Building Mistrust from Inconsistent Verbal and Non-Verbal Signals

W241.2 - Field Experiments Project
Kanitha Mann, Gaurav Khanna, Francis Shannon & Sudha Subramanian

RESEARCH MOTIVATION

In-person communication amongst the general population is a combination of what is spoken and what is conveyed through body language and facial expression. These non-verbal signals tend to be consistent with verbal signals, giving the audience an accurate sense of what the speaker means. However, when verbal messaging is inconsistent with the speaker's non-verbal signals, the audience may misunderstand or develop a mistrust of the speaker. Such a scenario appears often in work and social settings to varying outcomes. Some examples include:

- Someone who may be speaking but is distracted by their environment (for example, looking at items in their surrounding other than the audience) can make the audience feel like the weight of the words isn't as important as it would have been would the speaker make eye contact.
- A manager trying to say they're not upset with you for a mistake, but their voice is shaky and exasperated.
- A salesperson who is pitching a good product to an important prospect, but is nervous and soft spoken, therefore not being convincing enough to make the sale

Each of these examples leads to the question: do speakers who present inconsistent verbal and non-verbal signals lead their audience to mistrust them? Conducting an experiment pertaining to this question can grant us insight on what signals people rely on to decipher the "real" message the speaker meant. Furthermore, this exercise can help educate people on how to better conduct themselves in work and social situations so that they can communicate more effectively. Improved communication would lead to improved relationships by building trust and being transparent.

1

PROBLEM STATEMENT

1. Null Hypothesis: There is no difference in the sample population's trust of speakers whose verbal and non-verbal cues align, compared to speakers whose verbal and non-verbal cues are at odds with one another.

    a. It may be that there are other cues that would make a subject mistrust. This could be gender, age, race, or how well the subject already understands what the speaker is talking about.
2. Alternative Hypothesis: The sample population is more likely to distrust speakers whose verbal and non-verbal cues do not align with one another, compared to when those cues do align.

EXPERIMENTAL DESIGN

The experiment was conducted through Qualtrics, an online survey vendor. Subjects were recruited to take the survey through three different sources: Friends and Family, Amazon Mechanical Turk (MTurk), and Lucid Surveyors (Lucid). While the Friends and Family group were contributing to the experiment with no expected monetary compensation, MTurk workers were awarded $0.20 and Lucid workers were awarded $1 for a good effort completion of the survey. The recruitment of Friends and Family subjects began on December 1, 2018, with MTurk recruitment beginning on December 2, 2018, and Lucid recruitment beginning on December 6, 2018. All responses were collected on December 8, 2018.

Upon entering the survey, subjects were instructed to watch one video that was up to 3 minutes long. The videos were randomly assigned as follows:

1. A control video delivered by a male speaker (Male Control)
2. A control video delivered by a female speaker (Female Control)
3. A treatment video delivered by a male speaker (Male Treatment)
4. A treatment video delivered by a female speaker (Female Treatment)

After watching the video, the subject would be asked a compliance question, based on the content of the videos, that checked for their understanding. If subjects answered this question incorrectly, the survey would end and they would not be part of the experiment. Subjects who answered the compliance question correctly would then move

on to the next question, which was to mark if they trusted or did not trust the speaker they just watched.

Of the 785 randomized surveys received, only 33% were considered a "quality" survey for further analysis, as those subjects answered the compliance question correctly. A total of 262 surveys were used for the analysis. Figure 1 displays a flowchart of how subjects were grouped between sources and treatments.
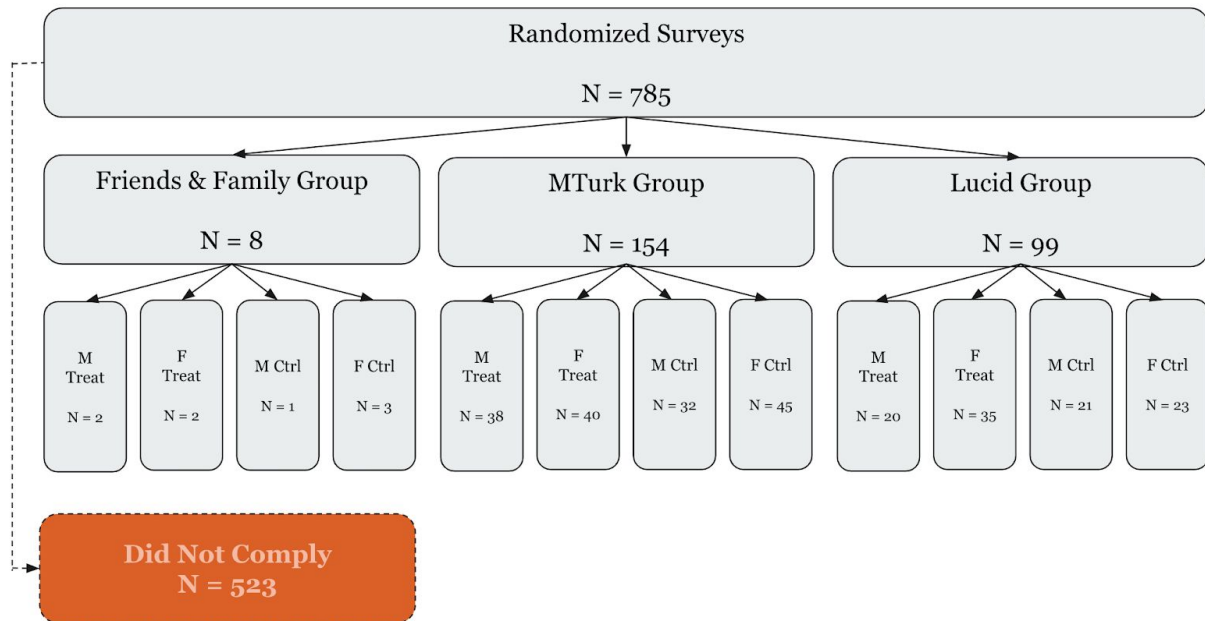

Figure 1. Flowchart of surveys divided into groups

VIDEOS

As the experiment sought to measure opinions on trust based on audio and body language, videos were used to deliver these varying effects. We originally sought to source a video from the internet, but either could not find good examples, or did not feel comfortable sharing videos that may have brought ridicule to the unsuspecting speaker. Instead, two of our researchers, one male and one female, agreed to record videos of themselves speaking from the same script about obesity. Furthermore, these researchers recorded two versions of their script. The control video displayed our researchers speaking with confidence. The treatment video displayed our researchers acting flippant about the topic at hand. During the middle of the videos, both speakers mentioned three named diets. This content was used to formulate the compliance question. Appendices B and C display links to the videos recorded for the experiment, as well as the script used.

SURVEY STRUCTURE

Figure 2 displays a flowchart of how the survey progressed for subjects, depending on their actions.
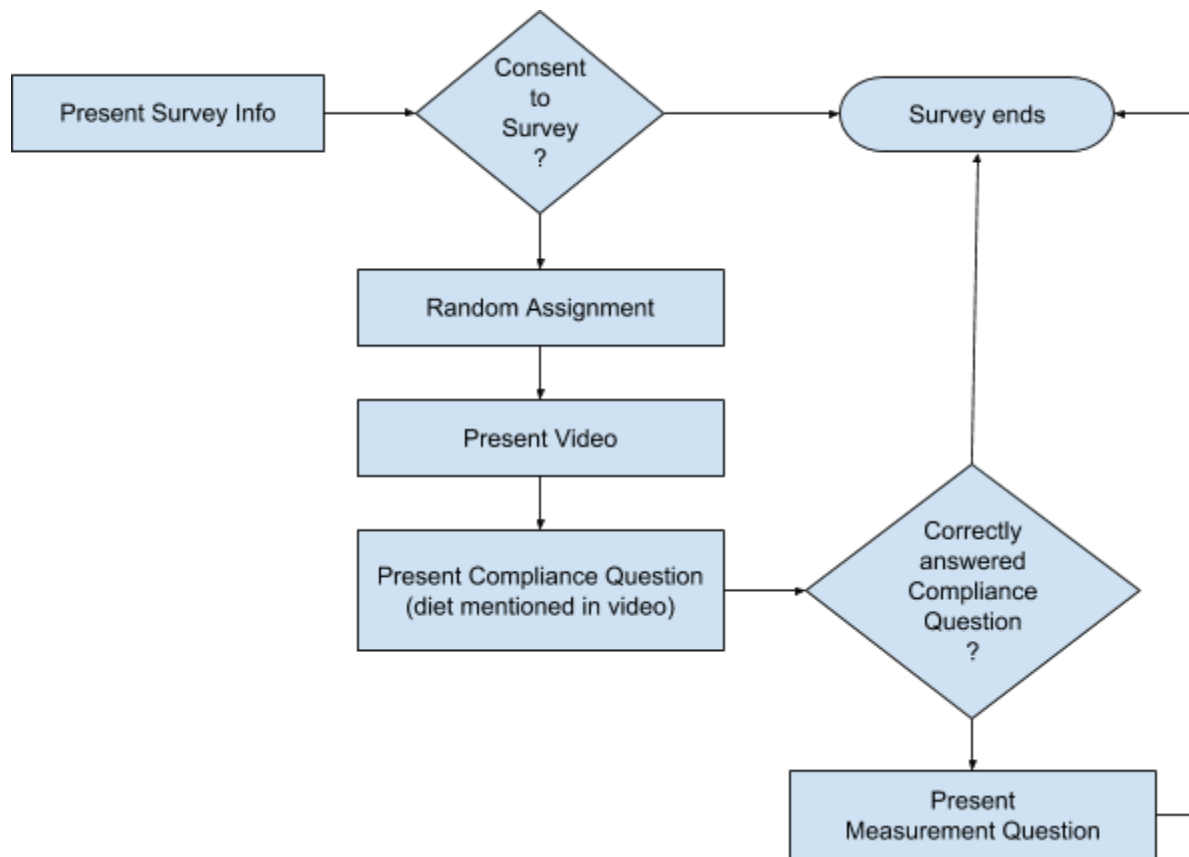

Figure 2. Flowchart of survey structure. Subjects who do not answer the compliance question correctly get directed up to Survey ends.

All recruited subjects were directed to take a Qualtrics survey, with an estimated completion time of 5 minutes. The survey was laid out as follows:

1. Subjects read an explanation of the survey, what to expect, and to consent by continuing on
2. Subjects were asked to watch one of the randomized videos presented to them in full
3. Subjects were informed that they are now to answer two questions
4. Subjects were asked the compliance question: *What is one of the diet plans mentioned in the video?* The subject could pick from one of three options: *Ketogenic, Paleo,* and *Pritikin.*

5. If the subjects picked *Ketogenic* or *Paleo,* they will receive a message that they did not comply to the consent set forth at the beginning of the survey (*The quality of your participation will be verified.*) and we would not ask them to complete the second question. Their participation ends at this point.

6. If the subjects picked *Pritikin*, they are asked our main question: *How would you rate the speaker using the options below?* The subject could pick from one of two options: *I trust what the speaker was saying* or *I do not trust what the speaker was saying.*

7. After submitting the answer, subjects either received an MTurk code or a Lucid code, depending on which source they came from, that they can use to receive payment.

Refer to Appendix D for screenshots of the survey.

DATA COLLECTION

Qualtrics collected and stored every subject's attempt at the survey. The parameters recorded include:

- Time metrics: Start Date/Time, End Date/Time, and Duration in seconds
- IP Address
- Latitude/Longitude Coordinates
- User Language
- Answer to the compliance question
- Answer to the measurement question
- Worker code generated if subject answered measurement question
- Which video the subject saw (FTreatment, FControl, MTreatment, MControl)

The results were downloaded into CSV format and subsequently imported into R for analysis.

PILOT

A pilot was performed from November 21, 2018 to November 22, 2018 in order to test a number of unknown variables related to the experiment, including the power to be obtained for the actual experiment, the mechanics of Qualtrics and crowdsourcing platforms like MTurk, and the quality of the crowdsource workers. The intention was to obtain 12 survey responses. The respondents were instructed to watch a video of one of our researchers speaking about homelessness for approximately 90 seconds. The treatment video was of the researcher displaying normal cues while the control video was of the researcher displaying misaligned cues. The treatment was randomized by the

survey. At the end of the video, the respondent was asked one compliance question (*What is the number of homeless people referenced for the LA metro area? Use commas.*) and one measurement question about trust.

The MTurk requirement of 12 responses was achieved in less than 12 hours, so the researchers closed the pilot at this time. Qualtrics actually recorded 76 responses. The inflated number was due to MTurk workers submitting multiple responses, usually because they got the compliance question incorrect and were making guesses by opening their survey in a different IP address, or through bot activity. Of the 12 responses that MTurk officially collected, 2 were rejected outright (one for a fake code, the other because she/he spent 13 seconds on the entire survey, which is unrealistic given that the video was 90 seconds long). 1 more MTurk was removed because she/he showed multiple attempts, indicated by the exact same latitude and longitude coordinates for these multiple attempts.

Of the 9 remaining MTurk, 7 watched the treatment video while 2 watched the control. 6 respondents in the treatment group trusted the speaker, and 1 respondent in the control group trusted the speaker.

Overall, the pilot was conducted so we could understand effect size & statistical power, the recruitment of subjects and timeliness of responses, additional categorization needed for the real experiment (such as the compliance question), and the covariates needed for the experiment (the speaker's gender).

See Appendix E for screenshots of the survey.

EXPERIMENT - DATA ANALYSIS /Recruited vs Useful (compliance verified)

The first part of our data analysis involved separating the "useful" surveys from the overall "recruited" surveys. Figure 3 displays the share of useful surveys per source, while Figure 4 displays the share overall.

Subjects (Recruited vs Useful)
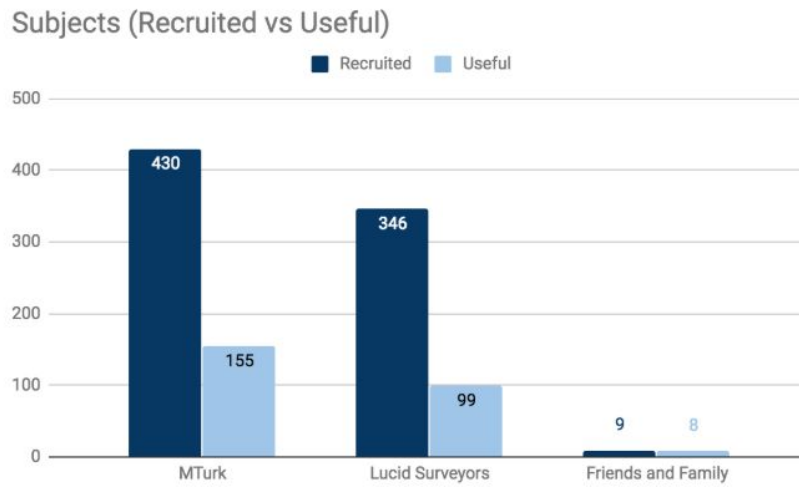
MTurk averaged 36% compliance
Lucid averaged 28% compliance

Figure 3. Useful vs. Recruited surveys per source

33%

Quality Surveys

(for analysis)

Figure 4. Useful vs. Recruited surveys overall

We wanted to verify that there were an equal amount of useful surveys for the treatment vs. control groups, as well as the gender covariate. Analysis of the survey data shows a pretty even distribution between treatment and control groups, shown below in Figure 5.
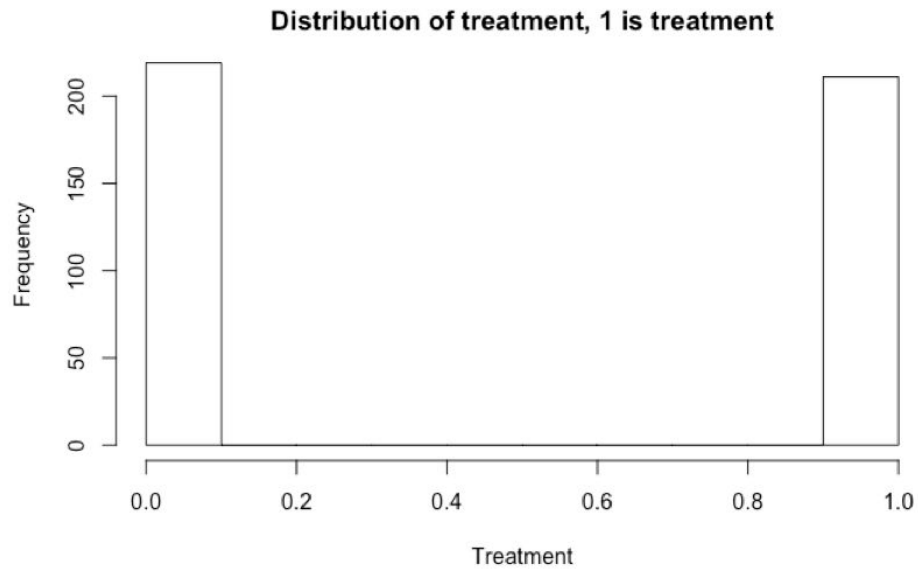

Figure 5. Distribution of treatment and control videos

Likewise, the gender distribution among survey respondents were also pretty evenly distributed, as shown below in Figure 6.

8
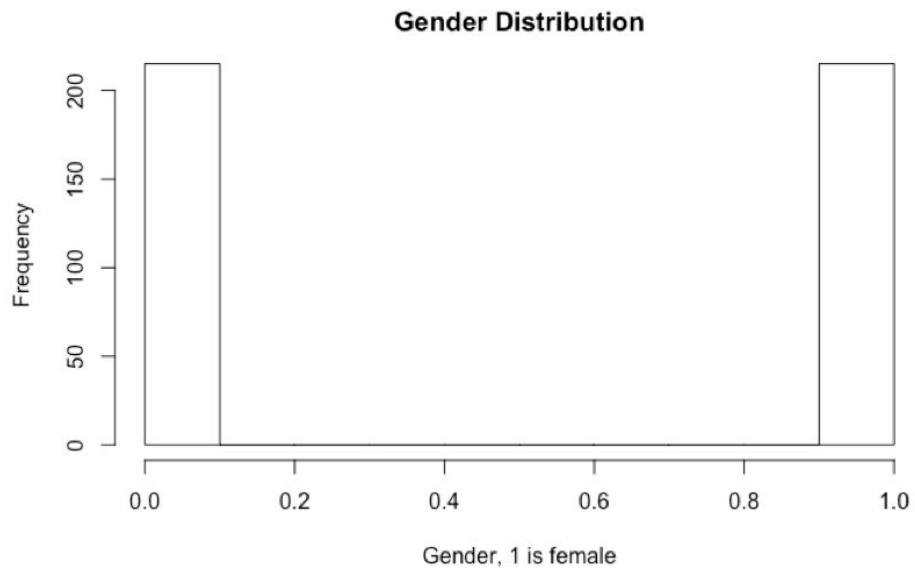
Figure 6. Distribution of male vs female speakers in the videos

Figure 7 displays a plot of the outcome variable from MTurk:
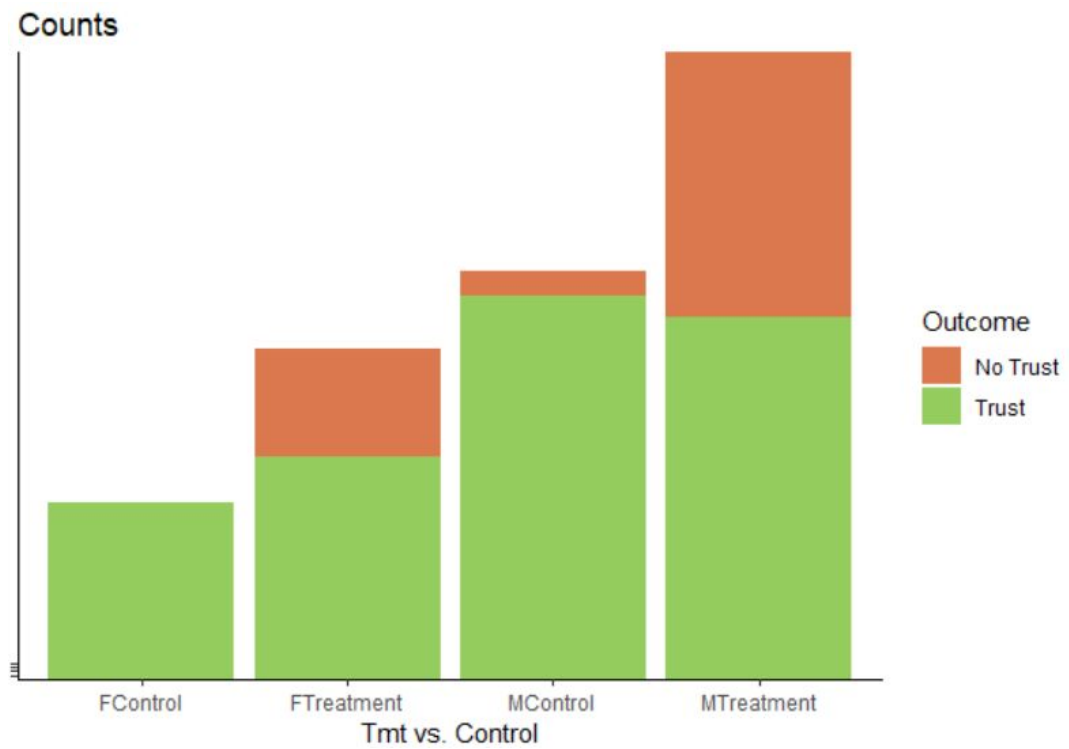
Figure 7. Number of surveys by video type and their outcomes for MTurk

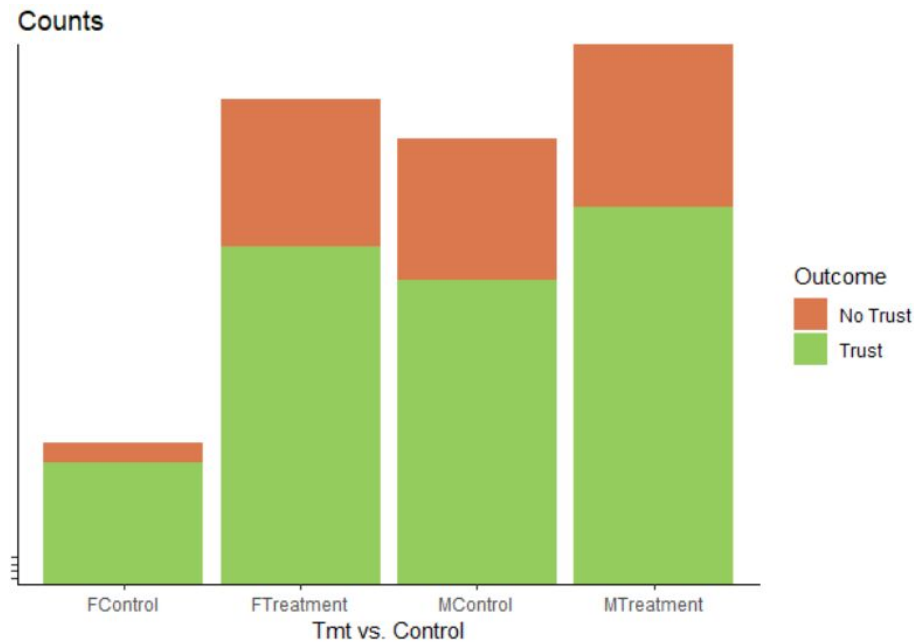Figure 8 displays a plot of the outcome variable from Lucid:

Figure 8. Number of surveys by video type and their outcomes for Lucid

Figures 7 and 8 show that subjects in female control groups performed similarly between MTurk and Lucid. Male control groups also favored the "Trust" answer more, although MTurk had a higher percentage of "Trust" answers than from Lucid.

Among subjects in the treatment (flippant video version) groups, there were more "Trust" responses than "No Trust" among both MTurk and Lucid surveyors, but featured more "No Trust" responses than from the control versions. There still exist differences between MTurk and Lucid for the treatment videos, however. A higher percentage of MTurk workers distrusted the male speaker treatment video than did the Lucid workers. These differences are reflected in the treatment effect that are displayed in Table 1 in the Results section.

Among the friends & family group, the response looks more along the lines expected in female treatment, female control and male control groups, as captured in Figure 9 below.
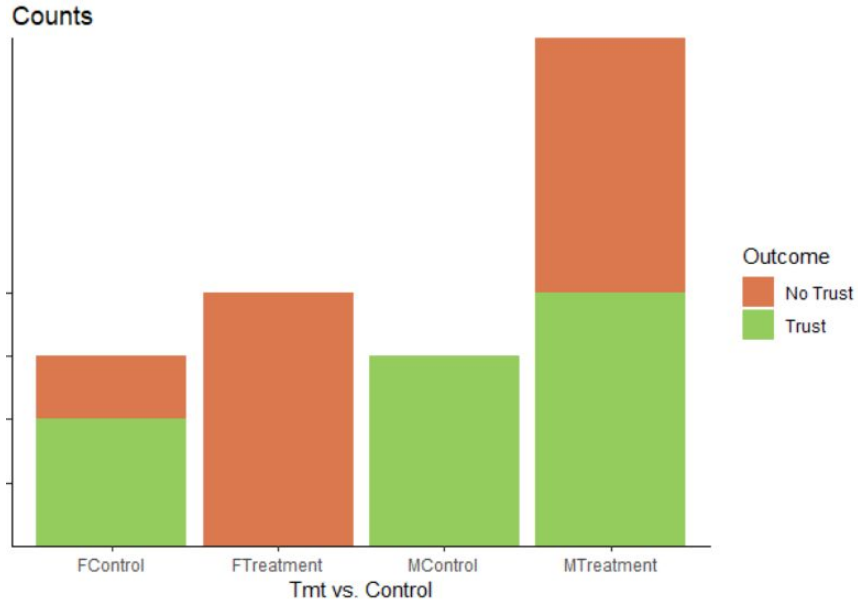
Figure 9. Number of surveys by video type and their outcomes for Friends and Family

RESULTS: Treatment Effect

As part of our analysis, we applied regression to compare treatment vs. control, and also taking the covariate (indicator variable for gender) into account as well. Table 1 describes the experimental results across all 3 sources as well as pooled as it pertains to the following model:

$$Y_i = \alpha + \beta T_i + \gamma F_i + \varepsilon_i$$

Where Y_i is the outcome variable "trust", T_i is a variable indicating the treatment or control video, F_i is a dummy variable indicating if the speaker was female or male, and ε_i is the error term. The subscript i indicates a per subject identifier.

11

Table 1: Experimental Results

	FriendsFamily	Lucid	MTurk	AllSources
	Dependent variable:			
	trust			
	(1)	(2)	(3)	(4)
treat	−0.500	0.018	−0.360***	−0.223***
	(0.661)	(0.139)	(0.087)	(0.076)
female	−0.333	0.175	0.061	0.101
	(0.624)	(0.137)	(0.084)	(0.074)
treat:female	−0.167	−0.181	0.035	−0.051
	(0.825)	(0.186)	(0.118)	(0.102)
Constant	1.000	0.682***	0.939***	0.839***
	(0.540)	(0.096)	(0.063)	(0.055)
Observations	8	99	154	261
R^2	0.417	0.023	0.196	0.094
Adjusted R^2	−0.021	−0.008	0.180	0.083
Residual Std. Error	0.540 (df = 4)	0.449 (df = 95)	0.364 (df = 150)	0.408 (df = 257)
F Statistic	0.952 (df = 3; 4)	0.748 (df = 3; 95)	12.157*** (df = 3; 150)	8.882*** (df = 3; 257)

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

RESULTS: Test for Sharp-Null Hypothesis

A simulated run of randomized experiments was conducted to test the ATE under the sharp-null hypothesis. This was done in 2 ways:

1. By performing randomized allocation of the subjects in treatment vs. control and plotting the results for 100,000 runs.
2. By performing randomized allocation of the subjects in treatment vs. control within the subgroups (Male and Female) and plotting the results for 50,000 runs

The results from the first run are shown in Figure 10. The ATE value of -0.223 (SE: 0.076), denoted as a blue line, lies very close to zero density (the ATE appeared in 3 out of 100,000 randomizations). We conclude that the results of the experiment's ATE is not easily reproduced by chance.
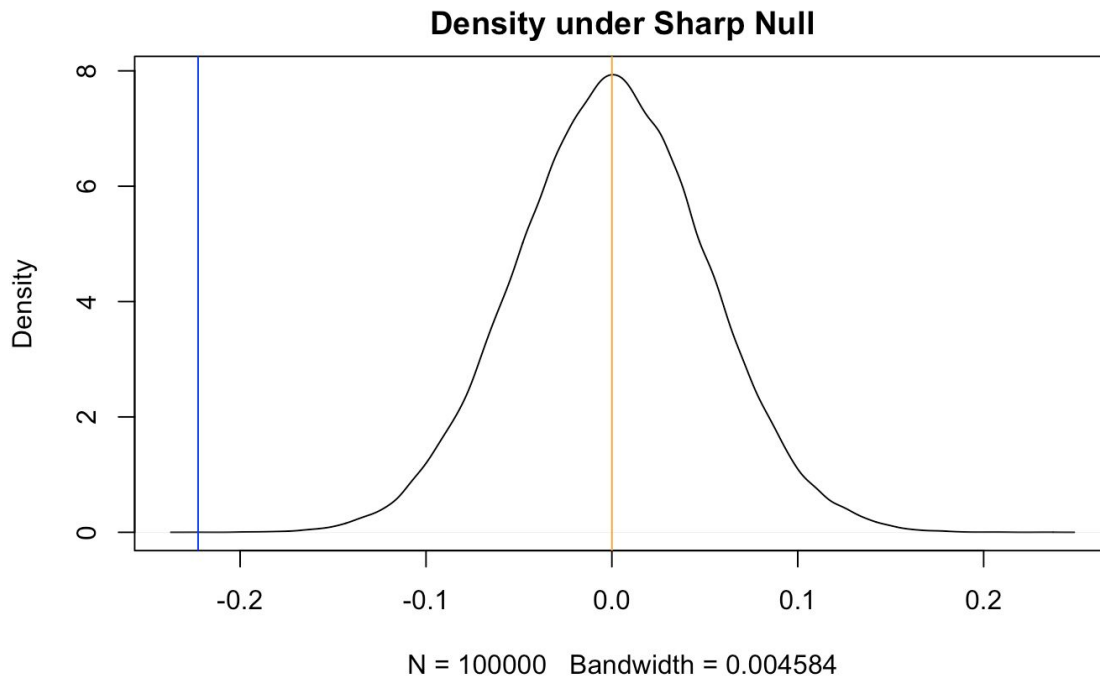
Density under Sharp Null

N = 100000   Bandwidth = 0.004584

Figure 10. Plot of the 100,000 randomized experiments with ATE denoted in blue

DISCUSSION: Study Results

The objective of this experiment is to verify our sharp null hypothesis: that there is no difference in the sample population's trust of speakers whose verbal and non-verbal cues align, compared to speakers whose verbal and non-verbal cues are at odds with one another. From the analysis, we have calculated an ATE value of -0.223 (SE: 0.076), indicating that a subject who watches our treatment video, the flippant version, is more likely to distrust the speaker. Upon plotting a simulation of 100,000 randomized experiments from out data collection, the ATE obtained from our regression happens in only 3 out of 100,000 simulations (p value of 3e-05). This provides sufficient evidence to reject the sharp-null hypothesis.

A significant percentage of our subjects did not make it through the compliance question. One design consideration was to not end the survey for them after the compliance question, but to ask them the measurement question and then calculate the Intent to Treat (ITT) along with the ATE. However, we chose not to calculate the ITT or the CACE, instead eliminating the subjects who didn't answer the compliance question correctly.

We consider that they didn't pay enough attention while watching the video, and thus would not be able to provide a reasonable input in the experiment. These subjects were never presented our actual measurement question about trust vs. non-trust. Because of this, our analysis only includes subjects who were presented the trust vs. non-trust question, leading to an ATE calculation and neither the ITT nor CACE.

The results present a strong evidence for the impact of mis-matched verbal and non-verbal cues in a communication (ATE of -0.223 (SE: 0.076) in all populations and -0.360 (.087) in the MTurk population). However, we can not generalize our results beyond the precise kind of communication, subject population, environment in which the survey was conducted, and the content that we used in the experiment. The majority of communication nowadays has shifted from being conducted in person and being able to see a person's body, to telecommunications and online mediums such as texting, emails, and web meetings. While we still believe that body language makes an impact on speaker trust, a person's reliance on it is not as great as it used to be. As such, more investigation would be needed to conduct the experiment through a different medium, outside of the typical survey tools, and to explore other covariates such as age.

Another reason for our statement on generalization (the fact that we do not feel confident of generalizing it to the entire human population) is that we got very different results among the 3 sub-groups:

1. Friends and Family group has a negative ATE (-0.5 (SE: 0.661)) but the effect is not statistically significant. This is our most trusted group (i.e. we are sure that they watched the video and shared true opinion) but the sample size is small (n = 8) and does not represent the population
2. Lucid group has a small, statistically insignificant, positive ATE (0.018 (SE: 0.139)). We had 99 subjects in this group after removing 72% that did not comply
3. The MTurk group has both a practical and statistically significant, negative ATE (-0.36 (SE: 0.087)). We had 154 subjects in this group after removing 64% that did not comply

We noted the differences, but did not try to correlate or explain them through this experiment. This could be a angle for more analysis.

CONSIDERATION FOR FUTURE STUDIES

Our experiment design presented the trust vs. non-trust question only if the subjects responded to the compliance question correctly. The design was such that the study ended for subjects who didn't answer compliance question correctly. Our analysis showed that such subjects re-attempted the survey until they correctly answered the compliance question and were thus presented with the measurement question. This behavior was deduced based on the subjects' latitude and longitude coordinates provided by Qualtrics. We assume that they re-attempt the survey until they can have a MTurk or Lucid code generated for payment.

Future studies may be designed, taking into consideration how subjects from MTurk, Lucid or any crowdsourcing platforms operate within the context of being able to "sample" the compliance question so that the subject can form a trust vs no-trust opinion after 2 views of the testing video (either control or treatment). We could not find any response patterns among subjects whom we suspected of taking 2 views to answer the compliance question correctly and therefore could not elaborate a suitable analytic plan for this subject strategy.

REFERENCES:

- https://www.theglobeandmail.com/report-on-business/careers/workplace-award/why-its-a-problem-if-your-body-language-doesnt-match-your-words/article3254957
9/
- https://www.businessinsider.com/common-body-language-mistakes-employees-make-2014-4
- https://www.youtube.com/watch?v=kGQ75mL2TOo

APPENDIX

A - GITHUB REPOSITORY

https://github.com/UCB-MIDS/ExperimentsForTrust.git

B - LINKS TO VIDEOS

1. Female speaker, treatment: https://vimeo.com/303927201
2. Female speaker, control: https://vimeo.com/303926147
3. Male speaker, treatment: https://vimeo.com/303932560
4. Male speaker, control: https://vimeo.com/303932013

C - VIDEO SCRIPT

Weight Loss Options in America 2018

Hi. I am Frank Shannon and I have been a family physician for 30 years. I would like to comment on the epidemic of obesity that is sweeping across the US and the world. The CDC recently reported that 40% of the US adult population are obese. Obesity is defined by the BMI, which is an index relationship between a person's body weight in kilograms divided by his or her height in centimeters squared. The threshold number for obesity is a BMI > 30. This would correspond to a 5'10" male who weighs 250 lbs. or a woman who is 5' 6" and weighs 180 lbs. Over and above this threshold, the 8% of the US population are greater than 100 lbs overweight.

On the international front, the World Health Organization has reported that obesity has tripled since 1975. In 2016, 18% have a BMI > 30 and an additional 40% are overweight with a BMI > 25. This epidemic is attributed to increased intake of fast food such as

McDonald's and KFC as well higher sugar and refined carbohydrate intake.

It is intuitively obvious that obesity causes health problems. In a recent long-term study of middle-aged adults living in Framingham, Massachusetts, the CDC documented the influence of obesity. They observed that people who eat a high fat and high sugar diet have twice the incidence of obesity, diabetes, cardiovascular disease and strokes. In addition, the incidence of early mortality was higher in this group.

This study is alarming because it is strong evidence that obesity is caused by an unhealthy diet and produces a shorter life expectancy. This trend calls for a change our diet. There are three general dietary approaches to weight loss. These are:

1. The Atkins Diet recommends high protein and fat intake with very little carbohydrate because excess sugar and refined carbohydrates are converted directly to fat and cause diabetes.
2. The Pritikin Diet recommends the opposite, namely a high intake of carbohydrates and foods that have a low caloric density. This diet is at the center of a lifestyle change which revolves around low stress and daily exercise.
3. Vegetarian/Vegan diet group recommends that all calories consumed should be plant-based. No animal or fish protein sources are allowed. There are different types of vegetarians depending on whether or not milk and/or eggs are allowed.
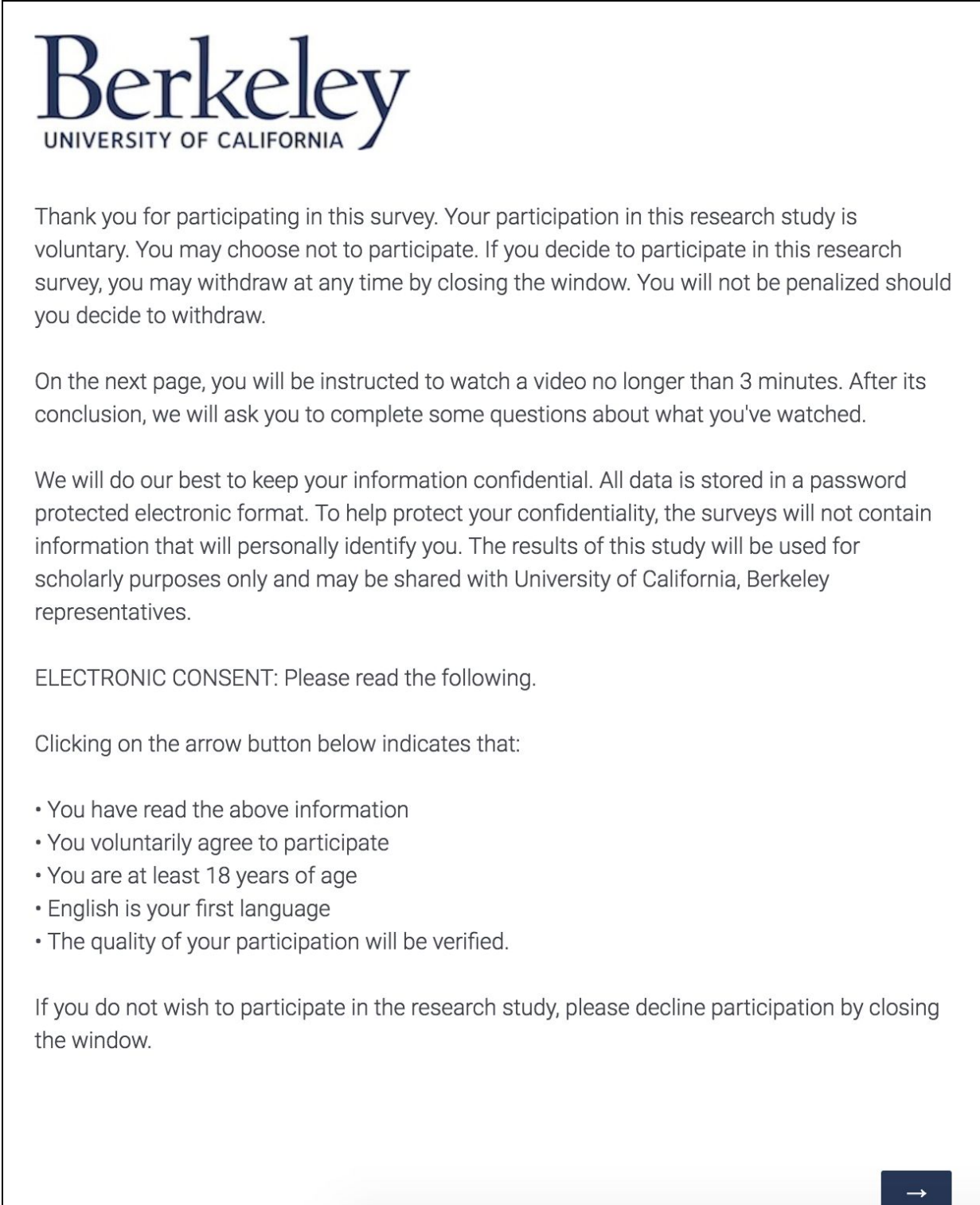
Of the three types of diets, the Atkins Diet has been around the longest and has the highest "total pounds lost". It is successful because a reasonable intake of protein and fats is more satisfying and results in less total caloric intake over the long haul in comparison to the more stringent vegetarian or Pritikin diets.

After hearing about the obesity problem in the US as well as the potential dietary solutions, where do you stand? In the normal weight 60% or obese 40%?

If you are in the "fat forty-percent" group, do you think you need to change your diet and lifestyle?

If so, do you think you need a new diet solution? Which diet type sounds the most appealing and feasible?

Berkeley
UNIVERSITY OF CALIFORNIA

Thank you for participating in this survey. Your participation in this research study is voluntary. You may choose not to participate. If you decide to participate in this research survey, you may withdraw at any time by closing the window. You will not be penalized should you decide to withdraw.

On the next page, you will be instructed to watch a video no longer than 3 minutes. After its conclusion, we will ask you to complete some questions about what you've watched.

We will do our best to keep your information confidential. All data is stored in a password protected electronic format. To help protect your confidentiality, the surveys will not contain information that will personally identify you. The results of this study will be used for scholarly purposes only and may be shared with University of California, Berkeley representatives.

ELECTRONIC CONSENT: Please read the following.

Clicking on the arrow button below indicates that:

• You have read the above information
• You voluntarily agree to participate
• You are at least 18 years of age
• English is your first language
• The quality of your participation will be verified.

If you do not wish to participate in the research study, please decline participation by closing the window.

→

Figure 11. First page of the survey (Consent)

Figure 12. Video to watch in the survey. Videos were randomized at this stage

Figure 13. Alert to answer two questions after watching the video
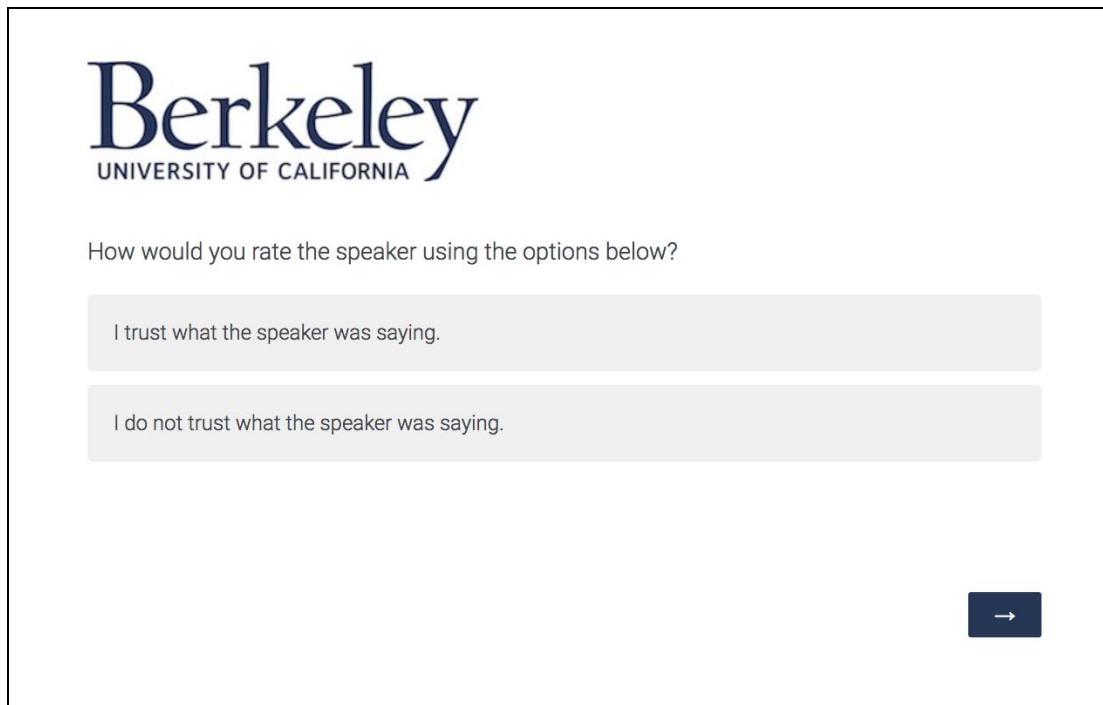
Figure 14. Compliance question. Subjects could pick one of three options

Figure 15. Treatment question. Subjects were shown this question only if they correctly answered the compliance question

Figure 16. End of survey for workers who answered the two questions. The code generated would then be used to receive payment on either MTurk or Lucid

Berkeley
UNIVERSITY OF CALIFORNIA

Thank you for taking our survey. As stated in the Consent Form, there are certain requirements that must be met in order to participate and receive compensation.
You are seeing this message because you are not eligible to complete the study and receive compensation. This may be due to any of the following reasons:

- You do not agree to participate.
- You are under 18 years old.
- English may not be your first language.
- You failed to answer a question that checked to see if you read and understood the instructions.

This follows Amazon Mechanical Turk policy, which states that "a Requester may reject your work if the HIT was not completed correctly or the instructions were not followed."
You may close this window or use your explorer bar to navigate back to the Amazon Mechanical Turk site

Figure 17. Notice that subjects were directly taken to if they did not answer the compliance question correctly.

Thank you for participating in this survey. Your participation in this research study is voluntary. You may choose not to participate. If you decide to participate in this research survey, you may withdraw at any time by closing the window. You will not be penalized should you decide to withdraw.

On the next page, you will be instructed to watch a short video. After its conclusion, we will ask you to complete some questions about what you've watched.

We will do our best to keep your information confidential. All data is stored in a password protected electronic format. To help protect your confidentiality, the surveys will not contain information that will personally identify you. The results of this study will be used for scholarly purposes only and may be shared with University of California, Berkeley representatives.

ELECTRONIC CONSENT: Please read the following.

Clicking on the arrow button below indicates that:

• You have read the above information
• You voluntarily agree to participate
• You are at least 18 years of age
• English is your first language
• The quality of your participation will be verified.

If you do not wish to participate in the research study, please decline participation by closing the window.

→

Figure 18. First page of the pilot survey (Consent)

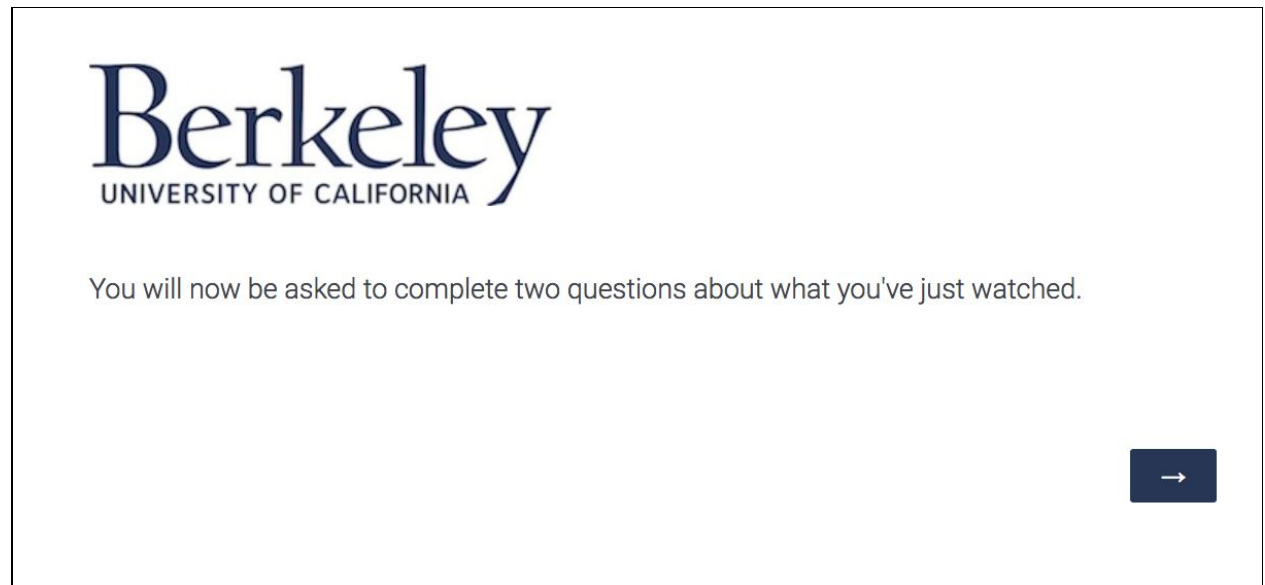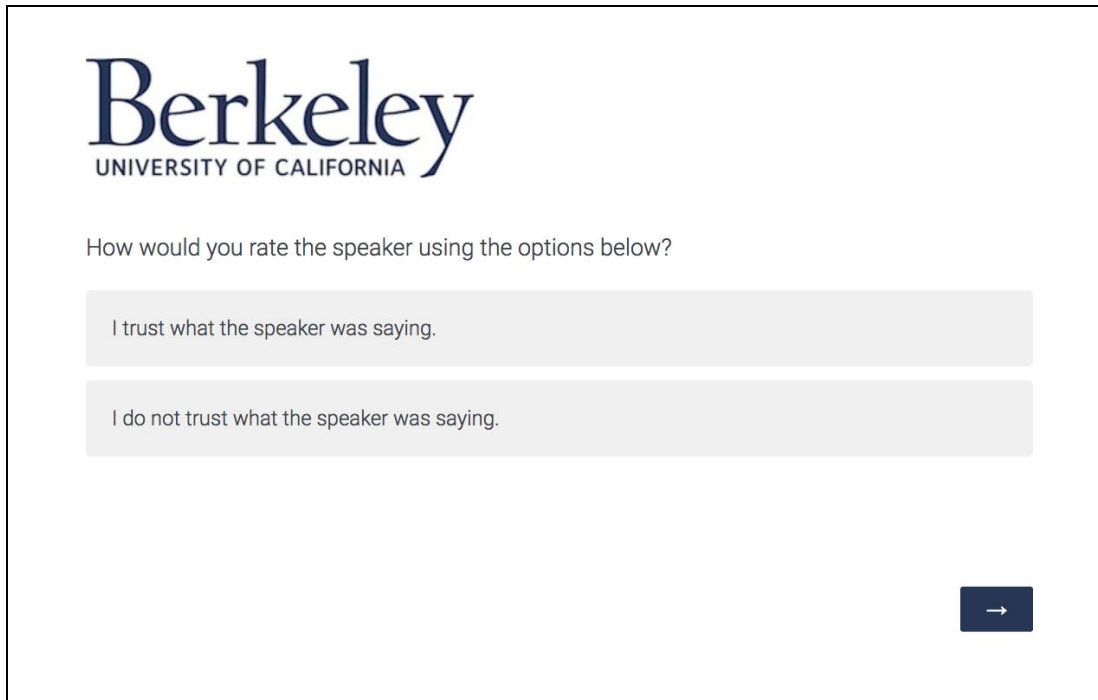Figure 19. Video to watch in the pilot survey. Videos were randomized at this stage

Figure 20. Alert to answer two questions after watching the video

Figure 21. Compliance question. Subjects were asked to answer in a particular manner

Figure 22. Treatment question. Subjects were shown this question only if they correctly answered the compliance question

Figure 23. End of survey for workers who answered the two questions. The code generated would then be used to receive payment on MTurk (MTurk was the sole source for the pilot)

Berkeley
UNIVERSITY OF CALIFORNIA

Thank you for taking our survey. As stated in the Consent Form, there are certain requirements that must be met in order to participate and receive compensation.
You are seeing this message because you are not eligible to complete the study and receive compensation. This may be due to any of the following reasons:

- You do not agree to participate.
- You are under 18 years old.
- English may not be your first language.
- You failed to answer a question that checked to see if you read and understood the instructions.

This follows Amazon Mechanical Turk policy, which states that "a Requester may reject your work if the HIT was not completed correctly or the instructions were not followed."
You may close this window or use your explorer bar to navigate back to the Amazon Mechanical Turk site

Figure 24. Notice that subjects were directly taken to if they did not answer the compliance question correctly.