

Short-Term Influences on Personality

W241 Experiments & Causal Inference Fall 2021

Allison Rozich, Allison Schlissel, Andy Tertzakian, Jenny Conde, Trevor Johnson

December 7, 2021

Abstract

How sensitive are personality test results to external stimuli, such as video and music? We conducted a study to investigate this question and identify the potential causal relationship between watching a sad video and personality test results. Our treatment group watched and listened to a short advertisement that is traditionally considered to evoke negative emotions, such as sadness, and our control group was provided with a neutral how-to documentary to watch. Both groups then answered a series of questions from the Five Factor Personality Test (FFPT), as well as additional questions about demographic covariates and “quiz questions” to identify compliers that actually watched the videos. We then evaluated the difference in quantitative scores across the five domains between treatment and control and determined that the treatment group scored significantly lower in the areas of agreeableness, extraversion, and neuroticism. While we may not be able to generalize these specific results to different video and music types or the larger U.S. population, these data imply that caution should be used when relying on “objective” personality tests in research as they are still prone to some subjective variation due to external factors.

Research Question & Background

Personality tests are used widely by individuals looking to better understand themselves, how they react to situations, and their strengths and weaknesses. Because of this, having objective, unbiased personality tests is important in providing accurate results. However, many personality tests have been found to be subjective, with results swinging based on the test-taker’s current mood and emotions. In 1996, Heide et al. found that several personality measurement scales were unstable when groups of people were exposed to different emotion-evoking videos.¹ This problem plagues the study of personality, causing many tests, such as Myers-Briggs, to be discarded in favor of more empirical and rigorously-developed personality tests, such as the Five Factor Personality Test (FFPT).²

Despite being considered the most scientifically robust personality test and a staple in personality research, the FFPT has even shown questionable results in experimental studies. In 2014, Querengässer et al. administered a personality test to experimental subjects twice over the course of one month: once with no treatment (the neutral state) and once in an emotional state, after exposure to positive or negatively associated films, songs, and memories.³ Subjects were asked to rate the strength of their happiness or sadness and then to complete a 60-question FFPT. Researchers measured and analyzed the difference between an individual’s scores for each of the five domains of the FFPT. These results suggest that a user’s emotional state influences

¹Heide, M., Grønhaug, K. The impact of mood on personality measures. *Qual Quant* 30, 1–17 (1996). <https://doi.org/10.1007/BF00139832>

²Chen, A. How accurate are personality tests? *Scientific American* (Online). 08 Oct, 2018. Retrieved from <https://www.scientificamerican.com/article/how-accurate-are-personality-tests/> 19 Sept 2021.

³Querengässer, J., Schindler, S. Sad but true? - How induced emotional states differentially bias self-rated Big Five personality traits. *BMC Psychol* 2, 14 (2014). <https://doi.org/10.1186/2050-7283-2-14>

personality test measures.⁴ However, we are investigating the sensitivity of this phenomenon: Can a *single* video influence mood and personality to such a degree that it is detectable via the FFPT?

In our experiment, we are interested in how a single video impacts the five factors tested and scored in the FFPT: agreeableness, extraversion, neuroticism, conscientiousness, and openness. Agreeableness describes the extent to which “a person puts others’ interests and needs ahead of their own.” Extraversion measures “how much a person is energized by the outside world.” Neuroticism captures an individual’s sensitivity “to stress and negative emotional triggers.”⁵ Conscientiousness scores an individual’s persistence, organization skills, and ambition. The final personality factor, openness, commonly referred to as “openness to experience,” describes creativity and curiosity.⁶

Hypotheses

- Null hypothesis (H_0): There is no difference in personality score for any of the five personality traits from the FFPT between those that watch a sad Sarah McLachlan animal cruelty video and those that watch a video on glass blowing.
- Alternative hypothesis (H_a): There is a difference in personality score for at least one of the five personality traits from the FFPT between those that watch a sad Sarah McLachlan animal cruelty video and those that watch a video on glass blowing.
- Significance threshold (α): 0.01

The FFPT is meant to be robust to temporary moods and measure one’s true personality traits. However, we feel that one’s responses might vary depending on their temporary mood at the time of the test.

To stay within the scope and budget of our research project, we decided a sad video on animal cruelty accompanied with sad Sarah McLachlan music would be the most effective at changing a participant’s mood right before taking the FFPT. We also decided that a video demonstrating glass blowing would have a minimal impact on one’s temporary mood. The sad video served as our treatment while the glass blowing video served as the placebo treatment.

We used a two-sided alternative hypothesis given that we were unsure the direction of the treatment effect. Because we would test the treatment’s impact five times (once for each personality trait), we used a Bonferroni correction to decrease the significance threshold to 0.01 to be conservative.

Before sending out our survey, we predicted that the treatment (the sad Sarah McLachlan video) would have the largest effect on subjects’ neuroticism scores. Neurotic people tend to have a high emotional reaction to stress and are more likely to experience mood swings. Thus, we expected those scores to change the most with the treatment. We also predicted that the treatment would have a minimal effect on subjects’ openness to experience scores. This personality trait seems the most unrelated to the sad video, and perhaps would hardly be swayed by this temporary mood change.

Experiment Details

Experiment Design & Treatment

In order to conduct our experiment, we utilized Qualtrics as our medium for survey delivery. Subjects were first asked some demographics questions, and subsequently were either assigned to treatment or control, which we describe in detail in the next section. We also asked subjects ahead of taking the FFPT if they had either taken or were aware of the FFPT. We hypothesized that respondents already familiar with the

⁴Querengässer, J., Schindler, S. Sad but true? - How induced emotional states differentially bias self-rated Big Five personality traits. BMC Psychol 2, 14 (2014). <https://doi.org/10.1186/2050-7283-2-14>

⁵The Big Five Personality Test. Retrieved from <https://www.truity.com/test/big-five-personality-test> 1 Dec 2021.

⁶Paul Costa and Robert McCrae and the Five-Factor Model of Personality. (2020, August 16). Lansing Community College. <https://socialsci.libretexts.org/@go/page/12242>

test would be less sensitive to our treatment video, which is why we collected this information. Additionally, this allowed us to filter out potentially problematic subjects who responded that they were not familiar with the FFPT but had actually taken it before.

Next, the treatment group was asked to watch the sad video advertisement video for the American Society for the Prevention of Cruelty to Animals (ASPCA) featuring Sarah McLachlan. We selected this video as our treatment exposure since it is cited as one of the most commercially successful nonprofit donation fundraising campaigns of all time.⁷ The control/placebo group was exposed to a neutral video on glass blowing. In this fashion both groups would be exposed to short videos ahead of taking the FFPT.

After watching their respective videos, participants were asked to answer a compliance question that could only be answered if they had indeed watched the entire video. These questions were made easy enough to answer so as not to be prohibitive, but hard enough to ensure one could not fake their way through it. Additionally, before moving past the videos, there was a timer which was implemented in Qualtrics to ensure that the participant spent time actually watching the video before moving on. In this way, non-compliers were able to be detected and removed from the dataset if they answered the post-video question incorrectly. We did not use the results from the compliance question as a covariate in our analysis, instead only used it for filtering purposes.

Finally, all respondents took the FFPT, consisting of 50 questions. Each question allowed participants to rank how strongly they agreed or disagreed with a statement. After finishing all of the questions, subjects were able to submit their answers.

Randomization Process

In order to split users into treatment and control groups randomly, we first asked them to provide us with their date of birth. If the subject was born on an even-numbered day of the month, they would be selected into the treatment group (i.e., shown the ASPCA video with Sarah McLachlan). If, on the other hand, the subject had an odd-numbered birthdate they were put into the control group (i.e., shown the neutral glass blowing video). We chose this method because we expected it to provide an unbiased and even split of our survey participants.

After carrying-out the randomization process, we found that there was a sufficiently random and equal split between the treatment and control groups.

Comparison of Potential Outcomes

The potential outcomes for this investigation were the treated and untreated personality factor scores for each of the five factors within the FFPT: extraversion, neuroticism, openness to experience, agreeableness, and conscientiousness. That is, each factor was measured for both those exposed to the sad video, and those exposed to the neutral video. In doing so, we were able to measure each of the potential outcomes for our project. Each of these factors is measured on a scale from 0 to 40 through the FFPT. Therefore, by analysing the changes to these factor values, we were able to compare potential outcomes.

Power Calculation

Before conducting our research, we ran a power analysis to determine how many survey participants we would need for our study. After taking the test ourselves, we conservatively concluded that a standard deviation of 15 and treatment effect of 3 was appropriate for each personality score. The power analysis concluded statistical significance if any one of the five factor personality tests showed a difference in score that produced a p-value less than our significance threshold of 0.01.

⁷Strom, Stephanie. "Ad Featuring Singer Proves Bonanza for the A.S.P.C.A." The New York Times, The New York Times, 26 Dec. 2008, <https://www.nytimes.com/2008/12/26/us/26charity.html>.

The graph in figure 1 shows how the power increases with sample size. After conducting the test, we calculated a power of .83 at 400 participants, and .89 at 500 participants. Given this information, we sought ways to find between 400 and 500 participants to complete our survey.

We used Lucid’s platform called Lucid Theorem^{8,9} to collect 504 survey responses total. We discarded 56 responses due to data discrepancies and 32 responses did not comply with the treatment. This left us with 416 survey responses that we feel answered questions correctly and complied with the treatment. More information about our data filtering and cleaning can be found in the Appendix. Given the adequate power of our study with 416 reliable survey responses, we have more confidence in the accuracy of our results.

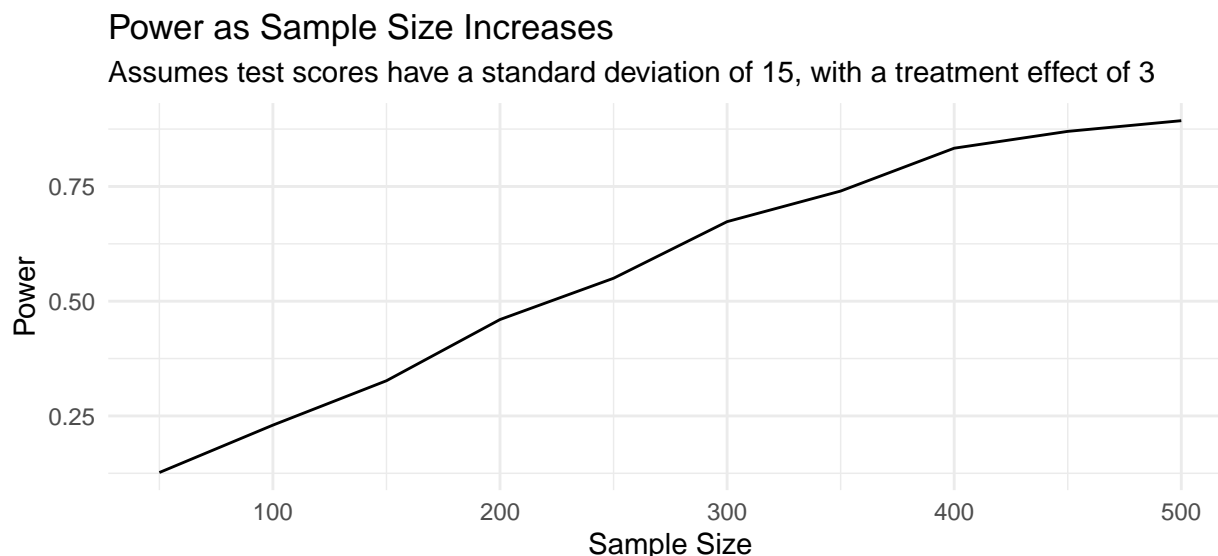


Figure 1: Power Analysis

Analysis

Data

For each of the 504 survey participants, we collected survey metadata, demographic data, and their responses to the 50 FFPT questions. Some of the covariates were explicitly asked in our Qualtrics survey while others were provided to us from the survey respondents’ profiles on Lucid. Not every covariate listed below was used in our analysis but was available if we needed. We’ve outlined the information collected in the survey below.

- Survey metadata:
 - Start and end date
 - IP Address
 - Survey progress out of 100%
 - Duration in seconds
 - Time taken for their first and last survey question clicks
- Survey questions:
 - Compliance “quiz” question to determine whether the participant complied with the treatment

⁸<https://luc.id/>

⁹<https://lucidtheorem.com/>

- Whether or not the participant was assigned treatment or control
- 50 FFPT survey questions (outcome being measured)
- Covariates:
 - Age grouped into 7 factor levels
 - Month of birthday
 - Gender
 - Highest level of education grouped into 8 factor levels
 - Whether or not the participant has taken the FFPT before
 - Whether or not the participant is familiar with the FFPT
 - Household income grouped into 24 factor levels
 - Ethnicity grouped into 4 factor levels
 - Hispanic
 - Political party
 - Region: west, midwest, south, and northeast
 - Granular location info: latitude, longitude, and zip code

CONSORT Document

We designed the experiment to have two parts: a video followed by a survey. Participants with odd birthday dates were assigned the Sarah McLachlan video (n=242), and participants with even birthday dates were assigned the neutral glass blowing video (n=262). After watching the video, the participants were assigned to complete The Five Factor Personality Test.

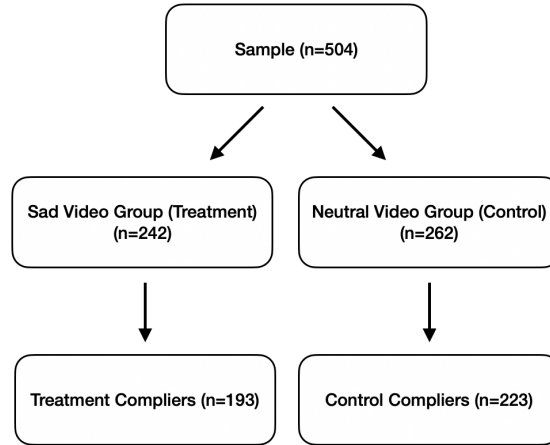


Figure 2: Consort Diagram

Scoring methodology

When filling out the survey, participants were asked whether they disagree, slightly disagree, feel neutral, slightly agree, or agree with 50 questions about their personality. These answers were converted to a numeric scale ranging from 1 to 5. Each of these numeric scores either positively or negatively impact the final score towards a personality trait. In the end, each of these scores were added up and rescaled to have a minimum possible score of 0 and maximum score of 40. These scores were then compared between treatment and control groups to determine significance.

Data Cleaning and Filtering

The original dataset included 504 observations with 94 variables. These variables consisted of the responses to the 50 FFPT survey questions, demographic data provided by Lucid, demographic data provided by Qualtrics, and some additional technical metadata from Qualtrics such as timestamps.

To begin, we renamed the columns to have more user-friendly names. We then went through every column and checked the data for reasonableness and searched for situations where we couldn’t trust the responses.

First, we found several participants that answered “Neutral” for every FFPT question—these users were not used in the final analysis. Users with the same answer across 45 of the 50 FFPT questions were filtered from our final dataset.

Second, we leveraged the demographic data that we collected in addition to the Lucid-provided data to check for consistency. For example, we asked respondents for their gender in our Qualtrics survey, but Lucid also provided this same data. After comparing the two, we found that for a few responses the two gender answers were different. We then performed a similar analysis for age. To ensure we had accurate data, we dropped respondents whose demographic data was not consistent across the Lucid-provided data and the data we collected.

Third, while the vast majority of survey respondents finished the survey in less than 15 minutes, a few took far longer to respond, causing a skewed distribution in completion time. We dropped those who took more than 1,000 seconds (16.67 minutes) to respond, which caused the distribution to be much more normal. We chose 1,000 seconds as a threshold because Lucid Theorem required our survey to have a maximum completion time of 15 minutes. Rounding up to 1,000 seconds allowed for some extra buffer time just in case users needed some additional time to complete the survey.

Finally, we collected data on respondents’ familiarity with the FFPT and whether or not they had taken the test before. Logically, it would not make sense for an individual to be unfamiliar with FFPT but respond as having taken the test before. We dropped these individuals from our analysis as well.

For additional data cleaning, we simplified several of the demographic covariates provided by Lucid. For example, the ethnicity and Hispanic covariates each contained 16 sub-categories. Lucid Theorem provided a variable mapping, which allowed us to simplify these few complex factor levels. We simplified ethnicity into four sub-categories (white, Black, Asian, other) and Hispanic into a binary variable.

Data Details and Covariate Checks

After discarding the 56 responses due to data discrepancies, we were left with 448 subjects in our dataset. Thirty two of these subjects were non-compliers. The table below shows the breakdown of treatment vs control for both the compliers and non-compliers.

Table 1: Compliers in treatment and control before filtering

	Control	Treatment
Complier	223	193
Noncomplier	10	22
Total	233 (52.01%)	215 (47.99%)

Next, we conducted covariate checks to check the distribution of covariates between our treatment and control groups.

Table 2 shows the output of the treatment assignment regressed on the covariates used in our models. Covariates showing a significant impact on whether the subject received treatment or control could indicate a potential covariate imbalance.

After running the test, we see that those in age group 18-24 years old show a significant impact at the .05 significance level on whether one receives treatment or control. To investigate further, we created a bar chart in figure 3 to show the breakdown of treatment vs control at every age group. Only 2 subjects were in the age group under 18—due to the small number of participants in this group, we can ignore the imbalance for these two subjects. However, only 31% received treatment of those in age group 18-24. Because this age group had 64 subjects, this imbalance might be concerning.

Based on the regression results and barchart, there does appear to be a slight imbalance of treatment vs control for those in age group 18-24. When we randomized subjects into treatment or control, we did not do so with any dependence on their age. Thus, we believe this imbalance is due to random chance.

Table 2: Covariate Balance Checks

	<i>Dependent variable:</i>
	Treatment
18-24 years old	−0.786** (0.362)
25-34 years old	−0.522 (0.360)
35-44 years old	−0.592 (0.362)
45-54 years old	−0.640* (0.363)
55-64 years old	−0.666* (0.365)
65+ years old	−0.626* (0.368)
Male	−0.086* (0.046)
High school diploma or GED	0.063 (0.099)
Some college but no degree	−0.048 (0.104)
Associate degree in college (2-year)	0.046 (0.114)
Bachelor’s degree in college (4-year)	−0.048 (0.111)
Master’s degree	−0.093 (0.146)
Professional degree (JD, MD)	0.227 (0.215)
Doctoral degree	0.064 (0.226)
Region - Northeast	−0.022 (0.073)
Region - South	−0.112* (0.064)
Region - West	−0.001 (0.072)
Familiar with test before	0.024 (0.062)
Taken test before	0.002 (0.079)
Asian	−0.143 (0.141)
Black or African American	0.049 (0.100)
White	0.087 (0.081)
Hispanic	0.042 (0.082)
Constant	1.108*** (0.368)
Observations	504
R ²	0.066
Adjusted R ²	0.022
<i>Note:</i>	
*p<0.1; **p<0.05; ***p<0.01	

Models

Using the cleaned and filtered data, we created models to estimate the average treatment effect of the Sarah McLachlan ASPCA video on the five personality types. Our models estimated the intent to treat effect (ITT) and the average treatment effect on the treated (ATET) since we administered a placebo video for those who in our control group. We drew conclusions primarily from the ATET but additionally show as an example that without the placebo, our estimates would have been biased. When creating models, we

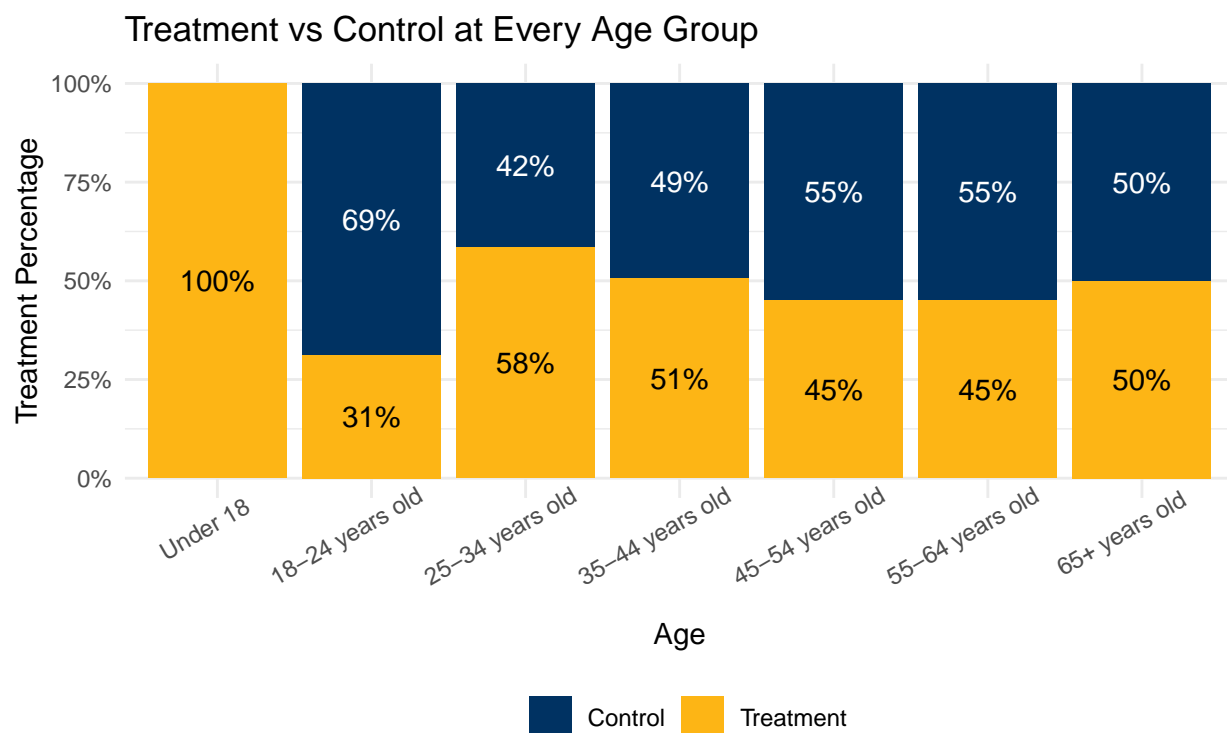


Figure 3: Treatment Assignment by Age

started with simple models, regressing only the personality outcomes on the treatment, and then modeled the outcomes additionally using covariate data we collected.

Simple Models

In our simple models, we regressed each personality type on a binary variable indicating if someone was assigned to the treatment group. We estimated the ITT for all subjects in the experiment and the ATET for compliers. Additionally, we applied robust standard errors in all of our models to account for any heteroskedasticity in the data.

The ATETs in these models indicate that extraversion and neuroticism were significantly impacted by the Sarah McLachlan video. As shown in table 4, the treatment video reduced the amount someone is extraverted by -2.14 on average with a standard error of 0.78. Similarly, the treatment video reduced an individual's level of neuroticism by -1.97 on average with a standard error of 0.69. The treatment variable on these two models have p-values less than 0.01, which we used as our Bonferroni-corrected threshold for significance. Agreeableness is moderately significant, but not enough to meet this threshold.

The simple models demonstrate the benefit of using a placebo design and having the ability to calculate the ATET instead of only the ITT. If we only had data to calculate the ITT, we would have incorrectly stated the significance level of agreeableness and openness. The ITT values for these personality types indicate that the treatment video did have a significant impact for agreeableness and a moderately significant impact for openness. The table showing our results with ITT is included in the Appendix.

Table 3: Simple Model ATET Results for Compliers

	<i>Dependent variable:</i>				
	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
	(1)	(2)	(3)	(4)	(5)
Treatment	-1.087* (0.521)	0.181 (0.653)	-2.141** (0.781)	-1.973** (0.694)	-1.097 (0.617)
Constant	21.305*** (0.339)	25.498*** (0.460)	20.493*** (0.508)	26.740*** (0.471)	26.677*** (0.389)
Covariates?	No	No	No	No	No
Treatment p value	0.0365	0.783	0.00607	0.00468	0.0734
Observations	416	416	416	416	416
R ²	0.011	0.0002	0.018	0.019	0.008
Adjusted R ²	0.008	-0.002	0.016	0.017	0.005

Note:

*p<0.05; **p<0.01; ***p<0.001

Models with Covariates

We then created models using the additional covariates that may influence personality type. The covariates we included in our model are age, gender, highest education, region, ethnicity, Hispanic, “familiar with test,” and “taken test before.”

Table 4: Model ATET Results for Compliers Using Covariates

	<i>Dependent variable:</i>				
	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
	(1)	(2)	(3)	(4)	(5)
Treatment	-1.374** (0.506)	-0.062 (0.659)	-2.084** (0.799)	-2.264*** (0.671)	-1.028 (0.614)
Constant	18.962*** (1.790)	25.566*** (2.085)	20.679*** (2.578)	29.065*** (1.942)	24.963*** (1.715)
Covariates?	Yes	Yes	Yes	Yes	Yes
Treatment p value	0.00818	0.926	0.00904	0.000918	0.0975
Observations	416	416	416	416	416
R ²	0.129	0.092	0.091	0.172	0.099
Adjusted R ²	0.078	0.039	0.037	0.124	0.046

Note:

*p<0.05; **p<0.01; ***p<0.001

In general, the precision in our estimates increased when we included covariates in our models. The new significance levels and covariate estimates are included in table 4. The significance of neuroticism and agreeableness both increased. Without covariates, neuroticism was only significant at the 1% level. Whereas after including covariates, neuroticism became significant at the 0.1% level. Similarly, agreeableness changed from being marginally significant to significant after including covariates with its p-value decreasing below our 1% threshold. The significance level of extraversion was not impacted, and conscientiousness and openness

remained insignificant.

Similar to our simple models, we observed that using the ATET was advantageous compared to the ITT. The ITT models with covariates indicated marginal significance (p-value less than 0.05) for openness and extraversion, significance (p-value less than 0.01) for agreeableness, and high significance for neuroticism (p-value less than 0.001). Both openness and extraversion had incorrect significance results using the ITT, further demonstrating the value in our placebo design. These full ITT results are included in the Appendix.

Model Results Discussion

Based on all the models we created, we drew conclusions from our models that estimate the ATET using covariates. These models are more precise than the simple models and are also unbiased, unlike our ITT models. We conclude that the Sarah McLachlan ASPCA video had a significant impact on three of the five personality types: agreeableness, extraversion, and neuroticism. On average, the treatment video decreased compliers' scores for agreeableness, extraversion, and neuroticism by 1.37 (0.51), 2.08 (0.8), and 2.26 (0.67) points, respectively, with each ATET's standard error included in parentheses.

Table 5 demonstrates the effectiveness of deploying our placebo design and utilizing covariates in our analysis. The table shows our full results for openness. The models that do not filter out noncompliers consistently estimate the ITT as -1.2, which is marginally significant (p-value less than 0.05). However, this estimate is biased. When we do filter out noncompliers and estimate the ATET, we calculate the treatment effect to be -1.0, which is not significant. Additionally, adding covariates successfully decreased the standard error for our treatment effect estimates for both the models estimating ITT and the models estimating ATET.

Table 5: Model Comparison for Openness

	<i>Dependent variable:</i>			
	Openness			
	ITT	ATET	ITT	ATET
	(1)	(2)	(3)	(4)
Treatment	-1.246*	-1.097	-1.200*	-1.028
	(0.590)	(0.617)	(0.584)	(0.614)
Constant	26.725***	26.677***	25.271***	24.963***
	(0.382)	(0.389)	(1.659)	(1.715)
Covariates?	No	No	Yes	Yes
Treatment p value	0.0343	0.0734	0.0429	0.0975
Observations	448	416	448	416
R ²	0.010	0.008	0.106	0.099
Adjusted R ²	0.008	0.005	0.058	0.046

Note:

*p<0.05; **p<0.01; ***p<0.001

Conclusions

In this experiment, we tested whether and how much mood impacting external stimuli can impact an individual's results on a personality test. By running regressions on 416 responses from our Qualtrics survey, we concluded that our treatment video, Sarah McLachlan's 2007 ASPCA commercial, had a significant impact on three factors of the Five Factor Personality Test: agreeableness, extraversion, and neuroticism. Each of these factors had a significantly lower value in our treatment group than control group.

However, our study had several limitations preventing us from generalizing to all populations and across longer time periods. As discussed throughout this paper, covariate checks, while approximately balanced between treatment and control, led us to believe that our sample demographics may be different compared to the large adult U.S. population. This was most notable in the distribution of household income among our survey respondents. For this reason, we believe the results of this study may not be generalizable to the larger U.S. population. Additionally, our valid respondents completed the survey in under 1,000 seconds (16.67 minutes), so while we concluded there was a significant effect on personality in this time period, we are uncertain how long this treatment effect lasts. Finally, we performed this experience using only one treatment video. It is likely that different videos will have different effects on personality, so we cannot generalize our specific ATET results to all sad videos. Future experiments should focus on reproducibility and generalizability, which would address these key limitations and lend additional confidence to our conclusions.

Despite these limitations, these results cause us to view personality tests like the FFPT more skeptically. The FFPT in particular is supposed to lack subjectiveness that plagues other personality tests. However, our results indicated that an individual's short-term mood will cause them to score differently on the FFPT. When evaluating personality, it will be important to keep in mind that FFPT results may not tell the full story about an individual's personality.

Appendix

ITT Model Results

Table 6: Simple Model ITT Results for All Participants

	<i>Dependent variable:</i>				
	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
	(1)	(2)	(3)	(4)	(5)
Treatment	-1.338** (0.506)	-0.159 (0.619)	-1.904** (0.731)	-1.936** (0.659)	-1.246* (0.590)
Constant	21.129*** (0.335)	25.545*** (0.449)	20.485*** (0.492)	26.704*** (0.458)	26.725*** (0.382)
Covariates?	No	No	No	No	No
Treatment p value	0.00829	0.798	0.0094	0.00346	0.0343
Observations	448	448	448	448	448
R ²	0.016	0.0001	0.015	0.019	0.010
Adjusted R ²	0.013	-0.002	0.013	0.017	0.008

Note:

*p<0.05; **p<0.01; ***p<0.001

Table 7: Model ITT Results for All Participants Using Covariates

	<i>Dependent variable:</i>				
	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
	(1)	(2)	(3)	(4)	(5)
Treatment	-1.588** (0.489)	-0.315 (0.616)	-1.838* (0.747)	-2.165*** (0.637)	-1.200* (0.584)
Constant	19.125*** (1.765)	25.989*** (1.982)	20.762*** (2.463)	28.745*** (1.874)	25.271*** (1.659)
Covariates?	Yes	Yes	Yes	Yes	Yes
Treatment p value	0.00156	0.615	0.0141	0.000828	0.0429
Observations	448	448	448	448	448
R ²	0.143	0.094	0.085	0.168	0.106
Adjusted R ²	0.097	0.045	0.035	0.123	0.058

Note:

*p<0.05; **p<0.01; ***p<0.001

Model Comparisons for All Personality Types

Sample Results with All Covariates

Table 8: Model Comparison for Agreeableness

	<i>Dependent variable:</i>			
	Agreeableness			
	ITT	ATET	ITT	ATET
	(1)	(2)	(3)	(4)
Treatment	-1.338** (0.506)	-1.087* (0.521)	-1.588** (0.489)	-1.374** (0.506)
Constant	21.129*** (0.335)	21.305*** (0.339)	19.125*** (1.765)	18.962*** (1.790)
Covariates?	No	No	Yes	Yes
Treatment p value	0.00829	0.0365	0.00156	0.00818
Observations	448	416	448	416
R ²	0.016	0.011	0.143	0.129
Adjusted R ²	0.013	0.008	0.097	0.078
<i>Note:</i>		*p<0.05; **p<0.01; ***p<0.001		

Table 9: Model Comparison for Conscientiousness

	<i>Dependent variable:</i>			
	Conscientiousness			
	ITT	ATET	ITT	ATET
	(1)	(2)	(3)	(4)
Treatment	-0.159 (0.619)	0.181 (0.653)	-0.315 (0.616)	-0.062 (0.659)
Constant	25.545*** (0.449)	25.498*** (0.460)	25.989*** (1.982)	25.566*** (2.085)
Covariates?	No	No	Yes	Yes
Treatment p value	0.798	0.783	0.615	0.926
Observations	448	416	448	416
R ²	0.0001	0.0002	0.094	0.092
Adjusted R ²	-0.002	-0.002	0.045	0.039
<i>Note:</i>		*p<0.05; **p<0.01; ***p<0.001		

Table 10: Model Comparison for Extraversion

	<i>Dependent variable:</i>			
	Extraversion			
	ITT	ATET	ITT	ATET
	(1)	(2)	(3)	(4)
Treatment	-1.904** (0.731)	-2.141** (0.781)	-1.838* (0.747)	-2.084** (0.799)
Constant	20.485*** (0.492)	20.493*** (0.508)	20.762*** (2.463)	20.679*** (2.578)
Covariates?	No	No	Yes	Yes
Treatment p value	0.0094	0.00607	0.0141	0.00904
Observations	448	416	448	416
R ²	0.015	0.018	0.085	0.091
Adjusted R ²	0.013	0.016	0.035	0.037
<i>Note:</i>		*p<0.05; **p<0.01; ***p<0.001		

Table 11: Model Comparison for Neuroticism

	<i>Dependent variable:</i>			
	Neuroticism			
	ITT	ATET	ITT	ATET
	(1)	(2)	(3)	(4)
Treatment	-1.936** (0.659)	-1.973** (0.694)	-2.165*** (0.637)	-2.264*** (0.671)
Constant	26.704*** (0.458)	26.740*** (0.471)	28.745*** (1.874)	29.065*** (1.942)
Covariates?	No	No	Yes	Yes
Treatment p value	0.00346	0.00468	0.000828	0.000918
Observations	448	416	448	416
R ²	0.019	0.019	0.168	0.172
Adjusted R ²	0.017	0.017	0.123	0.124
<i>Note:</i>		*p<0.05; **p<0.01; ***p<0.001		

Table 12: Sample Agreeableness ATET Model Showing All Covariates

	<i>Dependent variable:</i>
	Agreeableness
Treatment	-1.374** (0.506)
Constant	0.677 (0.957)
how_old_are_you35-44 years old	2.765** (0.942)
how_old_are_you45-54 years old	1.779 (0.970)
how_old_are_you55-64 years old	3.745*** (0.958)
how_old_are_you65+ years old	4.276*** (0.996)
genderMale	-1.391** (0.534)
education_highestBachelor's degree in college (4-year)	0.132 (1.004)
education_highestDoctoral degree	-2.505 (1.935)
education_highestHigh school diploma or GED	-0.709 (0.910)
education_highestLess than high school degree	-1.360 (1.252)
education_highestMaster's degree	0.961 (1.420)
education_highestProfessional degree (JD, MD)	-0.951 (1.445)
education_highestSome college but no degree	-0.876 (0.953)
regionNortheast	-0.321 (0.820)
regionSouth	0.287 (0.802)
regionWest	-0.459 (0.882)
familiar_with_testYes	0.578 (0.730)
taken_test_beforeYes	-1.571 (1.075)
ethnicityBlack or African American	2.012 (1.639)
ethnicityOther	0.746 (1.599)
ethnicityWhite	1.389 (1.407)
hispanic	0.804 (1.126)
Constant	18.962*** (1.790)
Covariates?	Yes
Treatment p value	0.00818
Observations	416
R ²	0.129
Adjusted R ²	0.078

Note:

*p<0.05; **p<0.01; ***p<0.001