Pre-analysis plan:
Foreign influencer operations: How TikTok shapes American perceptions of China

We measure the effect of pro-China social media influencer content on attitudes and behavior using an experimental design. We first expose participants to a baseline survey, then randomly assign participants to either:

(1) A group that is required to watch 4 minutes TikTok videos to which they are exposed.

(2) A group that is exposed to 4 minutes of TikTok videos, and is free to watch or skip any videos they choose, as they would on the real TikTok platform.

Within group (1), respondents will then be randomly assigned to a random sample of videos from Chinese state media accounts, videos from pro-China influencer accounts, or entertainment-related placebo videos unrelated to China. Within the pro-China videos, videos will be randomly assigned with equal probability on the subjects of politics, economics, and culture, from a larger corpus with equal probability of assignement. Respondents in this group *must watch all of the videos* (i.e., there is no option to skip videos).

Within group (2), respondents will also be randomly assigned to a random sample of videos from Chinese state media accounts, videos from pro-China influencer accounts, or entertainment-related placebo videos unrelated to China. Within the pro-China videos, videos each will be randomly assigned with equal probability from a larger corpus of videos on the subjects of politics, economics, and culture, with equal probability of assignment. However, respondents in this group *can freely choose to watch or skip any of the videos*, but will be required to spend at 4 minutes before proceeding.

Following treatment assignment and exposure, respondent will complete an endline survey.

A visual overview of the experimental design can be found in Figure A13.

This pre-analysis plan was originally registered prior to collection of data for a pilot study on 500 respondents. In response to findings from the pilot study, we made the following changes: (a) adding the factor analysis outlined below; (b) adding the continuous LATE/2SLS estimation strategy outlined below; (c) adding a series of questions on partisan identification; (d) changing the probabilities of assignment to the forced and free choice groups to increase sample size allocated to the free choice group; (e) updating the power analysis to reflect effect sizes from the pilot study. We then filed an updated pre-analysis plan before additional data collection.

Treatments

Our treatments take the form of pro-China TikTok videos created by influencers, pro-China TikTok videos created by official state media, or entertainment-related control videos unrelated to China. We will draw from a total corpus of approximately 150 China-related videos that will be randomly assigned to respondents. Within this corpus, we pre-classify whether the videos are cultural, political, or economic in nature, and randomly with equal probability across categories.

Treatment assignment

Respondents will be recruited from Cint Thoerem, and will be balanced on age, gender, ethnicity, and region. Respondents will then be randomly assigned to one of the two treatment groups (*Forced* or *Free choice*) with 20% probability of assignment to the *Forced* group and 80% to the *Free Choice* group).[12] Respondents are then randomly assigned to one of three conditions (*Placebo*, *State Media*, or *Influencers*) within these groups with equal probability. Finally, within the *State Media* and *Influencer* groups, respondents will be randomly assigned political, economic, and culture videos with equal probability of assignment. In short, at each step there will be equal probability of assignment using complete random assignment. See below for an overview of the assignment steps.

1. *Forced*:

    1a. *Placebo*: Randomly assigned to view either (1) nature or (2) viral entertainment videos with equal probability.

    1b. *State Media*: Randomly assigned to view Chinese state media videos.

    i. Among Chinese state media videos, randomly assigned politics, economics, and culture videos with equal probability of assignment.

    1c. *Influencers*: Randomly assigned to view pro-China influencer media videos.

    i. Among pro-China influencer videos, randomly assigned political, economic, and cultural videos with equal probability of assignment.

2. *Free choice*:

    2a. *Placebo*: Randomly assigned to view either (1) nature or (2) viral entertainment videos with equal probability.

    2b. *State Media*: Randomly assigned to view nature, entertainment and Chinese state media videos. Can skip videos at will.

    i. Among Chinese state media videos, randomly assigned political, economic, and cultural videos with equal probability of assignment.

    2c. *Influencers*: Randomly assigned to view nature, entertainment and pro-China influencer videos. Can skip videos at will.

[12] We randomly assign more respondents to the *Free Choice* group as we expect smaller treatment effects from this group, and therefore wish to increase sample size and therefore power in this group. See the power analysis below for justification of the selected sample sizes in each condition.

i. Among pro-China influencer videos, randomly assigned political, economic, and cultural videos with equal probability of assignment.

An overview of the experimental design can be found in the figure below.

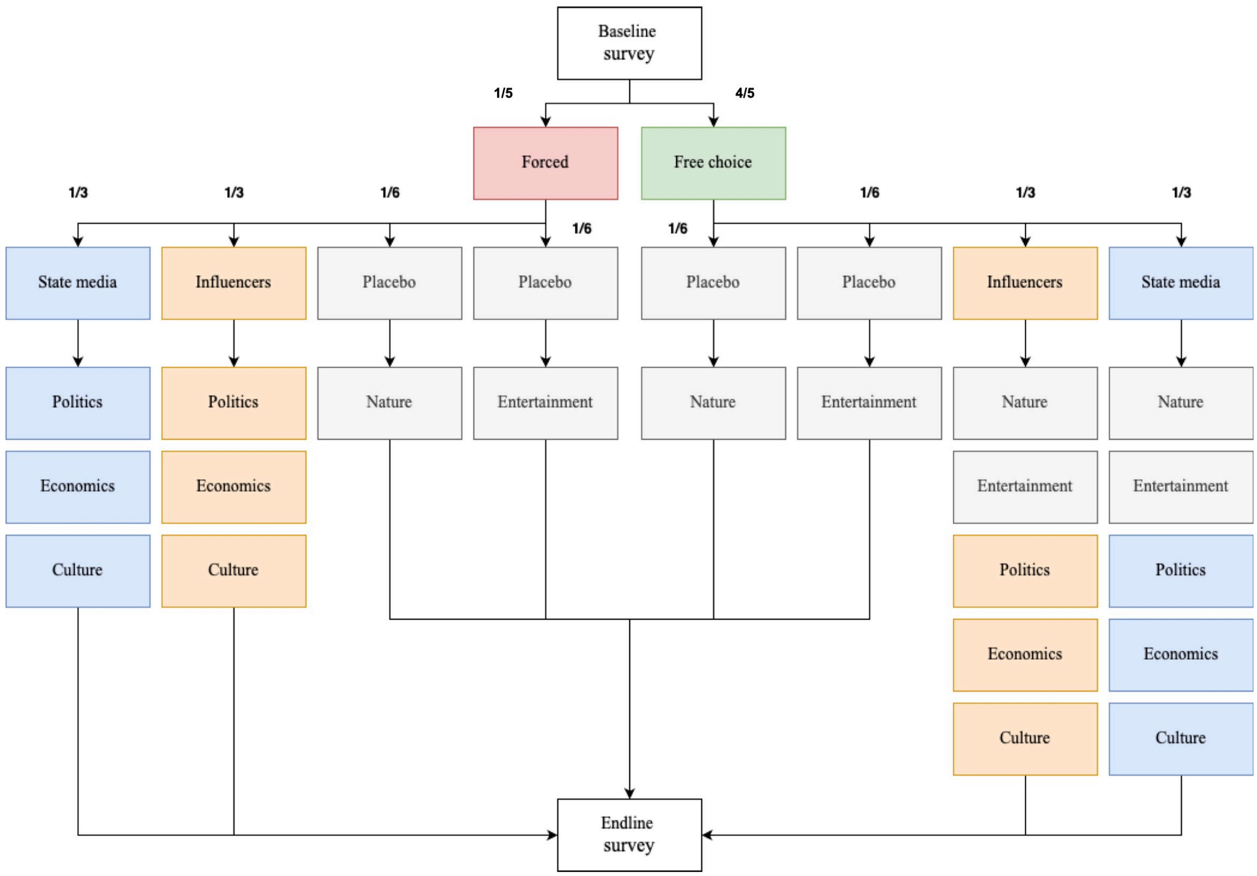

Figure A13: Overview of experimental design.

Note: Fractions indicate probability of assignment at each branch.

Compliance

The ability to watch or skip pro-China videos in the free choice treatment arms is equivalent to compliance in a one-sided noncompliance framework. As subjects can choose to watch any duration of pro-China videos they choose, this represents a continuous measure of compliance sometimes referred to as variable treatment intensity. Angrist and Imbens (1995) show that two-stage least squares (2SLS) applied to a causal model with variable treatment intensity and nonignorable treatment assignment identifies a weighted average of per-unit treatment effects.

We therefore pre-register the weighted average local average treatment effect (LATE)—i.e., the marginal effect of an additional second of watching pro-China videos—from a 2SLS estimation strategy with a continuous measure of treatment intensity defined as the total number of seconds spent watching pro-China videos as our primary compliance-adjusted estimator. Because the marginal effect of each second watched may vary, we will also provide a plot summarizing the effect for each second. This is essentially a binning analysis, which is a common technique to deal with non-linear effects in continuous compliance design, where each bin represents one additional second watched.

We also note that compliance could be defined in a binary or categorical manner. However, our primary definition of compliance will be the continuous measure described above. For a binary definition of compliance in a potential outcomes framework where d refers to treatment status and z refers to treatment assignment, compliance in the free choice treatment groups can be defined as: $d_i(z_i = 1) = 1$ (subject i is assigned to the pro-China treatment and watches *one or more* pro-China videos). Non-compliance can similarly be defined as $d_i(z_i = 1) = 0$ (subject i is assigned to the pro-China treatment but does not watch any pro-China videos). It is also possible to classify compliers in terms of categorical levels of compliance as respondents in the free choice group may choose to watch anywhere between 0 to N pro-China videos. We can therefore also alter our definition of compliance to $d_i(z_i \in \{1, 2, \ldots, N\})$ where $d_i(z_i = 1)$ now refers to subject i is assigned to the pro-China treatment and watches *1* pro-China video, $d_i(z_i = 2)$ refers to subject i is assigned to the pro-China treatment and watches *2* pro-China videos, and $d_i(z_i = 3)$ refers to subject i is assigned to the pro-China treatment and watches *3* pro-China videos, and so on.

In all of the approaches above, we calculate complier average causal effects using 2 stage least squares (2SLS) where treatment assignment Z is an instrument for treatment receipt D. Additionally, in all of the approaches described above, the exclusion restriction assumption in this setting is that random assignment of pro-China videos is correlated with the treatment of *watching* pro-China videos, and is only correlated with our dependent variables through actually *watching* the pro-China videos.

Hypotheses

We will test the following core hypotheses:

- *Hypothesis 1*: Viewing videos produced by pro-China influencers and state media will increase respondent affinity for China relative to the placebo condition. (See the key outcome question in Outcomes).
- *Hypothesis 2*: Viewing videos produced by pro-China influencers will cause larger shifts in affinity for China than viewing videos produced by Chinese state media. (See the key outcome question in Outcomes).
- *Hypothesis 3*: In the free-choice arm, compliance for those assigned to pro-China influencers will be higher than compliance for those assigned to watch state media.

We also test the following secondary hypotheses:

- *Hypothesis 3a*: On average, respondents in the forced viewing group will increase affinity for China more than respondents in the free choice group Outcomes).
- *Hypothesis 3b*: Compliers in the free choice group will increase affinity for China less than respondents in the forced choice group (based on the outcome variables defined in Outcomes).
- *Hypothesis 4*: Respondents assigned to watch pro-China influencers in the forced choice arm will have higher positive affect than those assigned to watch Chinese state media (based on the outcome variables defined in Outcomes).
- *Hypothesis 5*: Videos produced by Chinese state media will shift American respondents' views on China's political system, on the Chinese economy, and U.S.-China policy issues toward those preferred by the Chinese state. Viewing Chinese state media will have no effect on the question asking respondents to sign the petition memorializing the 1989 protests. (See outcome variables defined in Outcomes). Policy-related videos will be more effective at this than cultural videos.
- *Hypothesis 6*: Videos produced by Chinese influencers will shift American respondents' views on China's economy but will have null effects on attitudes to the Chinese political system and U.S.-China policy issues towards those preferred by the Chinese state. (See outcome variables defined in Outcomes). Viewing pro-China influencers will have no effect on the question asking respondents to sign the petition memorializing the 1989 protests. Policy-related videos will be more effective at this than cultural videos.
- *Hypothesis 7*: Attitudes will shift more for younger respondents and less educated respondents when compared to older respondents and higher educated respondents.
- *Hypothesis 8*: Influencer videos will cause respondents to feel inspired, relative to placebo. State media videos will cause respondents to feel upset.
- *Hypothesis 9*: Respondents to the center and right will be more likely to be persuaded by influencer videos than respondents to the left.

Outcomes

We examine the following primary outcome variables of interest:

Primary outcome: China favorability

- Do you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of China?

Behavioral measure: willingness to take political action related to China

- In June 1989, following student-led protests, hundreds of anti-government protesters and citizens were killed by People's Liberation Army troops. Would you like to sign a petition memorializing the Tiananmen incident and condemning the Chinese government? This petition is run by Amnesty International and is not affiliated with the researchers. [Yes, at the end of the survey, please take me to the petition. / No thank you, at the end of the survey, please do not take me to the petition.]

  We define the behavioral outcome as a binary indicator equal to 1 if a respondent both indicates that they would like to be taken to the petition *and* clicks on the link to the petition, and 0 otherwise.

Secondary outcomes: attitudes towards China

Attitudes towards China's economy:

- To what extent do you agree with the following statement? I admire the Chinese economy.
- To what extent do you agree with the following statement? China has an advanced economy.
- To what extent do you agree with the following statement? China is a world leader in technology.

Attitudes towards China's political system:

- To what extent do you agree with the following statement? I admire the Chinese political system.
- To what extent do you agree with the following statement? The Chinese political system is legitimate.
- To what extent do you agree with the following statement? The Chinese political system is responsive to the needs of the Chinese people.

Attitudes towards Chinese culture:

- To what extent do you agree with the following statement? I admire Chinese culture.

- To what extent do you agree with the following statement? Chinese culture has had a positive influence on the world.
- To what extent do you agree with the following statement? I am interested in learning more about Chinese culture.

Attitudes towards China-U.S. Policy

- To what extent do you agree with the following statement? China is an enemy of the United States.
- To what extent do you agree with the following statement? The United States should cooperate closely with China on trade issues.
- To what extent do you agree with the following statement? The United States should cooperate closely with China on security issues.

Mechanism outcomes:

Questions measuring negative affect and positive affect from the PANAS scale.

- Please indicate how strongly you are feeling the following emotions:
  1. Inspired (Not at all, a little, somewhat, rather strong, extremely strong
  2. Upset (Not at all, a little, somewhat, rather strong, extremely strong
  3. Interested (Not at all, a little, somewhat, rather strong, extremely strong
  4. Excited (Not at all, a little, somewhat, rather strong, extremely strong

  Analysis of open-ended responses including topic modeling approaches and qualitative scoring.

Our secondary outcomes will be combined into three indices—economy, politics, and culture—using factor analysis. The code used to conduct the factor analysis is provided below. Factor analysis will not be applied to the foreign policy attitudes questions. Multiple comparisons corrections will be applied to the secondary and exploratory outcomes using the Bonferroni, Holm-Bonferroni, and Benjamin-Hochburg procedures.

Factor analysis function:

```
create_factor <- function(data, dv_names, verbose = TRUE) {

# Keep variables of interest and ensure that they're numeric
sub <- data[, dv_names] %>% mutate_all(funs(as.numeric(as.character(.))))

# Impute missing values with median
impute <- sub
```

```
  impute <- sapply(impute, function(x) ifelse(is.na(x), median(x, na.rm = T), x))

  # Principal component, scaled to have mean 0/sd 1
  f <- princomp(impute, cor = TRUE)
  if (verbose) print(loadings(f))
  dv <- f$scores[, 1]
  dv <- as.numeric(scale(dv))

  # Make sure the variable points the correct way ()
  if (cor(dv, data$outcome, use = "complete") < 0) dv <- -1 * dv

  # If a row in the original data has more than 50% NAs, then replace the score
  # with NA
  bool <- apply(sub, 1, function(x) sum(is.na(x)) / ncol(sub) > 0.5)
  dv[bool] <- NA
  dv

}
```

Estimation procedures

Our primary estimands are: (1) the average treatment effect of watching a pro-China influencer video on the outcomes listed above when compared to the Placebo group, (2) the average treatment effect of state media videos on the outcomes listed above when compared to the Placebo group, and (3) the complier average causal effect of watching a pro-China influencer video on the outcomes listed above when compared to the Placebo group, and (4) the complier average causal effect of watching a pro-China state media video on the outcomes listed above when compared to the Placebo group.

We will also compare effect sizes among those in the free choice group and forced choice group. For the free choice group we will calculate both intent-to-treat (ITT) effects and complier average causal effects.

Average Treatment Effect

For those in the forced choice group, we will estimate the following average treatment effects (ATE), and the estimator will include covariate adjustment:

1. *State Media* vs. *Placebo*
2. *Influencer* vs. *Placebo*
3. *Influencer* vs. *State Media*

We include the following pre-treatment covariates in the regression specification: *gender, age, education, national pride, and left-right political orientation.* In the event of missingness, missing covariates will be imputed using the predictive mean matching method in the MICE package in R. We will compute HC2 robust standard errors.

The ATE will be estimated using the "lm_robust" function in the "estimatr" package in R (Blair, Cooper, Coppock and Humphreys 2019). The code that will be used is as follows:

```
lm_robust(outcome ~ treatment + covs, data = df)
```

where covs is the list of covariates above and country is an indicator for each of the 19 countries. In the event that standard covariate adjustment worsens precision, we will estimate treatment effects using the estimator outlined by Lin (2013) using the code below:

```
lm_lin(outcome ~ treatment, covs, data = df)
```

Results without covariate adjustment will be reported in the appendix. We expect these results to be similar but less precisely estimated due to the exclusion of prognostic covariates. When interpreting the results, we will rely primarily on the covariate-adjusted estimates. We will also calculate randomization inference p-values for the main outcome variables in the appendix.

Intent-to-treat effects

For those in the free choice group, we will calculate the same ITTs as the ATEs described above, substituting receipt of treatment for assignment to treatment in the regression specification. We will also compare the ITT in the free choice group to the ATE in the forced choice group. The ITT will be estimated using the "lm_robust" function in the "estimatr" package in R:

```
lm_robust(outcome ~ treatment_assignment + covs, data = df)
```

Local average treatment effects

We will also estimate local average treatment effects / complier average causal effects (CACE) as described in the "compliance" section above. We will calculate the same LATEs as the ITTs/ATEs described above. We will estimate the LATE using two-stage least squares using the "iv_robust" function in the "estimatr" package in R:

```
iv_robust(
outcome ~ treatment_receipt + covs | treatment_assignment + covs, data = df
)
```

Treatment effect heterogeneity

We will examine the following heterogeneous treatment effects:

1. Age - under 30 vs. over 30.
2. Education - college educated vs. non-college educated
3. Left-right political alignment - left vs. center vs. right
4. Partisan identification - Republican vs. Democrat

We will examine treatment effect heterogeneity by calculating conditional average treatment effects (CATEs). A CATE is an average treatment effect specific to a subgroup of subjects, where the subgroup is defined by subjects' attributes. Heterogeneous treatment effects will be estimated by regressing the outcome variables on treatments for each of the subgroups specified in Treatment effect heterogeneity. This will be conducted using the "lm_robust" function in the "estimatr" package in R.

```
lm_robust(outcome ~ treatment, data = df, subset = covariate == "subgroup")
```

Threats to inference

Attrition

We will examine whether there is differential attrition between the placebo and treatment groups, as well as between the forced choice and free choice groups. We will conclude that there is significant attrition if the difference in the rate of attrition between the groups is statistically significant ($p<0.05$). If there is significant differential attrition, we will present reweighted results and trimming bound results in the paper's appendix, as appropriate, following the procedures outlined in Gerber and Green (2012).

Attention checks

In light of recent evidence of decreased attention in online samples (Peyton, Huber and Coppock 2020), respondents will be screened according to pre-treatment attention checks and dropped from the sample of analysis if they fail the attention check. Our attention checks will take the following form:

1. "For our research, careful attention to survey questions is critical! To show that you are paying attention, please select 'I have a question'."
2. "People have different tastes in movies. For this question, however, we are not interested in your taste but want to test whether survey takers are reading questions carefully. Below, please select the options "Romance" and "Science Fiction.""

Potential Early Survey Stop

One concern with online samples and our survey partner is the possibility that as the fielding of the survey progresses, the quality of responses may decrease. While fielding the survey, we will monitor the rate at which surveys are completed and the length of survey completion. If we observe a significant slow-down in the rate at which survey takers are completing the survey, we will halt the survey. If we observe a significant decrease in the amount of time it is taking survey takers to complete the survey below a 10 minute baseline, we will halt the survey. At this point we will not analyze the outcome data, but if based on indications like the number of speeders we assess that the survey is no longer receiving high-quality responses, we will end the survey.

Power analysis

The power analysis below depicts the intent to treat effect of the three treatment groups (Placebo, State Media, and Influencers), separately for the forced and free choice groups. The analysis indicates that due to expectations of larger treatment effects sizes in the forced group, a larger sample should be expended on the free choice group. The power analyses below assume the distributions of potential outcomes / treatment effect sizes among respondent as found in the pilot study and were run using 10,000 simulations.

Forced choice

The power analysis for the forced choice group indicates sufficient power (defined here as the conventional 80% power) with a sample size of roughly 500 respondents. This is not surprising as significant results were uncovered from the pilot alone in the forced choice group. The distributions of p values across 500 simulations can be found in the figure below.

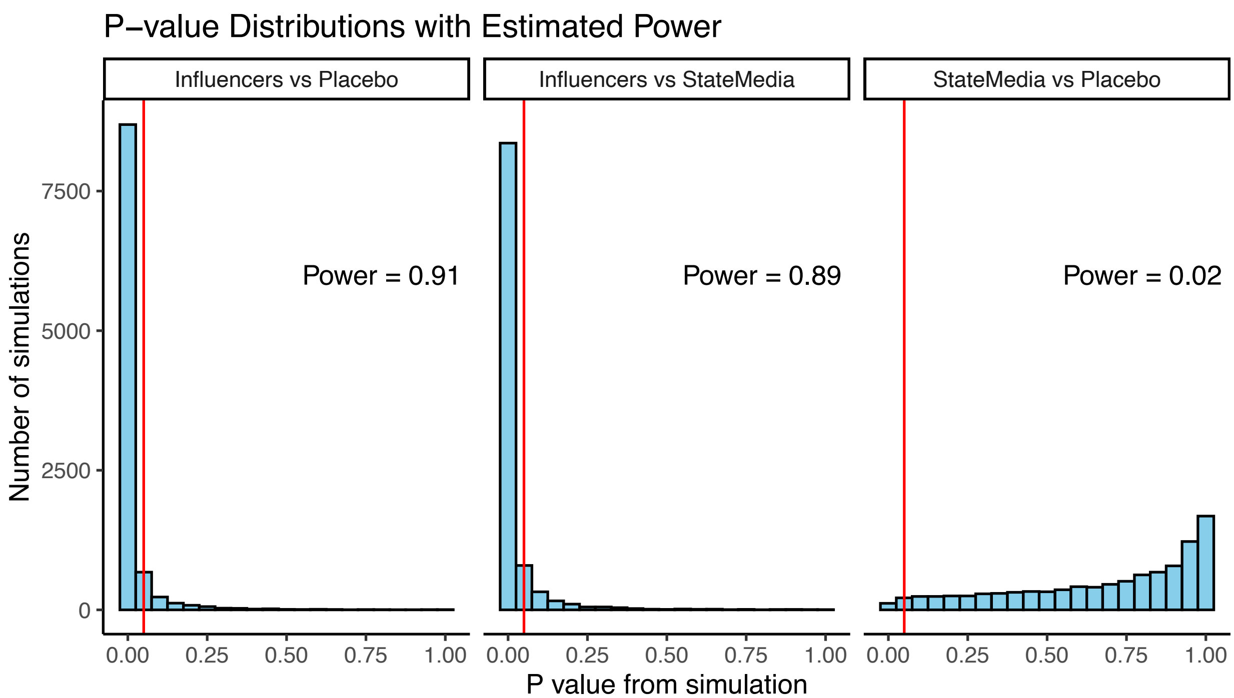

Figure A14: Distribution of p values in forced arm (N = 500 respondents)

Free choice

However, for the free choice group, due to high expected levels of noncompliance and a correspondingly lower ITT, the required sample size to reach 80% power is significantly larger. Using the distribution of potential outcomes from the pilot, we do not reach 80% power in each of the treatment arms until there is a sample size of 6000 respondents in the free choice group.

The distributions of p values across 500 simulations can be found in the figures below.

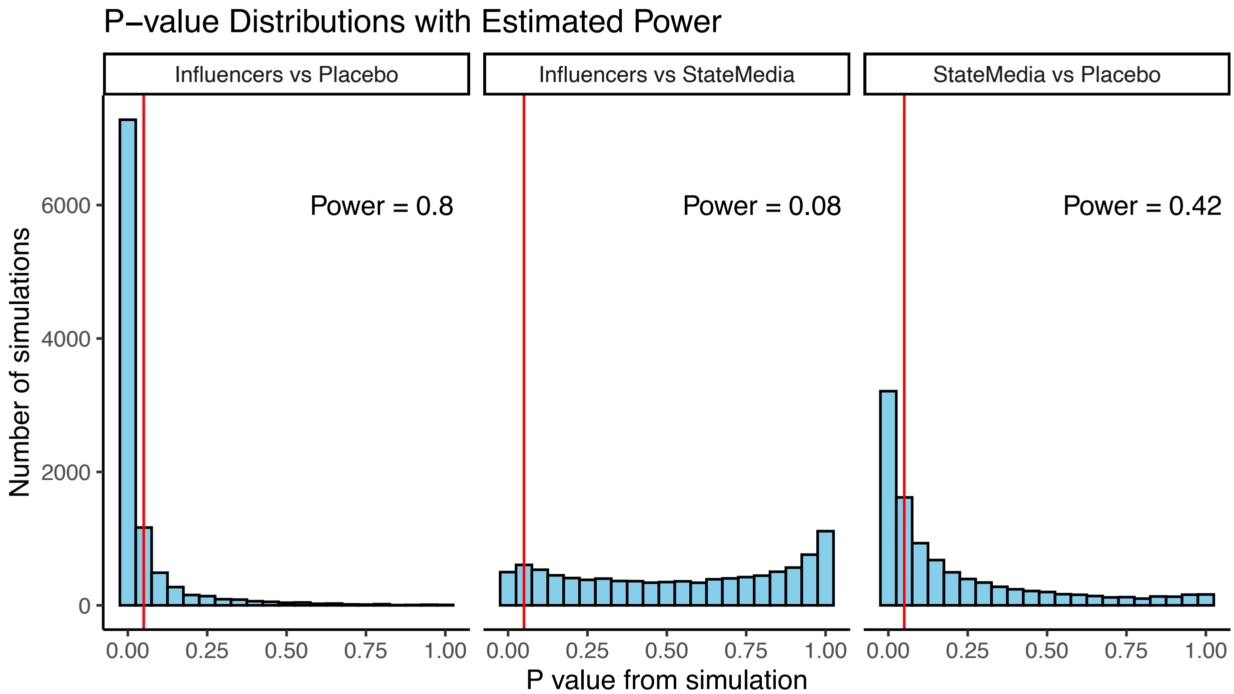

Figure A15: Distribution of p values in free choice arm (N = 6000 respondents)

These simulations do not cover additional sample size that would likely be necessary to detect differences between the different types of videos or heterogeneous effects by age, education, or pre-existing affinity for China. To ensure sufficient power to detect these effects, additional respondents would be necessary.