

Beauty is in the Eye of the Ad-Holder

Ashton Chevallier, Lisa Minas, Olivier Zimmer

Summer 2017 - MIDS W241 - Causal Experimental Design

Background

In a 2011 survey, Dove found that only 2% of women around the world consider themselves beautiful and 68% of women could not identify with the models in advertising. In response, Dove launched numerous marketing campaigns, created educational programs, and donated money “to help women everywhere develop a positive relationship with the way they look, helping them raise their self-esteem and realize their full potential”¹. But have these efforts been effective in changing how women are portrayed in media and marketing or how women view themselves?

One of Dove’s campaigns was Image_Hack² which aimed to “change the stereotypical way women are portrayed in media and marketing.” In Image_Hack, Dove and partnering company Mindshare uploaded numerous images of strong, active, everyday women to the site ShutterStock. Each of these non-stereotypical portraits of women was labeled “beautiful woman”. Given the quantity uploaded and the search algorithm on ShutterStock, the campaign claimed to have successfully changed the search results in Denmark to return this ‘real’ version of women.



Image 1: (left) stereotypical search results for beautiful woman; (right) non-stereotypical search results for beautiful woman

Research Question

We plan to research if the concept behind Dove’s ImageHack campaign would have a causal impact on women’s relationship with their self-image and self-esteem. Dove’s concept of emphasis of real women as the standard of beauty logically makes sense: if the average woman has a negative relationship with her body because she doesn’t fit the standard of beauty created by the media, then changing the standard of beauty to be a realistic ideal should improve her relationship with her body.

¹ <https://www.dove.com/us/en/stories/about-dove/our-vision.html>

² <http://imagehack.org>

An experiment is necessary to affirm this theory. By comparing women's self-esteem and self-image after seeing a stereotypical or non-stereotypical advertisement, we can quantify the impact of beauty in advertisements. Without running such an experiment, we won't be able to judge if Dove's alternative images change women's perception. For example, maybe women who aren't happy with the media seek out alternative images and stumble upon Dove's ImageHack campaign. The treatment does not cause these women's opinions but instead the women were predisposed to prefer the non-stereotypical images.

Hypothesis

We hypothesize that, after seeing a 30 second treatment video advertisement, our subjects will:

H_{1A} : Report feeling more confident

H_{1B} : Report feeling more beautiful

H_{1C} : Report feeling more related to the advertised models

H_{1D} : Prefer advertisements featuring non-stereotypical models

Study Design

To test these hypotheses, we designed a survey that would expose subjects to a 30 second video advertisement and then ask them a series of questions related to their opinions about their self-image, confidence, importance of beauty, and preferences of advertisements.

Treatment

Subjects were instructed to watch a 30 second video which features stereotypical models (control) or non-stereotypical models (treatment). Video advertisements were chosen as treatment to maximize treatment effect while maintaining a realistic advertisement setup. For example, a five-minute documentary on Dove's campaign would have a stronger treatment effect but would not be a realistic advertising experience. Print advertisements would better simulate the Dove campaign but may not be a strong enough effect to measure impact.

Ideally our videos would be identical except for the model. Unfortunately, we were unable to find such examples and we had neither the budget nor ability to create such advertisements. Instead, we used two advertisements that were relatively similar (product, recognizability of brand, quality of video) but with a stereotypical or non-stereotypical model.

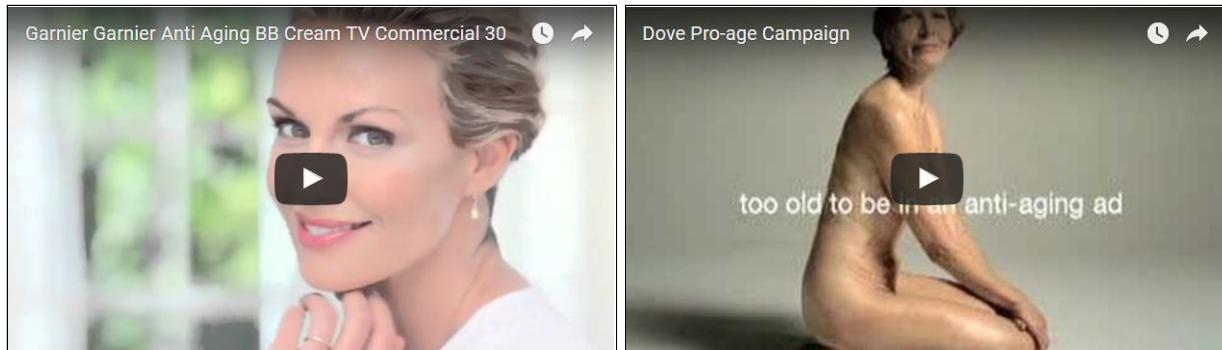


Image 2: (left) control video advertisement; (right) treatment video advertisement

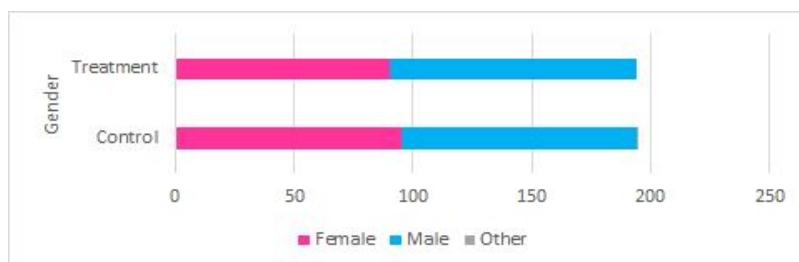
Subjects

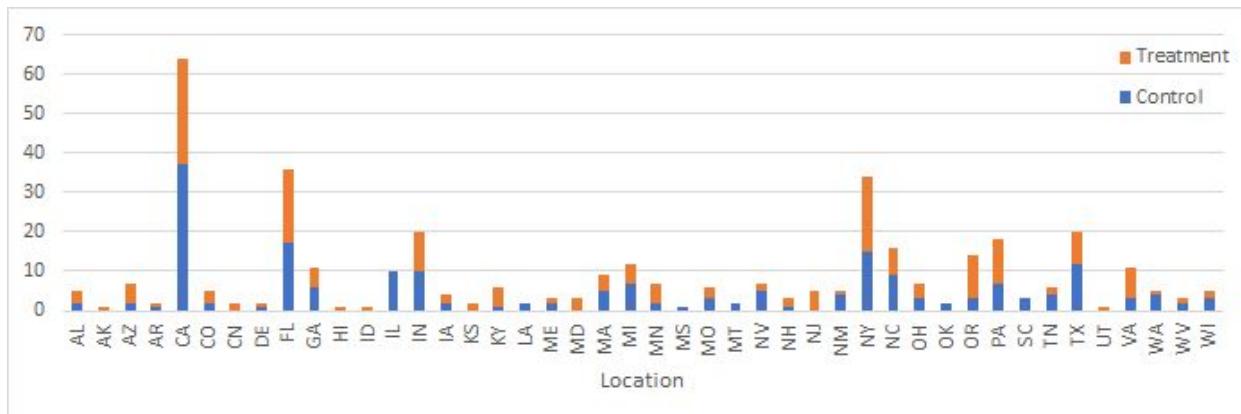
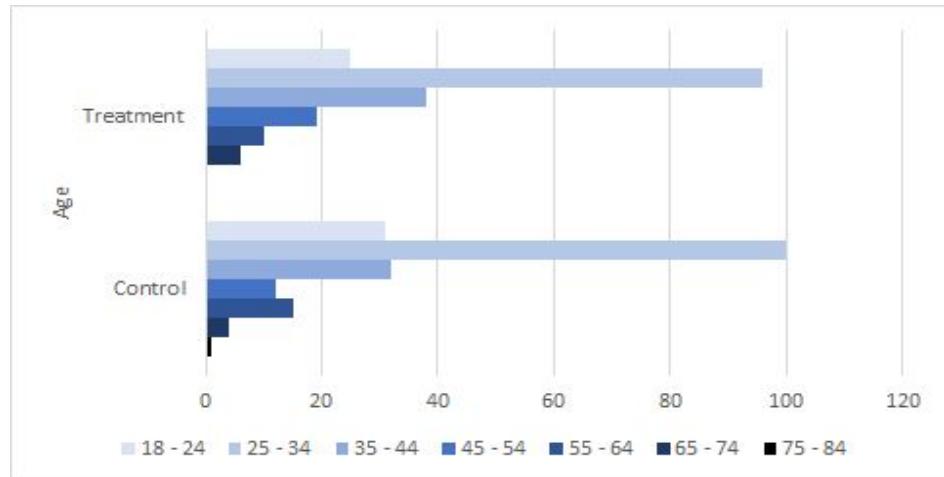
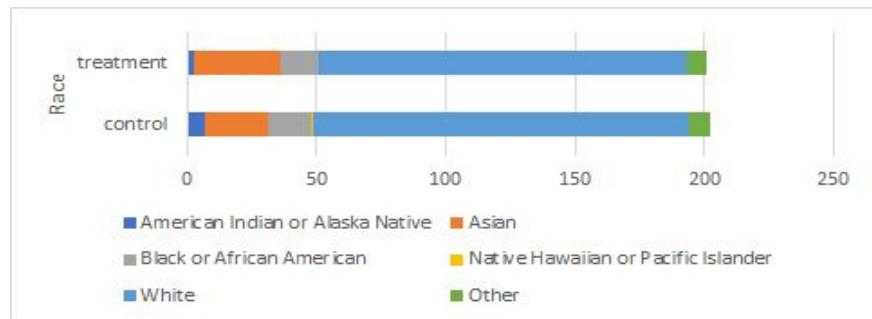
Dove's campaign focuses on women around the globe. We chose to gather our subjects from Amazon's Mechanical Turk because the workers are a similarly broad, diverse group of consumers. Furthermore, the workers are easy to recruit allowing us to quickly get a large sample. We accepted both men and women into our experiment because advertising is seen by the entire population and men's reactions to women's advertising would be an interesting baseline to compare to women's reactions.

To stay within budget, we restricted our subjects to those living in the United States. Mechanical Turk places a restriction that workers be 18 years of age or older. To ensure full effect from treatment, we rejected subjects who did not have audio or did not read the questions. Reading validation questions ensured active participation and intentional answers; we know Mechanical Turk workers will sometimes speed through surveys without considering their answer which would result in inaccurate data.

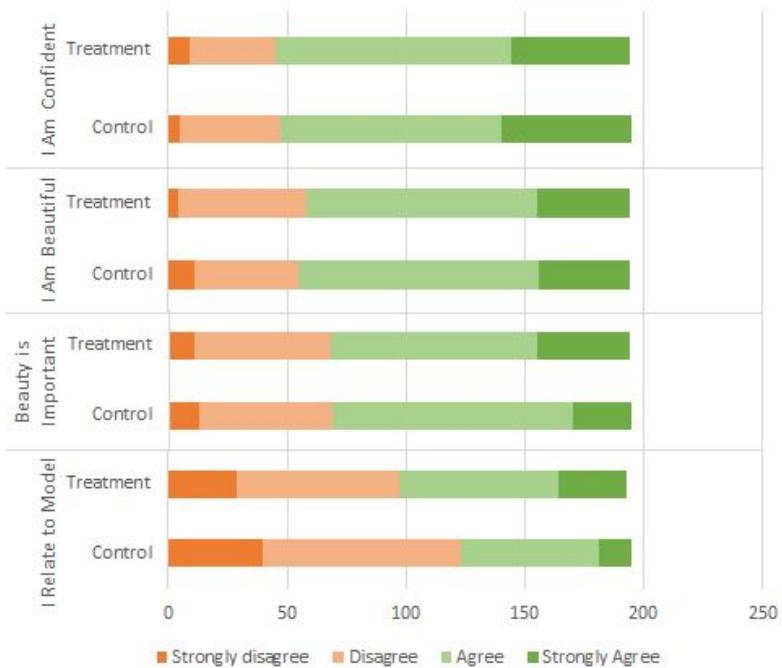
EDA and Randomization Validation

We used the randomizer element of Qualtrics to randomly assign treatment or control equally to our subjects blocked by gender. It appears that our randomization scheme worked well. We see even splits between genders, proportionally distributed across the United States, and race roughly comparable to the US population. A major demographic difference is that our subjects are younger compared to the US age distribution. We based our expectations on the 2010 US census.

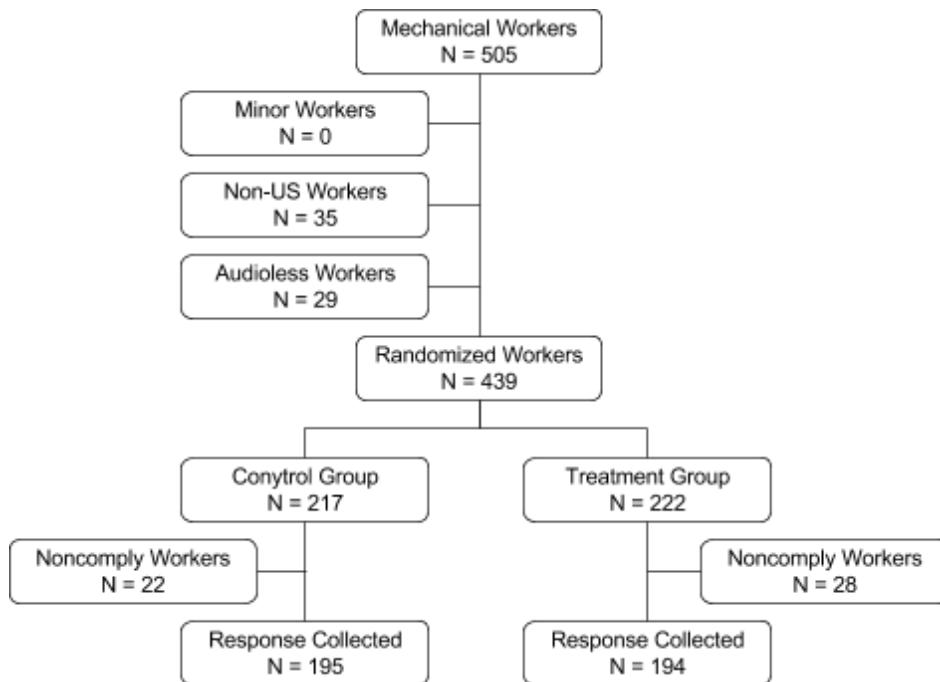




It is surprising to see that our subjects had very high opinions of themselves. 71% of our subjects agree that they are beautiful and 76% agree they are confident. We don't have insight into how Dove measured women's opinion on beauty or how they generated their sample but we believe there must be a large difference in study design; to go from 2% in 2011 to 76% in 2017 indicates to us a difference in sample more so than a change in the underlying population.



Observations Throughout the Experiment



Experimental Design

Our experiment (randomized assignment into treatment or control), shows subjects either a treatment or control 30 second video advertisement, then asks subjects a series of questions (see Appendix A). Questions consisted of demographics, direct answer, and conjoint analysis.

All response choices were randomized within each question and questions were randomized within each section. Randomization help avoid unintentional systematic bias (e.g. always preferring the advertisement on the left).

Experimental Design			
Treatment Group	R	X	O
Control Group	R	-	O

By randomizing subjects prior to treatment, we will eliminate systematic or confounding factors. We will then be able to assume the subjects within treatment and control groups are equivalent and any difference in response is due to treatment and not an unknown variable. For example, if a subject actively chose to watch the treatment video advertisement, they may be predisposed to prefer non-stereotypical advertisements and their results would reflect this predisposition instead of causal impact of the treatment.

Our observations are subject responses to survey questions. Each question aims to identify an aspect of the subject's view on beauty: how important is beauty, do they identify as beautiful, are they confident, how do they believe other's value beauty. Half of our questions ask subjects how strongly they agree with statements such as "I am beautiful" and "I am confident". The other half of our questions follow a conjoint analysis setup. Subjects are shown five pairs of print advertisements for a fake brand, one featuring a stereotypical model and the other featuring a non-stereotypical model. We created the print advertisements and are for a brand campaign instead of for a specific product (see Appendix A for all images). Subjects were asked to choose which advertisement they identify with and prefer. By asking subjects to choose between advertisements, we captured subconscious choices that subjects may not report when consciously making an explicit choice such as "Do you prefer stereotypical or non-stereotypical models?". Conjoint analysis simulates real experiences a subject may encounter and allows us deeper insight into the subject's true preferences. For example, although a subject may know she shouldn't prefer the photoshopped stereotypical model, she may still be drawn to a stereotypical model while easily ignoring a non-stereotypical model.

Comparing between-subject responses to post-treatment questions allows us to assess the causal impact the treatment has on subject perceptions. If the treatment group shows higher agreement with statements such as "I am beautiful" and "I am confident", then we can claim the treatment of non-stereotypical models caused improvements in self perception.

Analysis

Pilot survey

A first pilot study was conducted between 2017-07-04 and 2017-07-07 that collected responses from 70 within our personal networks. The main goal of the pilot was to ensure that the experimental pipeline was running properly. Based on pilot subject responses, we added a screening question to check if subjects enabled audio, as we noticed that some respondents

muted their speakers which diminished the treatment effect. The audio check asked subjects to correctly choose *pineapple* as the word said during a three-second video.

The pilot also provided preliminary estimates to treatment effect and standard deviation. Using the formula below, we calculated a sample size of $n = 400$ for the study to have sufficient power. We confirmed the adequate sample size using randomization inference on our collected data.

$$n = \left[\frac{Z_{\alpha/2} \sigma}{E} \right]^2$$

n = sample size
 σ = standard deviation
 E = effect size
 Z = critical value

First survey

The first survey ran on 2017-07-10 on Mechanical Turk and collected 390 valid responses. To analyse the responses to the direct questions (see Appendix A), we transformed the ordinal answers ("I strongly agree", "I agree", "I disagree", "I strongly disagree" into numerical answers (2, 1, -1, -2, respectively). We built regression models on the transformed response per question using several demographic covariates. We decided on a final model just using gender as a covariate as the more complicated models lacked statistical significance and provided little additional insight.

Dependent variable:				
	Beautiful	Confident	Importance	Relate
	(1)	(2)	(3)	(4)
GroupTreatment	-0.001 (0.094)	0.017 (0.090)	-0.058 (0.102)	0.244** (0.110)
GenderMale	-0.217** (0.095)	0.124 (0.090)	-0.181* (0.102)	-0.459*** (0.110)
GenderOther	-2.302** (0.935)	-1.153 (0.892)	-2.200** (1.007)	-1.703 (1.092)
Constant	0.302*** (0.082)	0.153* (0.079)	0.200** (0.089)	-0.297*** (0.096)
Observations	390	390	390	390
R2	0.027	0.010	0.020	0.058
Adjusted R2	0.020	0.002	0.012	0.051
Residual Std. Error (df = 386)	0.931	0.888	1.003	1.088
F Statistic (df = 3; 386)	3.597**	1.280	2.599*	7.954***

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 1. Results for the four direct questions - data from 2017-07-10

Model 1 ("I am beautiful") and Model 3 ("Being beautiful is important") highlights a significant relation with the covariate gender but not with treatment. Female subjects tended to feel more beautiful than male subject which makes sense as males might feel handsome rather than beautiful. Model 2 ("I am confident") is statistically insignificant across gender and treatment. Model 4 ("I relate with the model in the video") shows statistical significance of treatment at the 0.05 level. This is the only model where treatment shows significance and supports our hypothesis that seeing a non-stereotypical model increases relation with the model. We confirmed the significance of our results and p-values by running through a randomization inference.

For the second part of the experiment we ran a conjoint analysis where subjects are asked to choose amongst print advertisement pairs, one featuring a stereotypical model while the other features a non-stereotypical model. We averaged responses by question across the five ads displayed by subjects; instead of looking at a subject's response per ad where there may be too much noise, we decided to look at a subject's response across all advertisement pairs. This highlights overall treatment effect not specific to a given pair.

Dependent variable:				
	Identify (1)	data[[question_average]] I Prefer (2)	Others Prefer (3)	Validate (4)
data[["Group"]][Treatment]	-0.081 (0.139)	-0.154 (0.131)	0.165 (0.147)	-0.011 (0.008)
GenderMale	0.351** (0.139)	0.677*** (0.132)	0.584*** (0.147)	0.001 (0.008)
GenderOther	-0.392 (1.321)	-0.689 (1.249)	2.418* (1.402)	0.0004 (0.075)
Constant	1.392*** (0.123)	1.689*** (0.116)	2.582*** (0.130)	5.000*** (0.007)
Observations	360	360	362	362
R2	0.019	0.074	0.051	0.006
Adjusted R2	0.011	0.066	0.043	-0.003
Residual Std. Error	1.315 (df = 356)	1.244 (df = 356)	1.396 (df = 358)	0.074 (df = 358)
F Statistic	2.284* (df = 3; 356)	9.430*** (df = 3; 356)	6.415*** (df = 3; 358)	0.663 (df = 3; 358)

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 2. Results for print advertisement conjoint analysis- data from 2017-07-10

We did not see statistical significance for treatment within the three test questions or the validation question. The validation question proved treatment did not cause people to rush answering our survey. Statistical significant for test questions is seen with gender. We were surprised to see being male increased identifying, preferring, and believing others prefer the treatment advertisement within a pair, especially since all models used were female.

When running a regression on the 20 questions (five advertisements x four questions) independently, we did find a significant relation between treatment and the question “I prefer non-stereotypical advertisements” for the *Pursue Your Passion* advertisement.



Image 3. *Pursue Your Passion* advertisement pair

The issue here is whether we measured a real effect or if this was due to chance. Out of 20 questions, we would expect to get by chance one significant result (95% threshold). In order to confirm if this was a real effect, our team decided to replicate the experiment.

Second survey

The second survey ran on 2017-08-10 on Mechanical Turk collecting 405 valid responses. We ran similar analysis on respondents output. The direct questions regression model had a similar output confirming treatment caused an effect on the question “I can relate with the model in the video” .

For the print advertisement pairs, “I prefer non-stereotypical ads” for the *Pursue Your Passion* advertisement was not significant in the second survey. But as shown on Figure 3, the output from the second survey this time gave significant relation between the treatment and the ‘I believe other prefer...’ (model 3). It’s hard for us to make a causal claim that ‘Others Prefer’ is actually a real effect from treatment, since our first experiment had opposite results. It seems this is more likely luck of the draw.

Dependent variable:						
	data[[question_average]]					
	Identify I Prefer Others Prefer Validate	(1)	(2)	(3)	(4)	
data[["Group"]]		Treatment	-0.109 (0.134)	-0.104 (0.137)	-0.267** (0.135)	0.017 (0.012)
GenderMale			0.471*** (0.136)	0.609*** (0.139)	0.281** (0.137)	0.010 (0.012)
GenderOther			-1.520 (1.293)	-1.870 (1.321)	-0.012 (1.306)	0.028 (0.116)
Constant			1.520*** (0.124)	1.870*** (0.127)	3.012*** (0.126)	4.972*** (0.011)
<hr/>						
Observations			370	370	370	370
R2			0.038	0.058	0.022	0.007
Adjusted R2			0.030	0.050	0.014	-0.001
Residual Std. Error (df = 366)			1.287	1.315	1.300	0.116
F Statistic (df = 3; 366)			4.831***	7.504***	2.698**	0.891
<hr/>						
Note:	*p<0.1; **p<0.05; ***p<0.01					

Figure 3. Results for print advertisement conjoint analysis - data on 2017-08-10

We concluded that only the results indicating relatability of the video models were truly causal as this was the only finding statistically significant in both experiments. This approach permitted us to better test for the 'chance' effect that can occur when running a 24 questions survey and validate our findings.

Our primary conclusion is that, when exposed to non-stereotypical models, female respondents tend to relate more to the model. There are multiple consequences from this finding, one of them is that advertisers could see it as a way to improve their advertising effectiveness by including more non-stereotypical models.

Future Research

We expect our lack of significant results is partially due to the high baseline self-image and confidence shown by our subjects. If a subject already has a very positive relationship with herself, we would not have much head room to improve that relationship. In future research we would like to explore if we see the same patterns given alterations in the experimental design, sample size, or sample population.

Our treatment of one 30 second video advertisement may not be a strong enough treatment to cause noticeable change. Furthermore, there may have been issues with the content of the video advertisement which would weaken the treatment effect. Anti-aging skin care products would not appeal to a younger audience and may diminish impact of treatment. The models in

the treatment advertisement were not wearing clothes while the models in the control advertisement were wearing clothes; the lack of clothing was so upsetting for one subject that she wrote us a complaint (see Appendix B - it's amazing, don't skip it). Choosing a different advertisement that appeals to a broader audience or two advertisements that are more similar may enhance our results. If budget were of no concern, we could even go as far as to shoot 2 identical ads, with the only difference being the model. We could additionally, experiment with longer or more advertisements, but we would want to balance what is a realistic advertising exposure with intensifying treatment.

Another way to measure a stronger effect is to track which print advertisement pair was first seen after treatment. This first pair would have the closest relation to treatment in time and may be more indicative of impact. Further, if treatment only caused impact on the next advertisement seen, we would lose this impact given the 'noise' of the other pairs. Due to our randomization scheme, the order in which subjects were exposed to the print advertisement pairs was not recorded.

Although we calculated 400 subjects as a sufficient sample size for our experiment, we may have overestimated the treatment effect and actually needed a larger sample size for sufficient power. Increasing the sample size or better filtering our sample population may produce additional significant results.

Further research could also be conducted on market penetration of non-stereotypical advertisement. One 30 second video advertisement cannot compete with the daily, constant barrage of stereotypical advertising. Dove alone championing the non-stereotypical models may similarly not be sufficient to impact consumer perceptions. An experiment where the majority of advertisements experienced by a subject over an extended period of time may provide enough stimulation to make lasting impression.

Appendix A - Survey

Welcome

I agree to participate in a research study conducted by masters students from the University of California - Berkeley. In order to analyze responses to the questionnaire, my answers will be recorded. No identifying information about me will be made public, and any views I express will be kept completely confidential. My participation is voluntary. I am free to withdraw from the study at any time. Please select one of the following options. If you choose not to participate, the survey will end immediately.

I agree

I do not agree

Skip To: End of Survey If Welcome = I do not agree

What is your age?

Under 18

18 - 24

25 - 34

35 - 44

45 - 54

55 - 64

65 - 74

75 - 84

85 or older

Skip To: End of Survey If Age = Under 18

Which gender do you identify with?

Male

Female

Other

Choose one or more races that you consider yourself to be:

- White
 - Black or African American
 - American Indian or Alaska Native
 - Asian
 - Native Hawaiian or Pacific Islander
 - Other
-

In which state do you currently reside?

Alabama

Alaska

Arizona

Arkansas

California

Colorado

Connecticut

Delaware

District of Columbia

Florida

Georgia

Hawaii

Idaho

Illinois

Indiana

Iowa

Kansas

Kentucky

Louisiana

Maine

- Maryland
- Massachusetts
- Michigan
- Minnesota
- Mississippi
- Missouri
- Montana
- Nebraska
- Nevada
- New Hampshire
- New Jersey
- New Mexico
- New York
- North Carolina
- North Dakota
- Ohio
- Oklahoma
- Oregon
- Pennsylvania
- Puerto Rico

- Rhode Island
- South Carolina
- South Dakota
- Tennessee
- Texas
- Utah
- Vermont
- Virginia
- Washington
- West Virginia
- Wisconsin
- Wyoming
- I do not reside in the United States

Skip To: End of Survey If Location = I do not reside in the United States (53)

End of Block

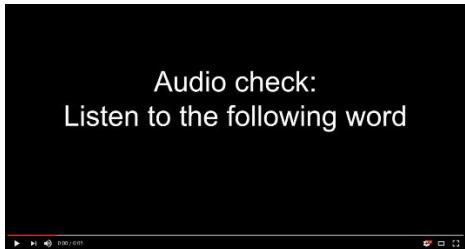
Task Description

Next, you will listen to a quick audio check and then watch a 30 second video advertisement for an anti-aging skin cream from a well-known beauty and skin-care company. Please turn up your volume as the commercial contains both audio and video. After, we will ask your opinion about five pairs of potential print advertisements for a new brand similar to this company. Please select one advertisement for each question, even if you aren't entirely sure.

End of Block

Audio Check





We are checking that your audio is functioning. What word was said during the audio check?

- Pineapple
- Elephant
- Water
- Bicycle
- Chicago
- Audio
- Leg
- Madame

Skip To: End of Survey If AudioCheck != Pineapple

End of Block

Video_Control



End of Block

Video_Treatment



End of Block

Personal Views



Please indicate how strongly you agree or disagree with the following statements.

	Strongly disagree	Disagree	Agree	Strongly Agree
I am confident	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am beautiful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being beautiful is important	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can relate with models in the video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block

Passion



Please consider these advertisements

	
Ad 1	Ad 2

I identify most with the model in...

Ad 1

Ad 2

I prefer...

Ad 1

Ad 2

I believe others will prefer...

Ad 1

Ad 2

Reading test: Select Ad 1.

Ad 1

Ad 2

Skip To: End of Survey If Passion_validate_1 != Ad 1

End of Block

Coffee



Please consider these advertisements

 <p>Vida</p> <p>Walking up to your beautiful life</p>	 <p>Vida</p> <p>Walking up to your beautiful life</p>
Ad 1	Ad 2

I identify most with the model in...

Ad 1

Ad 2

I prefer...

Ad 1

Ad 2

I believe others will prefer...

Ad 1

Ad 2

Reading test: Select Ad 1.

Ad 1

Ad 2

Skip To: End of Survey If Coffe_validate_1 != Ad 1

End of Block

Couple

X

Please consider these advertisements



Ad 1

Ad 2

I identify most with the model in...

Ad 1

Ad 2

I prefer...

Ad 1

Ad 2

I believe others will prefer...

Ad 1

Ad 2

Reading test: Select Ad 1.

Ad 1

Ad 2

Skip To: End of Survey If Couple_validate_1 != Ad 1

End of Block

Work

X

Please consider these advertisements



Ad 1



Ad 2

I identify most with the model in...

Ad 1

Ad 2

I prefer...

Ad 1

Ad 2

I believe others will prefer...

Ad 1

Ad 2

Reading test: Select Ad 1.

Ad 1

Ad 2

Skip To: End of Survey If Work_validate_1 != Ad 1

End of Block

Fitness

X

Please consider these advertisements

	
Ad 1	Ad 2

I identify most with the model in...

Ad 1

Ad 2

I prefer...

Ad 1

Ad 2

I believe others will prefer...

Ad 1

Ad 2

Reading test: Select Ad 1.

Ad 1

Ad 2

Skip To: End of Survey If Fit_validate_1 != Ad 1

End of Block

Validation

Thank you for completing this survey. Please enter the code below in Mechanical Turk to receive payment.

JCy2dRc3hQ

End of Block

Appendix B - Subject Response to Treatment

Hi, Just for the record, as there was nowhere in the survey to say anything about the video ad, I found the nudity unnecessary and inappropriate somehow. I am a woman of a certain age and I am well-preserved, as they say, but the idea of exposing myself at my age to the public nude just seems silly, inappropriate, exhibitionistic, and in poor taste, and what would I be trying to prove? Therefore I can't identify with the models at all, so the ad wouldn't work for me. Just sayin.' Thank you for the opportunity to participate. Kindest regards, Kim

