Pre-analysis plan:
The effect of social media framing and cognitive dissonance on citizens' responses to the Covid-19 in Brazil

Ernesto Calvo *   Tiago Ventura †

May 15, 2020

Abstract

In this proposal, we describe a survey experiment designed to assess the effects of social media exposure on perceptions of personal risk and government responses to the COVID-19 pandemic. We implement a framing experiment treating respondents with distinct social media messages on Twitter. Our experiment manipulates the author of the message, rotating between two prominent politicians, and the content of the tweet, a neutral tweet with equal wording, and a negative polarized view about the crises in which each side blames its opponent for the crises. We measure the effect of the social media exposure against the risk perceptions regarding personal health and job security during the crises, as well as perceptions about government response to the pandemic. Our design allows us to measure the effects of cognitive dissonance on social media and the effects of blame attribution on political behavior and policy preferences during this crucial time.

*University of Maryland, Government and Politics, UMD. Address: 3140 Tydings Hall, College Park, MD 20742, USA. Email: ecalvo@umd.edu. Webpage: http://gvptsites.umd.edu/calvo/

†University of Maryland, Government and Politics, UMD. Address: 4118 Chiconteague, College Park, MD 20742, USA. Email: venturat@umd.edu. Webpage: http://tiagoventura.rbind.io/

Contents

1 Introduction **1**

2 Experimental Design **2**

2.1 Treatment Conditions 2

2.2 Outcome Variables 6

3 Hypothesis **6**

3.1 The effect of negative message about the COVID-19 7

3.2 Extensions: Emotional Response and Attention 8

4 Analysis **9**

4.1 Estimation 10

1 Introduction

During the first weeks of January, news about the rapid spread of a new virus, COVID-19, in the Hubei province of China took the world by surprise. As the Chinese government initiated a lockdown of millions of its citizens, early responses by governments around the world focus their attention on restricting trade and travel to and from the affected areas. By mid-March however, efforts to contain the outbreak proved insufficient, with the COVID-19 spreading widely across the globe. On March 11, 2020, the World Health Organization declared the rapidly-spreading coronavirus outbreak a pandemic, acknowledging that the virus would likely spread to every country on the globe.

While some governments have promptly adopted policies to enforce social distancing as a strategy to mitigate the consequences of the pandemic, political leaders in a few countries resisted calls for swift action. The populist President of Brazil, Jair Bolsonaro, provides a textbook example of such behavior. As the cases of COVID-19 in Brazil grew exponentially, President Bolsonaro asked citizens to go back to work and help the economy, blaming the media for ”hysterical” reporting on the spread of the virus. Bolsonaro accused the opposition of profiting from the crisis for political gains, and urged mayors and state governors to roll back lockdown measures.

By the time his nationally televised presidential message concluded, the government and the opposition rapidly developed distinct narratives on social media (Aruguete and Calvo, 2018; Calvo, 2020; Lin et al., 2014). Bolsonaro’s supporters applauded President Bolsonaro on social media, echoing his ”business-as-usual” message while minimizing the risks of the COVID-19. The opposition, the media, and most institutional actors, by contrast, criticized the president for a risky and polarizing message that failed to answer to the challenges of the pandemia.

In this pre-registration proposal, we describe a survey experiment that models the effects of these competitive frames on risk perceptions and perceptions of government response to the COVID-19 pandemic. We implement the survey using an online probabilistic sample of Brazilian respondents. We use edited tweets to prime respondents on the author and the content of the tweets. In the former, respondents are exposed to tweets from a like-minded, prominent political figure. For the content priming, respondents are exposed to a neutral tweet with the same wording from each politician, or a negative tweet from each politician blaming the other side for the pandemic crises.

2 Experimental Design

Our experiment implements a four-arm treatment assignment in which each respondent is randomly exposed to one of four different tweets, with a variation on the content and the author of the message. Each respondent will be exposed to only one tweet, and after the treatment assignment, will respond to our outcome variables. Below, we describe the treatment conditions and the outcomes. The experiment was included in a national online survey in Brazil. The survey is fielded by Netquest-Vanderbilt, with probabilistic samples draw by the LAPOP team in Vanderbilt from users registered with Netquest.

2.1 Treatment Conditions

We edited tweets to prime respondents in our experiment. Although we reduce the external validity of the experiment by not using real tweets for our treatment conditions, we carefully chose the wording of the tweets based on actual public statements and social media activity to maximize the validity of each treatment conditions. The randomization procedures guarantee internal validity [1].

We vary only two features of each tweet, the author and the content. For the author, we use two prominent political figures: Eduardo Bolsonaro, congressmen and son of President Jair Bolsonaro, and Fernando Haddad, the front-runner in the 2018 national election. We choose high-level politicians to ensure congruence or dissonance between the message and the respondents' preferences.

For the content, we vary between a neutral and a negative framing for the COVID-19. In the neutral, we use precisely the same wording for each author, in which the tweets mainly highlight the existence of a crisis and the importance of President Bolsonaro to lead the institutional efforts to fight the pandemic. For the negative tweets, we created one for each sender, mimicking their political preferences, thus maximizing external validity for the experiment. For Eduardo Bolsonaro, the tweets reinforce the argument that the crisis is not that serious, and that the opposition and the media are responsible for the "hysteria" around the spread of the virus. For Fernando Haddad, the tweet criticizes the government and Bolsonaro's statements minimizing the consequences of the crises.

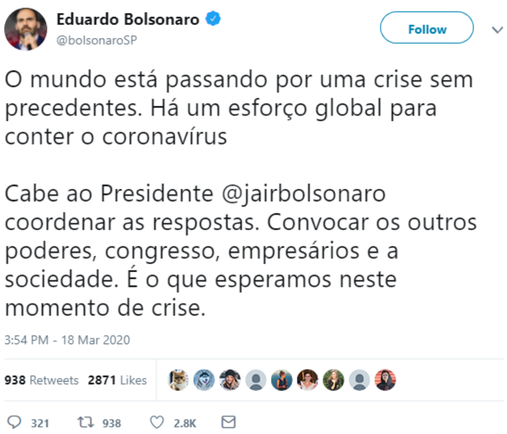
Table 1 presents the treatment conditions. In the following, figure 1 shows the tweets as the respondents will read, in Portuguese, for each four treatment conditions. Each respondent will be exposed to only one of the four arms of the treatment.

[1] The experiment received the approval of the University of Maryland Institutional Board Review 1552091-3

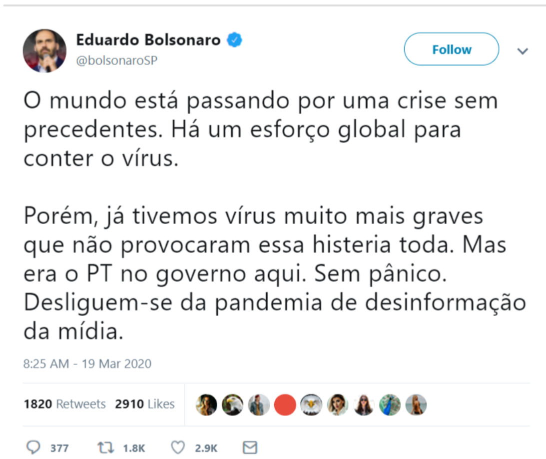
Table 1 Treatment Conditions

	Neutral Tweet	Negative Tweet
Eduardo Bolsonaro	The world is currently living an unprecedented crisis. Countries all over the world rally together to fight against Coronavirus. It is the responsibility of President @jairbolsonaro to coordinate our answers. He needs to act together with Congress, Business leaders, and civil society. This is what we expect in such critical times.	The world is currently living an unprecedented crisis. Countries all over the world rally together to fight against Coronavirus. However, we have already had these types of the virus before, and it did not lead to all this hysteria. But it was the PT in the government here. No panic. Switch off from the pandemic of misinformation from the media
Fernando Haddad	The world is currently living an unprecedented crisis. Countries all over the world rally together to fight against Coronavirus. It is the responsibility of President @jairbolsonaro to coordinate our answers. He needs to act together with Congress, Business leaders, and civil society. This is what we expect in such critical times.	The world is currently living an unprecedented crisis. Countries all over the world rally together to fight against Coronavirus. President @jairbolsonaro is delayed in answering. He is more concerned about attacking his opponents and take part in protests that put in risk the health of the Brazilian people.

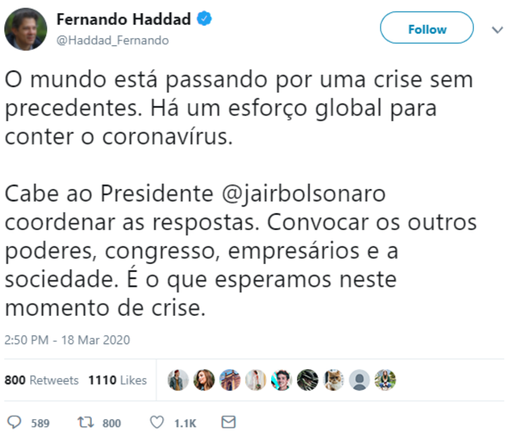
Figure 1 Tweets for the Treatment Conditions



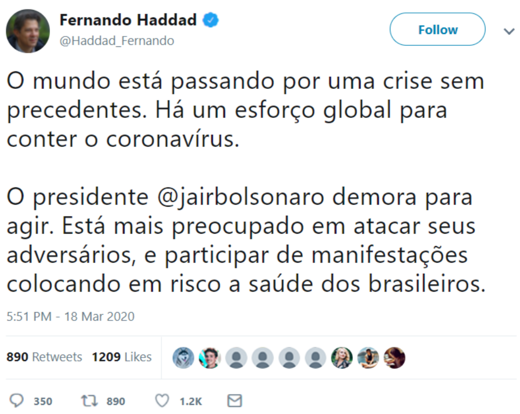
a) Eduardo Bolsonaro x Neutral Tweet (T1)



b) Eduardo Bolsonaro x Negative Tweet (T2)



a) Fernando Haddad x Neutral Tweet (T3)



b) Fernando Haddad x Negative Tweet (T4)

2.2 Outcome Variables

We use three main questions as our outcome variables. The complete wording is presented below. These questions capture perceptions about personal risk during the COVID-19 pandemic and the respondents' opinions about the government's performance during the crises.

- **Question 1:** How likely is that your health would be affected by the COVID-19? (very likely, somewhat likely, somewhat unlikely, very unlikely)
- **Question 2:** Given the current health and economic crisis produced by the Coronavirus COVID-19, how likely is it that you could lose your job? (very likely, somewhat likely, somewhat unlikely, very unlikely)
- **Question 3:** Has the government response been appropriate when faced with the corona COVID-19? (Very appropriate, somewhat appropriate, somewhat unappropriated, very inappropriate).

3 Hypothesis

The hypotheses of the model test for the effect of social media content on perceptions of risk and government performance. We consider the effects of negative and neutral messages and the extent to which the effect interacts with cognitive congruence and dissonance between the authors of the tweet and the respondents' preferences. Let us introduce our hypothesis for each of these dimensions

3.1 The effect of negative message about the COVID-19

The first hypothesis of the experiment expects negative messages to increase perceptions of individual risk and decrease support for the government's response to the pandemic.

- *Hypothesis 1:* We predict that negative messages, compared with the neutral tweets, will increase perceptions of risk and decrease support for the government's response to the COVID-19 pandemic.

We also expect cognitive dissonance between the respondents' preferences and the author of the tweets to interact with the effects of social media. When cognitive dissonance occurs together with a neutral message, as in our tweet with the fixed content, we expect the effects of the social media message to be mitigated. In other words, we argue that respondents who observe a misaligned politician being neutral about the issue (T1 and T3), in our case calling for a national front to fight the crises, will report lower perceptions of risk, and greater support for the government response.

- *Hypothesis 2:* We predict that cognitive dissonance, combined with neutral content in the tweets, will decrease perceptions of risk, and increase support for the government response.

However, we expect effects in the opposite direction when dissonance interacts with negative content. As previous research has shown (Banks et al., 2020), exposure to dissonant social media messages increases *contrast* (Merrill et al., 2003) and renders greater perceived polarization. After being exposed to cognitive dissonant information, the user increases ideological contrast between the user and the politician, and considers not-aligned political messages more distant from their own beliefs.

Following this intuition, we expect that to the extent respondents observe a dissonant partisan signal, combined with a negative message, the framing effects of going negative about the pandemic will be exacerbated (Adida et al., 2018). Therefore, supporters of the opposition will show higher risk perceptions and be more dissatisfied with the government when exposed to a cognitive dissonant source. In contrast, the opposite effects are expected for the partisans of Bolsonaro. We present the hypothesis below:

- *Hypothesis 3:* We predict that cognitive dissonance combined with negative messages will increase framing effects compared with a congruent messages. When exposed to cognitive dissonant tweets:
  - *H3a:* Respondents aligned with the opposition will feel more at risk and show greater negative views about the government.
  - *H3b:* Respondents aligned with the government will feel less at risk and show greater support for the government response to the crises.

We have three question to capture respondents' policy preferences: a) partisanship, b) ideological placement from left to right, and c) vote choice in the 2018 presidential election. We will estimate models using all the three measurement strategies for robustness.

3.2 Extensions: Emotional Response and Attention

After each of the tweet treatments, each respondent is asked to indicate how the content "makes you feel". There is consistent evidence that emotions such as anger and disgust yield stronger political responses (mobilization) while emotions such as sadness elicit disengagement. We expect anger and disgust to mediate our treatment effects.

HE: Negative emotions that elicit anger and disgust will increase "mobilization", boosting the framing effect from the previous hypothesis to a larger extent, compared to than emotions such as "sadness".

We also expect attention to the treatment conditions to moderate the effects of framing and cognitive dissonance. Recent scholarship in both political science and psychology suggests that the amount of time spent on a survey page works as a measure of respondent effort, specifically rendering stronger treatment effects for those who are more cognitively engaged with the treatment (Berinsky et al., 2014; Wise and Kong, 2005; Malhotra, 2008). Similar effects for latency, measured as the time reading tweets, have been shown to increase the effects of social media framing in polarization (Banks et al., 2020).

To capture the effects of attention, we will measure the time the respondents spent reading each of the treatment condition. We expect:

HT: Lower latency (more time reading the tweets) will increase the effects of framing and cognitive dissonance in compliance by agents and in trust by principals.

4 Analysis

In this study, we aim to estimate the average treatment effect for negative messages and the conditional effects with cognitive congruence and dissonance. Specifically, we are interested in two kinds of estimates:

- the average treatment effect of negative messages, abstracting from the author of the tweet, and the effects of cognitive dissonance.
- the conditional average treatment effect for cognitive dissonance with positive and negative messages.

4.1 Estimation

Our key independent variable is the respondent's treatment assignment. We manipulate our four treatment arms to identify the estimates previously described. To test hypothesis 1, we will combine the treatment conditions of hypothesis 2 and 4. The equation below describes our more saturated statistical model:

$$Y = \alpha + \tau \text{Neg} + \zeta \mathbf{X} + \epsilon \tag{1}$$

Where Y is a vector of the three outcomes, α is the intercept, τ is the treatment effect for the dummy of negative tweets, and is a matrix of coefficients for individual-level socio-demographic covariates, and ϵ is a vector of error terms.

To assess the hypothesis 2, we estimate the conditional average treatment effect for cognitive dissonance when tweets are neutral. The equation below describes our saturated statistical model:

$$Y = \alpha + \tau \text{Neutral} \cdot \text{D} + \beta_1 \text{Neutral} + \beta_2 \text{D} + \zeta \mathbf{X} + \epsilon \tag{2}$$

In equation 2, *Neutral* and D are dummy variables for the treatment conditions 1 and 3, and when the author of the tweet is dissonant to the respondent, respectively. Thus, the parameter τ measures the conditional treatment effect for cognitive dissonance when tweets are neutral. All the other variables are the same as equation 1.

To assess hypothesis 3, we use the same equation from hypothesis 2, but split the sample between respondents that are aligned with the government and with the opposition, and interact with the negative content from the treatment conditions from hypotheses 2 and 4. The equation goes below with all the parameters indexed by v to represent we will be perfoming the analysis conditional on political alignment of the voter.

$$Y_v = \alpha_v + \tau v \text{Neg} \cdot \text{D} + \beta_{v1}\text{Neg} + \beta_{v2}\text{D} + \zeta_v \mathbf{X} + \epsilon_v \tag{3}$$

For the conditional AMCE of hypothesis 2 and 3, we will measure cognitive dissonance using the three variables described on the hypotheses section. For the partisanship and vote choice varibles, we will match the respondents' choice and the author of the tweet, building a binary variables between dissonance and congruence. For the ideological placement variable, we will split the sample between the treatment conditions with Bolsonaro and Haddad, and interact the continuous variable measuring ideology with the dummy for negative tweets.

For all the models, we will perform subgroup analysis to capture the effect of attention, as in HT. We will use the natural log of the time the respondent took to read the treatment as our measure for attention. Finally, to assess our expectation about the effect of emotions, as in HE, we will report results from simple linear interactive models between each behavioral response with the treatment assignments, as well as make use of non-parametric techniques for mediation analysis (Imai et al., 2011).

Our primary analysis will be based on the pre-specified statistical model. However, we commit to present unadjusted estimates, and alternative regression specifications to increase robustness checks, and to attend requests by referees. We will make sure to indicate the primary model proposed in the PAP in the final report.

References

Adida, C. L., Dionne, K. Y., and Platas, M. R. (2018). Ebola, elections, and immigration: how politicizing an epidemic can shape public attitudes. *Politics, Groups, and Identities*, pages 1–27.

Aruguete, N. and Calvo, E. (2018). Time to #Protest: Selective Exposure, Cascading Activation, and Framing in Social Media. *Journal of Communication*, 68(3):480–502.

Banks, A., Calvo, E., and Karol, David, a. S. T. (2020). #polarizedfeeds:two experiments on polarization, framing, and social media. *Working Paper*.

Berinsky, A. J., Margolis, M. F., and Sances, M. W. (2014). Separating the shirkers from the workers? making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, 58(3):739–753.

Calvo, Ernesto, W. S. V. T. A. N. (2020). Winning! electoral adjudication and dialogue in social media. *Working Paper*.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4):765–789.

Lin, Y.-R., Keegan, B., Margolin, D., and Lazer, D. (2014). Rising tides or rising stars?: Dynamics of shared attention on Twitter during media events. *PloS one*, 9(5).

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public opinion quarterly*, 72(5):914–934.

Merrill, S., Grofman, B., and Adams, J. (2003). Assimilation and contrast effects in voter projections of party locations: Evidence from norway, france, and the usa. *European Journal of Political Research*, 40(2):199–221.

Wise, S. L. and Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2):163–183.