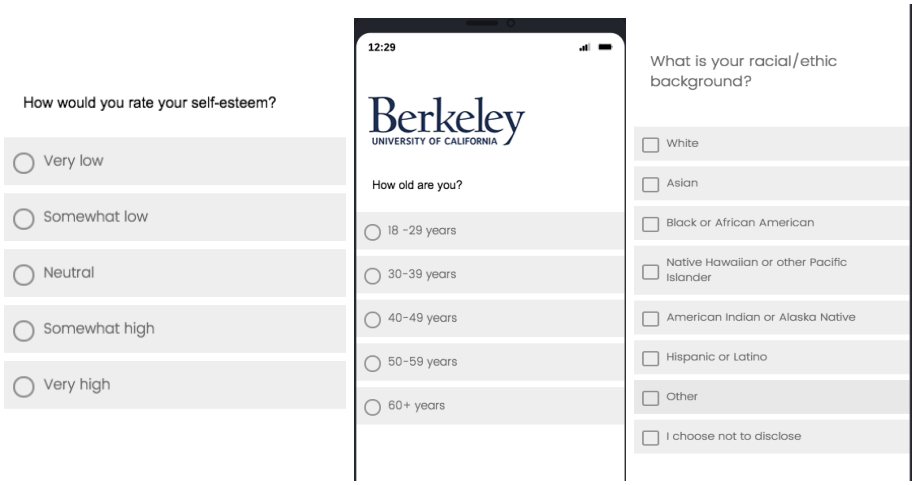
Hey everyone! I am working on a group project for W241. If you are woman-identifying, please fill out the survey below. My teammates and I appreciate your time. Thank you! 🙂 https://berkeley.qualtrics.com/jfe/form/SV_6FOQmUFsVyEttgq (edited)

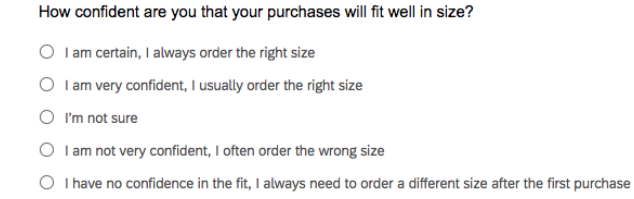
Sample Images for 3 of 20 total outfits

Top row images were given to participants in the control group and consist of size 0-4 models. Bottom row images were given to participants in the treatment group and consist of plus size models.

A few of the Survey Questions + Prompts:







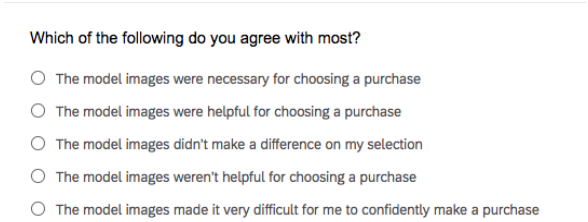


Table A:

Table 4: Table of Average Purchase Intent by Outfit

Outfit	Control	Treatment	Difference
1	2.930556	2.430556	-0.5000000
2	1.708333	1.875000	0.1666667
3	2.583333	2.291667	-0.2916667
4	2.152778	2.361111	0.2083333
5	2.333333	2.527778	0.1944444
6	2.250000	2.430556	0.1805556
7	2.569444	2.611111	0.0416667
8	2.583333	2.694444	0.1111111
9	2.638889	2.236111	-0.4027778
10	2.500000	2.291667	-0.2083333
11	3.805556	3.083333	-0.7222222
12	2.777778	2.138889	-0.6388889
13	2.388889	2.013889	-0.3750000
14	3.305556	3.097222	-0.2083333
15	2.694444	2.597222	-0.0972222
16	2.708333	2.583333	-0.1250000
17	2.013889	2.069444	0.0555556
18	2.208333	2.277778	0.0694444
19	3.305556	3.027778	-0.2777778
20	2.750000	2.375000	-0.3750000

Summary Statistics

treated		0		1		
Variable	**N**	**Percent**	**N**	**Percent**	**Test**	
age	72		72		X2=1.034	
... 18 -29 years	14	19.4%	13	18.1%		
... 30-39 years	24	33.3%	21	29.2%		
... 40-49 years	11	15.3%	13	18.1%		
... 50-59 years	17	23.6%	16	22.2%		
... 60+ years	6	8.3%	9	12.5%		
race	72		72		X2=3.081	
... Asian	7	9.7%	8	11.1%		
... Black or African American	1	1.4%	5	6.9%		
... Hispanic or Latino	3	4.2%	2	2.8%		
... Mixed/Other	5	6.9%	5	6.9%		
... White	56	77.8%	52	72.2%		
self_esteem_rate	72		72		X2=0.56	
... High	30	41.7%	32	44.4%		
... Low	22	30.6%	18	25%		
... Neutral	20	27.8%	22	30.6%		
shopping_style	72		72		X2=4.584	
... shop 2-4 times per month	25	34.7%	32	44.4%		
... shop 5+ times per month	19	26.4%	9	12.5%		
... shop less than once per month	28	38.9%	31	43.1%		
shop_online	72		72		X2=0.482	
... about equal	24	33.3%	28	38.9%		
... mostly in person	13	18.1%	12	16.7%		
... mostly online	35	48.6%	32	44.4%		
shopping_preferences	72		72		X2=0.379	
... high browsing	44	61.1%	46	63.9%		
... low browsing	3	4.2%	4	5.6%		
... medium browsing	25	34.7%	22	30.6%		
shopping_exp_description	72		72		X2=1.096	
... sometimes satisfied	29	40.3%	25	34.7%		
... usually frustrated	8	11.1%	12	16.7%		
... usually satisfied	35	48.6%	35	48.6%		

size_bucket	72		72		X2=0.044
... 0-4	13	18.1%	13	18.1%	
... 12-16	18	25%	17	23.6%	
... 18+	10	13.9%	10	13.9%	
... 6-10	31	43.1%	32	44.4%	
size_bucket_condensed	72		72		X2=0.034
... Above Modal Size	28	38.9%	27	37.5%	
... Below Modal Size	13	18.1%	13	18.1%	
... Modal Size	31	43.1%	32	44.4%	
size_bucket_condensed_binary	72		72		X2=0
... Above Modal Size	28	38.9%	27	37.5%	
... Modal and Below Modal Size	44	61.1%	45	62.5%	
size_purchase_confidence	72		72		X2=1.054
... confident	49	68.1%	48	66.7%	
... not confident	12	16.7%	9	12.5%	
... sometimes confident	11	15.3%	15	20.8%	
model_image_usefulness	72		72		X2=9.88**
... The model images didn't make a difference on my selection	12	16.7%	15	20.8%	
... The model images made it very difficult for me to confidently make a purchase	1	1.4%	0	0%	
... The model images were helpful for choosing a purchase	43	59.7%	28	38.9%	
... The model images were necessary for choo sing a purchase	13	18.1%	18	25%	
... The model images weren't helpful for choosi ng a purchase	3	4.2%	11	15.3%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01