Can social backlash stifle free speech?*

Juan S. Morales† Margaret Samahita‡

Wave 3 – Pre-Analysis Plan

March 29, 2023

*All errors are our own.

†Department of Economics, Lazaridis School of Business and Economics, Wilfrid Laurier University. E-mail: jmorales@wlu.ca.

‡School of Economics and Geary Institute for Public Policy, University College Dublin. E-mail: margaret.samahita@ucd.ie.

1 Introduction

We study how perceived social pressure affects the public expression of opinion through a shift in publicly stated views towards a norm (*conformity*) or by inducing self-censorship (*silence*). In two earlier pre-registered online experiments (total N=1,650), we studied these social dynamics in the context of two often debated topics in the US: gender and race. We elicited participants' views on two divisive statements and their willingness to publish these views online (in an incentivized manner). Our main intervention, the "CancelCulture" treatment, exposed a random group of participants to a priming text informing them about "cancel culture" and examples of individuals who lost their jobs due to negative backlash over something they posted on social media.

Priming participants in our CancelCulture treatment led to modest effects in conformity to the norm, defined to be the majority opinion of those willing to publish their views.[1] However, participants also became more willing to publish their stated views. These results suggest that heightened awareness about cancel culture may increase individuals' conviction in their own opinions or their expected social reward from publicly expressing dissenting views. This "backlash" induced by our CancelCuture prime is strongest among Independents.

To better understand the mechanism(s) behind the above results, we propose a new study based on the earlier design with the following modifications:

- We introduce a new treatment, "WeakPrime", which exposes participants to the same text about online backlash but without explicit mention of "cancel culture". This will allow us to study whether the earlier backlash effect is due to the labeling of the description with the term "cancel culture", a prevalent narrative in conservative media which may have induced strong reactance among subjects.

- We introduce a second treatment dimension by informing participants how many others had decided to publish their opinion, varying this percentage to be either high or low. This will allow us to study whether willingness to publish exhibits strategic substitutability (subjects wanting to speak up when many others are unwilling to do so) or strategic complementarity (subjects wanting to speak up when many others are willing to do so).

[1] Perhaps unsurprisingly, empirically these norms coincided with the liberal/progressive views on both race and gender topics.

2 Design

The experiment timeline is shown in Figure 1. We start by collecting data on participants' demographics, risk attitude, political preferences and social media use. To elicit political preference, we ask: "In political matters, people talk of 'the left' and 'the right'. How would you place your views on this scale, generally speaking?". To elicit social media usage, we ask participants how much time per day they spend on Facebook, Twitter, Instagram and other social media platforms. We also ask how often on average participants post personal opinion on social media (never, less than once a month, a few (1-3) times a month, a few (1-6) times a week, at least once a day).

2.1 Attitude elicitation

Participants are next asked to consider two statements in random order:

- In my opinion, trans women should be allowed to participate in women's sports competitions.
- In my opinion, many people nowadays are too sensitive about things to do with race.

Participants are asked what they think of each statement, choosing from: strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, or strongly agree (coded as 1-7). Immediately after, they are asked "How important is the issue discussed in the statement to you?" and participants answer from 1 (not important at all) to 5 (extremely important).

2.2 Treatments

Following the demographic questionnaire and attitude elicitation, participants are randomized into one of 3×2 treatment groups, as shown in Figure 1. The first treatment dimension (Treatments 1-3) randomises the text shown to each participant. The Strong-Prime treatment is identical to our original intervention:

> The public nature of social media has resulted in individuals sometimes experiencing negative consequences as a result of their posts, in a phenomenon that some people refer to as **"cancel culture"**.

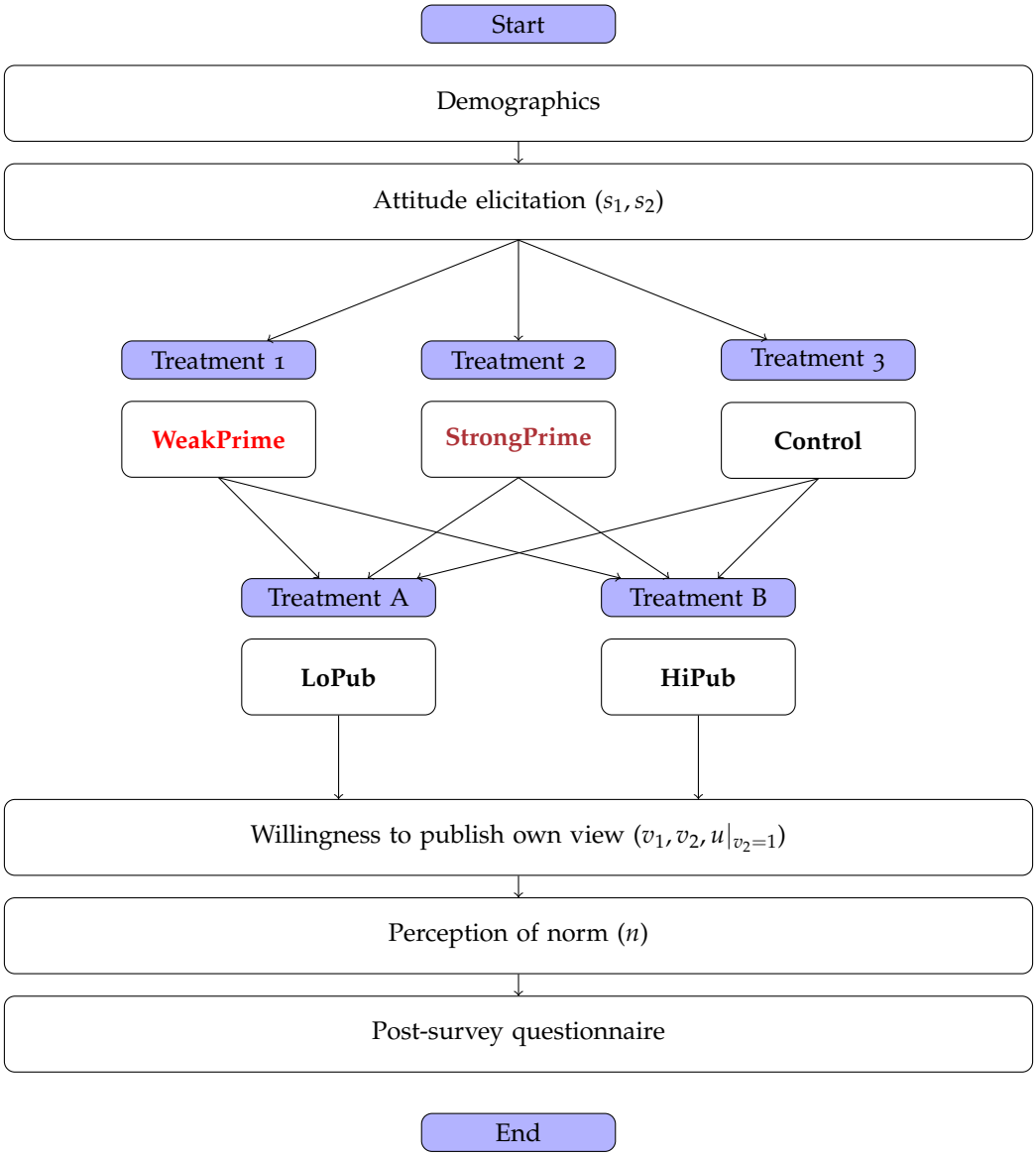
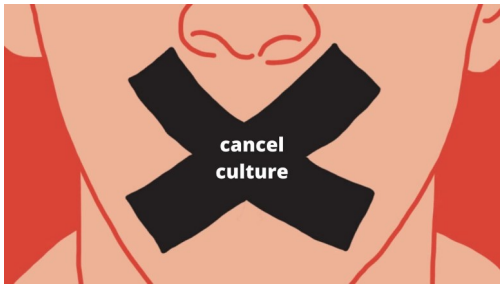


Figure 1: Experiment timeline

"those most vulnerable to harm tend to be **individuals previously unknown to the public**, like the communications director who was **fired** in 2013 after posting on social media, from her personal account, **an ill-thought-out joke** about Africa, AIDS and her own white privilege ... or the data analyst who was **fired** last spring after posting on social media, after the death of George Floyd in police custody, a study that suggested that riots depressed rather than increased Democratic Party votes"

These cases highlight the risk of **public backlash from social media.**



The WeakPrime treatment uses the same text while removing boldfont, image, and the text "in a phenomenon that some people refer to as **"cancel culture"**". We include an attention check after the text, asking: "To check that you are paying attention, what does the text say [cancel culture/social media backlash] can result in?" Participants select from the following alternatives: losing a job, lower voter turnout, or toppling a famous figure, and they cannot proceed unless they select "losing a job". In the Control treatment, participants are shown a text about University College Dublin (UCD) followed by an attention check about the text.

In the second treatment dimension, we vary the percentage figure shown to participants ostensibly representing how many of previous participants were willing to publish their opinion regarding the race or gender statements. The text is shown below the elicitation of willingness to publish, described in the next section, and reads:

> When we asked a previous group of participants, X% of them were willing to publish their opinion on the above statement with their name.

where X is determined by randomisation into either the LoPub or HiPub treatment. In LoPub, X equals [13/7] for the [gender/race] statement. In HiPub, X equals [60/67] for the [gender/race] treatment. These figures are taken from our previous experiments,

selecting US states with similarly high or low proportion of participants willing to speak up. Participants are debriefed with this information at the end of the experiment.[2]

2.3 Elicitation of willingness to publish

Following the T1-3 text, we next ask ask: "Would you be willing to let us post on social media, together with your name, your response to the previous statement:" (for example)

> *[Your name here]*
> *"**I somewhat disagree**" that many people nowadays are too sensitive about things to do with race."*

We inform participants the following:

- If you select Yes, we will create a tweet containing the above response and may post it on a public Twitter page created once data collection is complete (* see below)
- *We will contact Prolific to request your first and last names. Note that while in general Prolific does not allow researchers to collect personal information, Prolific does encourage researchers to get in touch in cases such as this, where the study design requires the collection of personal data (see https://researcher-help.prolific.co/hc/en-gb/articles/360015378834-Can-I-ask-Participants-for-their-Personal-Information-Identifiers-).
- The tweet will only contain a text of your name without any hyperlink, the public Twitter page will potentially contain the names and opinions of many participants.
- The link to the public Twitter page will be made available to participants who contact the researcher to ask for it, but it will not be otherwise advertised. The public Twitter page will be deleted after 30 days.

This is followed by the text about how many of previous participants were willing to publish their opinion, as described above.

Given that we study the effect of perceived social cost on stated opinion, it is crucial that participants seriously consider the possibility that their opinions will be shown to

[2]For LoPub, [13%/7%] of [Maryland/Indiana] participants were willing to publish their opinion about the [gender/race] statement. For HiPub, [60%/67%] of [Arizona/Oregon] participants were willing to publish their opinion about the [gender/race] statement.

others. However, actual publication with names is neither something we want to nor can do given the potential for negative consequences for the participants. Additionally, we seek to follow the standard of no deception in experimental economics. We therefore truthfully inform participants that, should they say Yes, we will attempt to obtain their names for the purpose of publication, but that publication is conditional on an event that (as we explain in the debrief) has an extremely low chance of happening. Previous inquiries about accessing participants' names from Prolific have, as we expected, been turned down.

2.3.1 Value of publishing

We also elicit a measure of participants' willingness to publish their stated view with their name. Due to time constraint, we only do this for one of the two statements. We therefore randomize participants into either a Race or Gender condition, which determines which of the two statements appears last and is used for the value of publishing. We first endow all participants with 10 tickets for a USD 100 bonus lottery. Participants are informed at the start that their chance of winning is approximately 1 in 2000. After participants are asked the above question on willingness to publish with their name, if they select "Yes, I would like to", they are then asked whether, in exchange for this post, they would be willing to give up 10, 5, or 1 of their lottery tickets. These questions are asked sequentially starting from the highest value. If/when they select Yes, they move on to the next section. If they do not ever select Yes, we code their willingness to pay (WTP) for publication as 0.

If participants state No to the question about publication with their name, they are then asked whether they would change their mind in exchange for a higher chance of winning the USD 100 lottery. We ask if they would be willing to let us post their response if we give them 1, 5, 25, or 50 additional lottery ticket. These questions are asked sequentially starting from the lowest value. If/when they select Yes, they move on to the next section. If they do not ever select Yes, we code their willingness to pay (WTP) for publication as -100.

Next, participants are asked about whether they would hypothetically publish their opinion on the above (second) statement on their own Twitter page in exchange for lottery tickets (and if so, how many). This is followed by a question on how much knowledge the participant has in the topic area (1 I have little to no knowledge to 5 I am an expert in this topic area).

2.4 Norm elicitation

We pre-registered defining the norm in our analyses as the *majority* opinion among those willing to publish with name. However, perception of this norm may differ across individuals. Therefore, we proceed by eliciting beliefs about others' stated opinion, denoted n_i. Again, due to time constraint, we only do this for the second of the two statements. After showing the statement, we ask participants:

- Considering ALL participants (in this US-based survey), what do you think **the majority** opinion is?
- Considering those participants (in this US-based survey) who stated that they WOULD be willing to let us post their opinion, together with their name, on social media (without any additional payment), what do you think **the majority** opinion is?

We incentivize participants by rewarding each correct answer with 5 additional lottery tickets for the USD 100 bonus.

2.5 Post-survey questionnaire

At the end of the study, we ask participants a number of questions: how many people they anticipate would see the public Twitter page, their reason for publishing any or none of the two opinions, and the below questions on perceived political correctness:

- "How often do you worry that things you post on social media can be misinterpreted?" (7-point scale, Never - Always)
- "The political climate these days prevents me from saying things I believe because others might find them offensive." (7-point scale, Strongly disagree - Strongly agree)
- "Are you worried about losing your job or missing out on job opportunities if your political opinions become known?" (4-point scale, Not at all worried - Worried a lot).
- "How often do you think social pressure causes people to misrepresent or lie about their political opinions on social media?" (7-point scale, Never - Always)

- "How often do you think social pressure causes people to refrain or abstain from expressing political opinions on social media?" (7-point scale, Never - Always)
- "How important is free speech to you?" (5-point scale, Not important at all - Extremely important)

This is followed by the Hong psychological reactance scale (Hong and Faedda, 1996), how common participants think their first and last name are (each on a 7-point scale, 1 Extremely uncommon - 7 Extremely common), and how likely they think the opinion of those willing to publish would indeed be posted on social media (1 Extremely unlikely - 7 Extremely likely). Finally participants can give feedback on the study using a text box.

2.6 Debrief

We end by debriefing participants about the purpose of the study and the sources of information we provided in the various treatments. We explicitly state that we do not anticipate publishing any participant's opinion with their name, even if they stated that they would like us to do this, since previous requests to Prolific had been turned down. Regardless, if the participant was willing to publish in exchange for lottery tickets, they would still get these additional tickets and the winner of the lottery would be paid after a few weeks. The full survey is provided in the Appendix.

2.7 Implementation

We use the data collection platform Prolific Academic and recruit 1500 participants, sampling 400 Democrats, 700 Independents (including Unaligned) and 400 Republicans. We over-sample Independents to verify our previous main finding of a backlash effect in this group. We randomise participants evenly into one of the six treatments.

3 Theory and Hypotheses

The utility of speaking up is given by:

$$u_i = -[\alpha(s_i - o_i)^2 + \beta(s_i - n)^2] + \kappa(\beta, v_{-i})(s_i - n)^2$$

where $s_i \in [0,1]$ denotes the individual's public stance, $o_i \in [0,1]$ denotes their private opinion, and $n \in [0,1]$ represents the social norm or appropriate public stance. β is

the cost or risk from social disapproval and α is a "cognitive dissonance" cost from expressing a public stance which differs from the individual's private opinion.

Individuals get a social reward for speaking up $\kappa(\beta, v_{-i})(s_i - n)^2$ which increases with the distance from the norm and is a function of β and v_{-i}, the proportion of others who speak up.[3] Desirable properties for $\kappa(\beta, v_{-i})$:

- $$\frac{\partial u_i}{\partial \beta} = (s_i - n)^2 \left(\frac{\partial \kappa(\beta, v_{-i})}{\partial \beta} - 1 \right)$$

  should be negative for low β (silencing in WeakPrime) and positive for high β (backlash effect in StrongPrime)

- $$\frac{\partial u_i}{\partial v_{-i}} = \frac{\partial \kappa(\beta, v_{-i})}{\partial v_{-i}} (s_i - n)^2$$

  One possibility is that this value is positive for low β (complementarity) and negative for high β (substitution). Another is that this value is uniformly positive.

Both of these effects are amplified by the distance from the norm, $(s_i - n)^2$. One possible function which takes into account the interaction between β and v_{-i} is, e.g., $\kappa(\beta, v_{-i}) = c_1\beta^2 - c_2 v_{-i}\beta + c_3 v_{-i}$, with $c_1, c_2, c_3 > 0$. If we assume that the effect of v_{-i} is uniformly positive, a possible alternative is $\kappa(\beta, v_{-i}) = c_1\beta^2 - c_2\beta + c_3 v_{-i}$.

With the above model, and with some restrictions on the parameters c_1, c_2, c_3, we can hypothesise the following:

Given that the WeakPrime increases the perceived cost of social disapproval, we hypothesise that:

Hypothesis 1a. *Individuals are less willing to publish their opinion when they are exposed to the WeakPrime than if they are not primed.*

The StrongPrime treatment explicitly talks about cancel culture, an issue which we hypothesise (based on our previous studies) to generate a backlash effect as individuals perceive an even greater need (or use it as an excuse) to voice their opinion publicly. Thus:

Hypothesis 1b. *Individuals are more willing to publish their opinion when they are exposed to the StrongPrime than if they are not primed.*

Consequently,

[3] κ may be heterogeneous (and assume negative values) depending on e.g. individual privacy concerns.

Hypothesis 1c. *Individuals are more willing to publish their opinion when they are exposed to the StrongPrime than the WeakPrime.*

Our second treatment dimension varies whether a high or low proportion of previous participants choose to speak up. Related to the literature on protest participation, participants may consider speaking up on a controversial topic to be risky and only do so when many others do (strategic complementarity).

Hypothesis 2. *In the Control and WeakPrime conditions, individuals are more willing to publish when informed that the proportion of previous participants who publish is high rather than low.*

However, in the StrongPrime treatment, in which we expect a backlash effect (due to the perceived negative consequences of cancel culture on free speech), this same mechanism may cause an increase in speaking up when the proportion of others who speak up is low rather than high (strategic substitution). Alternatively, strategic complementarity may still hold even in this condition. The overall effect is ambiguous and it is an open empirical question which we will attempt to answer in our experiment.

4 Analyses

4.1 Main analysis

Our main outcome in this section is individuals' willingness to publish their opinion with their name. We define v_i as a binary variable which takes value 1 if subject i is willing to publish their opinion (without additional lottery tickets as incentive).

To test Hypothesis 1, we estimate the following regression:

$$v_{iq} = \alpha + \beta_1 WeakPrime_i + \beta_2 StrongPrime_i + \delta_q + \varepsilon_{iq}$$

where $WeakPrime_i$ and $StrongPrime_i$ are the relevant treatment dummies. We include topic fixed effects δ_q. We hypothesise that $\beta_1 < 0$ and that $\beta_2 > 0$.

To test hypothesis 2, we interact the WeakPrime and StrongPrime treatments with HiPub to estimate the following regression:

$$\begin{aligned} v_{iq} = \alpha &+ \beta_1 WeakPrime_i + \beta_2 StrongPrime_i + \beta_3 HiPub_i \\ &+ \beta_4 WeakPrime_i \times HiPub_i + \beta_5 StrongPrime_i \times HiPub_i + \delta_q + \varepsilon_{iq} \end{aligned}$$

where $WeakPrime_i$, $StrongPrime_i$ and $HiPub_i$ are the relevant treatment dummies. We hypothesise that $\beta_3 > 0$ and $\beta_4 > 0$. The sign of β_5 is theoretically ambiguous.

In some specification(s) we will include a vector of controls $\mathbf{X_i}$ described below, which may increase the precision of our estimates (but should be orthogonal to our treatment since it is randomized). In all specifications we use robust standard errors (clustered at the individual level).

4.2 Complementary analyses

Other outcomes

We use the number of lottery tickets the participant is willing to pay/accept for publication as another outcome variable. We define $lottery_i$ to be the value at which the participant responds Yes to publishing. For example, if they are willing to pay 10 tickets, $lottery_i = 10$, while if they are willing to accept 5 tickets, $lottery_i = -5$. Note that these values are potentially the lower bound: a participant who answers Yes to paying 10 tickets may have been willing to pay a higher amount. For those unwilling to accept 50 tickets, we impute a value of -100 for the analysis. In addition, we will define categorical variables for each of these groups.

Additionally, we will use the participant's hypothetical willingness to publish their views on their own social media account as an outcome.

Heterogeneity by distance to the norm

As in our previous experiments, we will interact our treatments (WeakPrime, StrongPrime, HiPub) with participants' distance to the norm ($|v_i - n|$). As in previous waves, we define the norm to coincide with "left-wing" views (7 for the gender question, and 1 for the race question). We will also use alternative norm definitions as robustness tests, including our previously defined norms based on the majority opinion of those participants willing to speak up, as well as local and perceived norms.

4.3 Exploratory analyses

- Heterogeneity in various dimensions, including: political affiliation, local norm (Republican vs Democratic state), social media use, age, education, psychological reactance. In particular, based on previous waves we expect that complementarity effects would be strongest for Democrats, priming effects would be strongest

for Independents, and welfare effects (heterogeneity by importance) would be strongest for Republicans.

- Norms: perceived norms of those willing/unwilling to publish will allow us to examine the channels through which complementarity/substitutability works. For example, if willingness to publish is higher in the LoPub than HiPub condition, is it because subjects believe the silent majority agree with them, or do they want to make a stand together with the publishing minority against the opposing majority?
- Welfare analyses using topic importance/expertise for those willing/unwilling to publish.
- How treatments change distribution of published opinion.

4.4 Control variables

Our baseline specification includes:

- Age: coded continuously
- Gender: coded as a dummy for Man, Woman, Non-binary/Other ("Prefer not to say" as the omitted category)
- Race: coded as a series of dummies for White, Hispanic or Latino, Black of African American, Native American or American Indian, Asian/Pacific Islander ("Other" as the omitted category)
- Education: coded as a dummy for having at least a 2-year college degree
- Employment: coded as a dummy
- Risk attitude (Falk et al., 2018): coded on a 0-10 Likert scale and standardised
- Political leaning: coded on a 0-10 Likert scale and standardised
- Social media use: coded as a dummy for spending more than 60 minutes daily on social media OR frequency of posting on social media
- Topic fixed effect
- Stated opinion

We may also control for the following variables which are elicited post-treatment:

- Perceived audience size
- Concerns about free speech and social pressure
- Response to the psychological reactance scale
- Perceived popularity of own name
- Perceived likelihood of publication

4.5 Robustness checks

We will check the robustness of our results to:

- Dropping subjects who do not answer the attention check correctly in the first attempt. We will also check whether the proportion of subjects correctly answering in the first attempt is significantly different depending on the first prime shown.
- We will check for selective attrition from those who withdraw after study is done.

References

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global evidence on economic preferences." *The Quarterly Journal of Economics*, 133(4): 1645–1692.

Hong, Sung-Mook, and Salvatora Faedda. 1996. "Refinement of the Hong psychological reactance scale." *Educational and Psychological Measurement*, 56(1): 173–182.

Appendices

A Full Survey

Begins on the next page.

Introductory Statement

This study is conducted by Dr Margaret Samahita from the School of Economics, University College Dublin and Dr Juan S. Morales from the Department of Economics, Wilfrid Laurier University.

What is this research about?

This study is part of a research project to study the opinions of Americans and will include demographic questions, as well questions about media use and current topics, among others. As is standard in online research studies, you will not be informed about the specific hypothesis tested or the methodology used. However, once the study is completed, you can get in touch with the researchers below to find out more.

Why have you been invited to take part?

You have been invited to take part since you meet the research requirement: you are an adult aged over 18 years living in the US.

How will your data be used?

Unless otherwise noted, your data will be analysed and aggregate results will be reported in a future research paper for publication in an academic journal. The data will be stored indefinitely. As per the publication policy of most economics journals, upon publication data will need to be made available for viewing or use by future researchers.

What will happen if you decide to take part in this research study?

You will fill out a 15 minute survey through Prolific using your desktop computer.

How will your privacy be protected?

We will collect your Prolific participant ID as is standard procedure, ensuring the data is anonymous.

What are the benefits of taking part in this research study?

Your responses will help researchers better understand the opinions of Americans and how these are formed. You will be paid a participation fee as is standard on Prolific. You will also have the possibility of earning an additional $100 bonus payment through a lottery. You start this survey with 10 tickets and your chance of winning is approximately 1 in 2000.

What are the risks of taking part in this research study?

There are no foreseeable risks to taking part in this study beyond that arising from everyday activities such as browsing online content. You may be asked to consider statements pertaining to events and situations that may evoke a range of emotional responses in different people. However, if you have any concern and wish to withdraw at any point, simply close the survey window.

Can you change your mind at any stage and withdraw from the study?

Yes, if you wish to withdraw at any point, simply close the survey window. Your data will not be used if you choose not to complete the study.

How will you find out what happens with this project?

Future updates to the project will be available by contacting the researcher. Researcher contact details for further information

margaret.samahita@ucd.ie

jmorales@wlu.ca

This project has been reviewed and approved by the University Research Ethics Board at the University College Dublin (REB# HS-22-50-Samahita) and at Wilfrid Laurier University (REB #8354).

REB contact details for further information

research.ethics@ucd.ie

REB@wlu.ca

It is advised that you print or save this consent information and/or record the researcher contact information in case that you have any questions or concerns.

If you consent to the above information sheet, please select Yes below.

I have read and understood the above and want to participate in this study.

o Yes

o No

What is your Prolific ID? Please note that this response should auto-fill with the correct ID

_____

DEMOGRAPHICS

What is your age (in years)?

_____

What is your gender?

o Man

o Woman

o Non-binary/Other _____

o Prefer not to say

Please specify your ethnicity.

o White

o Hispanic or Latino

- o Black or African American
- o Native American or American Indian
- o Asian / Pacific Islander
- o Other _____

In which state do you currently reside?

What is the highest level of school you have completed or the highest degree you have received?

- o Less than high school degree
- o High school graduate (high school diploma or equivalent including GED)
- o Some college but no degree
- o Associate degree in college (2-year)
- o Bachelor's degree in college (4-year)
- o Master's degree
- o Doctoral degree
- o Professional degree (JD, MD)

Which statement best describes your current employment status?

- o Working (paid employee)
- o Working (self-employed)
- o Not working (temporary layoff from a job)
- o Not working (looking for work)
- o Not working (retired)
- o Not working (disabled)
- o Not working (other) _____
- o Prefer not to answer

Please tell us, in general, how willing or unwilling you are to take risks.

- o 0 Completely unwilling to take risks
- o 1
- o 2

- o 3
- o 4
- o 5
- o 6
- o 7
- o 8
- o 9
- o 10 Very willing to take risks

In political matters, people talk of ‘the left’ and ‘the right’. How would you place your views on this scale, generally speaking?

- o 0 The Left
- o 1
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7
- o 8
- o 9
- o 10 The Right

How much time per day do you spend... [never/no account, less than 30 minutesfrom 30 minutes to 1 hour, from 1 hour to 2 hours, more than 2 hours]

On social media (Facebook, Twitter, Instagram, Tik Tok, Snapchat, etc)

Watching, reading or listening to news about politics and current affairs

How often on average do you post personal opinions on social media?

- o Never
- o Less than once a month
- o A few (1-3) times a month

- o A few (1-6) times a week
- o At least once a day

ATTITUDE ELICITATION

[Gender and Race statements are shown in random order]

You will now be asked to state your opinion on a number of questions.

Please consider the following statement.

In my opinion, trans women should be allowed to participate in women's sports competitions.

What do you think about the above statement?

- o Strongly disagree
- o Disagree
- o Somewhat disagree
- o Neither agree nor disagree
- o Somewhat agree
- o Agree
- o Strongly agree

How important is the issue discussed in the statement to you?

- o 1 Not important at all
- o 2
- o 3
- o 4
- o 5 Extremely important

Please consider the following statement.

In my opinion, many people nowadays are too sensitive about things to do with race.

What do you think about the above statement?

- o Strongly disagree
- o Disagree
- o Somewhat disagree
- o Neither agree nor disagree
- o Somewhat agree
- o Agree
- o Strongly agree

How important is the issue discussed in the statement to you?

- o 1 Not important at all
- o 2
- o 3
- o 4
- o 5 Extremely important

TEXT

[Subjects are shown one of: WeakPrime, StrongPrime or Control texts]

[WeakPrime]

Please read the following text.

The public nature of social media has resulted in individuals sometimes experiencing negative consequences as a result of their posts.

"Those most vulnerable to harm tend to be individuals previously unknown to the public, like the communications director who was fired in 2013 after posting on social media, from her personal account, an ill-thought-out joke about Africa, AIDS and her own white privilege ... or the data analyst who was fired last spring after posting on social media, after the death of George Floyd in police custody, a study that suggested that riots depressed rather than increased Democratic Party votes."

These cases highlight the risk of public backlash from social media.

To check that you are paying attention, what does the text say social media backlash can result in?

- o losing a job
- o lower voter turnout
- o toppling a famous figure

[StrongPrime]

Please read the following text.

The public nature of social media has resulted in individuals sometimes experiencing negative consequences as a result of their posts, in a phenomenon that some people refer to as **"cancel culture"**.

“Those most vulnerable to harm tend to be **individuals previously unknown to the public**, like the communications director who was **fired** in 2013 after posting on social media, from her personal account, **an ill-thought-out joke** about Africa, AIDS and her own white privilege ... or the data analyst who was **fired** last spring after posting on social media, after the death of George Floyd in police custody, a study that suggested that riots depressed rather than increased Democratic Party votes.”

These cases highlight the risk of **public backlash from social media**.

To check that you are paying attention, what does the text say cancel culture can result in?

- o losing a job
- o lower voter turnout
- o toppling a famous figure

[Control]

Please read the following text.

University College Dublin (commonly referred to as UCD) is a research university in Dublin, Ireland, and a member institution of the National University of Ireland. With 33,284 students, it is Ireland's largest university. Five Nobel Laureates are among UCD's alumni and current and former staff. UCD's main campus is located on a 133-hectare (330-acre) campus at Belfield, four kilometres to the south of the city centre. In 1991, it purchased a second site in Blackrock. This currently houses the Michael Smurfit Graduate Business School.

A report published in May 2015 showed the economic output generated by UCD and its students in Ireland amounted to €1.3 billion annually.

To check that you are paying attention, where does the text say UCD's main campus is located?

o Smurfit

o Belfield

o Blackrock

WILLINGNESS TO PUBLISH

[The order of topics follows that of the attitude elicitation section]

The next set of questions are especially important for our study, so please consider your answers carefully.

Would you be willing to let us post on social media, together with your name, your response to the previous statement:

[Your name here]

*"**I [subject's response from above, strongly disagree – strongly agree]** that trans women should be allowed to participate in women's sports competitions."*

-If you select Yes, **we will create a tweet** containing the above response and may post it on a public Twitter page created once data collection is complete (* see below)

-***We will contact Prolific to request your first and last names.** Note that while in general Prolific does not allow researchers to collect personal information, Prolific does encourage researchers to get in touch in cases such as this, where the study design requires the collection of personal data (see https://researcher-help.prolific.co/hc/en-gb/articles/360015378834-Can-I-ask-Participants-for-their-Personal-Information-Identifiers-).

-The tweet will only contain **a text of your name without any hyperlink**, the public Twitter page will potentially contain the names and opinions of many participants.

-The link to the public Twitter page will be **made available to participants** who contact the researcher to ask for it, but it will not be otherwise advertised. The public Twitter page will be **deleted after 30 days**.

When we asked a previous group of participants, [13/60]% of them were willing to publish their opinion on the above statement with their name.

Would you be willing to let us post your response above?

o Yes, I would like to

o No, I'd rather not

You will now be asked the same question regarding the **other** statement you were asked about.

Would you be willing to let us post on social media, together with your name, your response to the previous statement:

[Your name here]

*"**I [subject's response from above, strongly disagree – strongly agree]** that many people nowadays are too sensitive about things to do with race."*

-If you select Yes, **we will create a tweet** containing the above response and may post it on a public Twitter page created once data collection is complete (* see below)

-***We will contact Prolific to request your first and last names.** Note that while in general Prolific does not allow researchers to collect personal information, Prolific does encourage researchers to get in touch in cases such as this, where the study design requires the collection of personal data (see https://researcher-help.prolific.co/hc/en-gb/articles/360015378834-Can-I-ask-Participants-for-their-Personal-Information-Identifiers-).

-The tweet will only contain **a text of your name without any hyperlink**, the public Twitter page will potentially contain the names and opinions of many participants.

-The link to the public Twitter page will be **made available to participants** who contact the researcher to ask for it, but it will not be otherwise advertised. The public Twitter page will be **deleted after 30 days**.

When we asked a previous group of participants, [7/67]% of them were willing to publish their opinion on the above statement with their name.

Would you be willing to let us post your response above?

o Yes, I would like to

o No, I'd rather not

[If subject selects Yes above]

You stated that you would like us to post on social media, together with your name, your response to the previous statement:

[Your name here]

*"**I [subject's response from above, strongly disagree – strongly agree]** that many people nowadays are too sensitive about things to do with race."*

In exchange for this post, we want to know if you would be willing to **give up some of your lottery tickets** for the $100 bonus (remember that you start with 10 tickets).

*Would you be willing to give up **all [10, 5, 1 asked sequentially, stopping when subject selects Yes] lottery tickets** in exchange for this public post?*

o Yes

o No

If you select Yes,

-**We will contact Prolific to request your first and last names.** Note that while in general Prolific does not allow researchers to collect personal information, Prolific does encourage researchers to get in touch in cases such as this, where the study design requires the collection of personal data (see https://researcher-help.prolific.co/hc/en-gb/articles/360015378834-Can-I-ask-Participants-for-their-Personal-Information-Identifiers-).

-Note, we will only reduce your lottery tickets if we do publish the above text with your name.

[If subject selects No above]

You would rather not let us post on social media, together with your name, your response to the previous statement:

[Your name here]

*"**I [subject's response from above, strongly disagree – strongly agree]** that many people nowadays are too sensitive about things to do with race."*

We would now like to ask whether you would be willing to **change your mind** in exchange for a **higher chance of winning the $100 lottery**. Remember that you start with 10 tickets.

*Would you be willing to let us post the above if we give you **[1, 5, 20, 50 asked sequentially, stopping when subject selects Yes] additional lottery ticket?***

o Yes

o No

If you select Yes,

-You will get 1 additional ticket in the lottery.

-**We will contact Prolific to request your first and last names.** Note that while in general Prolific does not allow researchers to collect personal information, Prolific does encourage researchers to get in touch in cases such as this, where the study design requires the collection of personal data (see https://researcher-help.prolific.co/hc/en-gb/articles/360015378834-Can-I-ask-Participants-for-their-Personal-Information-Identifiers-).

Hypothetically, would you post your response to the previous statement on your Twitter account in exchange for lottery tickets (if yes, please state how many)::

---

*"**I [subject's response from above, strongly disagree – strongly agree]** that many people nowadays are too sensitive about things to do with race."*

o No

o Yes. How many? (If you are willing to post your response for free, please write 0) _____

Please consider the following statement.

Many people nowadays are too sensitive about things to do with race.

How much knowledge would you say you have in this topic area?

o 1 I have little to no knowledge

o 2

o 3

o 4

o 5 I am an expert in this topic area

PERCEPTION OF NORM

As earlier mentioned, you have the chance to win an additional bonus of $100 through a lottery.

You will now see **2 questions**. You will earn **5 additional lottery tickets** for each question you answer correctly, in addition to your existing tickets.

Therefore, please consider your answers carefully since each correct answer will increase your chance of winning the $100 bonus.

Remember, you will earn 5 additional lottery tickets for each correct answer, so please consider your answers carefully.

Please consider the following statement.

In my opinion, many people nowadays are too sensitive about things to do with race.

Considering ALL participants (in this US-based survey), what do you think the **majority** opinion is?

- o Strongly disagree
- o Disagree
- o Somewhat disagree
- o Neither agree nor disagree
- o Somewhat agree
- o Agree
- o Strongly agree

Considering those participants (in this US-based survey) who stated that they WOULD be willing to let us post their opinion, together with their name, on social media (without any additional payment), what do you think the **majority** opinion is?

- o Strongly disagree
- o Disagree
- o Somewhat disagree
- o Neither agree nor disagree
- o Somewhat agree
- o Agree
- o Strongly agree

POST-SURVEY QUESTIONNAIRE

If it was published, how many people do you think would see the tweet containing your response with your name on our Twitter page?

_____

Why did you decide **[to let us post one or more/not to let us post any]** of your opinions on social media?

_____

_____

_____

_____

_____

How often do you worry that things you post on social media can be misinterpreted?

o 1 Never

o 2

o 3

o 4

o 5

o 6

o 7 Always

The political climate these days prevents me from saying things I believe because others might find them offensive.

o Strongly disagree

o Disagree

o Somewhat disagree

o Neither agree nor disagree

o Somewhat agree

o Agree

o Strongly agree

Are you worried about losing your job or missing out on job opportunities if your political opinions become known?

o Not at all worried

o Not very worried

- o Worried a little
- o Worried a lot

How often do you think social pressure causes people to **misrepresent or lie** about their political opinions on social media?

- o 1 Never
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7 Always

How often do you think social pressure causes people to **refrain or abstain from expressing** political opinions on social media?

- o 1 Never
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7 Always

How important is free speech to you?

- o 1 Not important at all
- o 2
- o 3
- o 4
- o 5 Extremely important

Please indicate how much you agree or disagree with the following statements. [Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree]

Regulations trigger a sense of resistance in me.

I find contradicting others stimulating.

When something is prohibited, I usually think, “That’s exactly what I am going to do.”

The thought of being dependent on others aggravates me.

I consider advice from others to be an intrusion.

I become frustrated when I am unable to make free and independent decisions.

It irritates me when someone points out things which are obvious to me.

I become angry when my freedom of choice is restricted.

Advice and recommendations usually induce me to do just the opposite.

I am contented only when I am acting of my own free will.

I resist the attempts of others to influence me.

It makes me angry when another person is held up as a role model for me to follow.

When someone forces me to do something, I feel like doing the opposite.

It disappoints me to see others submitting to society’s standards and rules.

The most common first name in the US is James. How common do you think your **first name** is in the US?

- o Extremely uncommon
- o Uncommon
- o Somewhat uncommon
- o Average
- o Somewhat common
- o Common
- o Extremely common

The most common last name in the US is Smith. How common do you think your **last name** is in the US?

- o Extremely uncommon
- o Uncommon
- o Somewhat uncommon
- o Average
- o Somewhat common

- o Common
- o Extremely common

For participants who agreed to let us publish their opinion together with their name, how likely do you think it is that the response will end up being posted on social media?

- o Extremely unlikely
- o Unlikely
- o Somewhat unlikely
- o Neutral
- o Somewhat likely
- o Likely
- o Extremely likely

Do you have any feedback or comments on the study? Please let us know if anything was unclear.

_____

_____

_____

_____

_____

DEBRIEF

Thank you for participating in our study.

This study aims to investigate the impact of cancel culture on self-expression. We are interested in how willing you would be to let us post your opinion on social media. As data collection is ongoing, we would like to ask you not to talk about this study with others for now.

You were shown one of the following two texts:

- The text about UCD was modified from https://en.wikipedia.org/wiki/University_College_Dublin and serves as a filler.
- The text about online backlash was modified from https://www.nytimes.com/2020/12/03/t-magazine/cancel-culture-history.html

In a previous survey,

- 13% of Maryland participants were willing to publish their opinion about the gender statement
- 7% of Indiana participants were willing to publish their opinion about the race statement
- 60% of Arizona participants were willing to publish their opinion about the gender statement
- 67% of Oregon participants were willing to publish their opinion about the race statement

Regarding the publication of your opinion on social media:

- Previous requests to Prolific asking for participant's names in a similar study design have been turned down; so we do not anticipate that we will publish your opinion with your name, even if you stated that you were willing to let us to do this. Regardless, if you stated that you were willing to publish the opinion with your name in exchange for lottery tickets, you will still get these additional lottery tickets.

If you wish to remove your data from the study, please contact us through your Prolific account.

If you win the bonus payment, it will be paid through Prolific in the next few weeks.

If you have any questions about the study, please feel free to contact Margaret Samahita (margaret.samahita@ucd.ie).

This project has been reviewed and approved by the University Research Ethics Board at the University College Dublin (REB# HS-22-50-Samahita) and at Wilfrid Laurier University (REB #8354)

REB contact details for further information

research.ethics@ucd.ie

REB@wlu.ca

Please click the button below to be redirected back to Prolific and register your submission. In case needed, the completion code is 86F1B35E.