UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2021

Can ♥s Change Minds?
Social Media Endorsements and Policy Preferences

Pierluigi Conzo, Juan S Morales and Andrea Gallice, University of Turin
Laura K Taylor and Margaret Samahita, University College Dublin

WP21/04

February 2021

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Can ❤s Change Minds? Social Media Endorsements and Policy Preferences*

Pierluigi Conzo[†] ⓡ Laura K. Taylor[‡] ⓡ Juan S. Morales[§] ⓡ
Margaret Samahita[¶] ⓡ Andrea Gallice[||]

February 4, 2021

Abstract

We investigate the effect of social media endorsements (likes, retweets, shares) on individuals' policy preferences. In two online controlled experiments (N=1,384), we exposed participants to non-neutral policy messages about the COVID-19 pandemic (emphasizing either public health or economic activity as a policy priority) while varying the level of endorsements of these messages. Our experimental treatment significantly shifted the policy views of active social media users by about 0.12 standard deviations. The treatment effect for these users is heterogeneous depending on their pre-existing views. Specifically, message endorsements reinforce pre-existing attitudes, thereby increasing opinion polarization. The effect appears concentrated on a minority of individuals who correctly answered a factual manipulation check regarding the endorsement metrics. This evidence suggests that though only a fraction of individuals pay conscious attention to these metrics, they may be easily influenced by these social cues.

JEL Codes: D83; L82; L86; O33
Keywords: social media; social conformity; political polarization; COVID-19

*We thank Anna Dreber Almenberg, Karen Arulsamy, Stefan Müller and Johannes Wohlfart as well as seminar participants at UCD and the Collegio Carlo Alberto for comments. Funding from UCD, Collegio Carlo Alberto and the Einaudi Institute for Economics and Finance is gratefully acknowledged. The experiments were approved by University College Dublin Office of Research Ethics (reference numbers: HS-E-20-110-Samahita and HS-E-20-134-Samahita). Authors' order has been randomized using the AEA Author Randomization Tool (reference: eTjbUATa_zKY), denoted by ⓡ. All errors are our own.

†Department of Economics and Statistics, University of Turin and Collegio Carlo Alberto
‡School of Psychology and Geary Institute for Public Policy, University College Dublin
§Collegio Carlo Alberto and ESOMAS Department, University of Turin
¶School of Economics and Geary Institute for Public Policy, University College Dublin
||ESOMAS Department, University of Turin and Collegio Carlo Alberto

1 Introduction

Social media has been hypothesized to have broad effects on politics (Zhuravskaya, Petrova and Enikolopov, 2020). However, the magnitude of these effects, and the mechanisms through which they arise, remain debated. This article studies how social media affects individuals' policy preferences. In particular, we study endorsements, a central feature of social media, as evinced by common metrics of engagement: *likes*, ♥s, 👍s, *retweets*, and *shares*. Can the perceived support of social media messages affect how individuals evaluate policies?

To answer this question, we conducted two pre-registered online experimental studies in Europe (Ireland, N=305, and Italy, N=300) and the US (N=779) in the context of the COVID-19 pandemic and its policy trade-offs (public health vs. economic activity). While the importance of these trade-offs were highly debated, COVID-19 was a prevalent issue in 2020 and presents a good test case to address our research question. The experiment allows us to isolate the effects of perceived support for policy choices in a controlled environment different from individuals' own social media. We therefore study endorsements, a specific feature of social media, without conflating issues of social image, peer effects, or selective exposure. Instead, we exposed individuals to strangers' tweets and endorsements, and examined their effects on individuals' policy preferences in an anonymous survey. More specifically, we exposed participants to non-neutral policy messages about the COVID-19 pandemic, manipulated the perceived level of endorsements of these messages, and examined how this affected their policy attitudes.

Our study reveals that perceived endorsements can affect policy preferences, but only among a specific subgroup of the population. In particular, our experimental treatment shifts the policy views of active social media users —defined and pre-registered as those who use Facebook or Twitter for one hour or more each day— by about 0.12 standard deviations. In addition, treatment effect for these users is heterogeneous depending on their pre-existing views. Specifically, message endorsements reinforce pre-existing attitudes, thereby increasing opinion polarization. The effect appears concentrated on a minority of individuals who correctly answered a factual manipulation check regarding the endorsement metrics, about 10 percent of our survey respondents.

Studies have emphasized how social media exposes individuals to echo-chambers of predominantly like-minded information (Bakshy, Messing and Adamic, 2015; Barberá, 2015; Halberstam and Knight, 2016; Peterson and Kagalwala, 2021) and thereby amplifies political polarization (Settle, 2018; Sunstein, 2018; Levy, Forthcoming; Allcott et al., 2020*b*). Media concerns about the influence of social media on elections are also common,[1] yet many contend that these concerns may be overblown (Gentzkow and Shapiro, 2011; Boxell, Gentzkow and Shapiro, 2017; Allcott and Gentzkow, 2017; Eady et al., 2019; Guess, Nyhan and Reifler, 2020; Guess, Forthcoming; Scharkow et al., 2020). As a way to sharpen our understanding of these issues, we propose studying precise mechanisms —motivated by previous work in the social sciences— through which social media may affect political dynamics.

A series of recent studies have documented that information can shift individual's political attitudes (Grigorieff, Roth and Ubfal, 2020; Haaland and Roth, 2020; Alesina, Stantcheva and Teso, 2018). Social pressure is also known to shape behaviour and views (Cialdini and Goldstein, 2004; Bursztyn and Jensen, 2017; Carlson and Settle, 2016) and is an important channel through which social media may affect policy preferences, especially in situations of evolving public opinion (Bursztyn, Egorov and Fiorin, 2020). More specifically, online social endorsements and perceptions of support have been shown to affect whether individuals select to read content (Messing and Westwood, 2014), *like* messages (Egebark and Ekström, 2018), or self-report voting (Bond et al., 2017), and can have broader implications for online political dissent (Morales, 2020). By examining how the perceived endorsements attached to social media messages affect policy attitudes, our study reveals one important mechanism through which social media affects politics.

2 Experimental design

Our first survey (N=605) was conducted using nationally representative samples in Ireland and Italy. The survey was sent out by the data collection company Dynata on 8 July 2020. Our second survey was conducted in the US, representative in terms of age, gender and census regions (N=1,519). The second survey was sent out by Dynata on 31 July 2020. The main analyses presented below pool the two surveys, while in the supplementary materials we show that our main results are quantitatively similar when analysing the samples separately.[2] The studies were respectively pre-registered on the AsPredicted and the AEA RCT Registry platforms and we follow these analysis plans unless otherwise noted.[3]

[1]See for instance: https://www.nytimes.com/2020/03/29/technology/russia-troll-farm-election.html and https://www.scientificamerican.com/article/how-twitter-bots-help-fuel-political-feuds/

[2]Supplementary materials are available at `https://doi.org/10.1257/rct.6254-4.0`.

[3]Available at `https://aspredicted.org/blind.php?x=5t367e` and `https://www.socialscienceregistry.org/trials/6254`.

We first measured participants' pre-treatment attitudes using statements about COVID-19 policy responses, including *"The government's highest priority should be saving as many lives as possible even if it means the economy will recover more slowly."* and *"Sweden's government has so far avoided implementing a lockdown in order to keep the economy going. What do you think of this policy?"*, among others.[4] Participants indicated their agreement to these statements on a 1-7 Likert scale. We standardized these responses and coded positive values as being pro-economy. In addition, we combined the questions into one index through principal component analysis.

We next randomised participants into one of three treatments: control, pro-economy or pro-health.[5] In each treatment, participants are shown six tweets about COVID-19 policies, of which three are pro-economy and three are pro-health. In the control condition, all tweets have low endorsements (a low number of likes and retweets). In the pro-economy (pro-health) condition, the three pro-economy tweets are given high (low) endorsements while the three pro-health tweets are given low (high) endorsements. The tweets were preceded by the following text: *"The algorithms used on social media may sometimes present you with posts by complete strangers. You will now be shown 6 tweets. As if you were going through your own social media feed (eg Twitter or Facebook), please consider whether you would "like" or "retweet" each of the following 6 tweets."*

Figure 1 shows an example of the experimental variation. The tweets were generated using https://www.tweetgen.com/ using the following input:

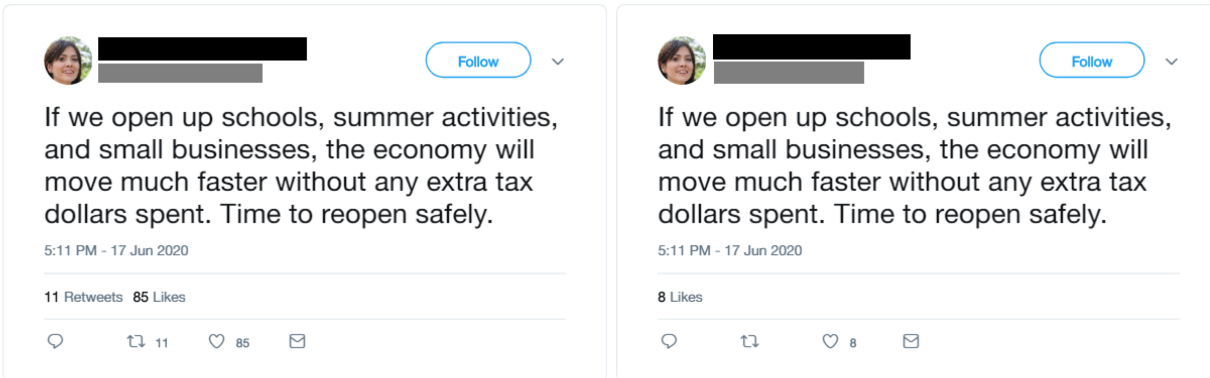
- **Text:** We ran a search of COVID-19 related tweets on Twitter and selected six tweet messages, three pro-health and three pro-economy.
- **Metrics:** "Low" endorsement tweets have between 0-10 likes and 0-1 retweets. "High" endorsement tweets have between 50-100 likes and 10-20 retweets.
- **User:** The profile pictures are generated by an algorithm using the website https://thispersondoesnotexist.com/. No username is shown.[6]
- **Time:** We randomly picked times and dates in the weeks before data collection.

[4]The EU sample included two pre-treatment statements, and the US sample included five statements. See Figure A2 for the full list of pre-treatment questions.

[5]Our European sample was only exposed to the pro-health and pro-economy treatments.

[6]We did not randomise the gender in the profile pictures, the same text is always assigned to the same profile picture (two males and one female for the pro-economy tweets, and two females and one male for the pro-health tweets). However, since every respondent sees the same set of tweets and profile pictures, our effect is not driven by the gender assignment of the tweets.

Figure 1: Example of experimental variation



Notes: Individuals are exposed to the same message with different levels of endorsements, depending on treatment. Left tweet appears more popular than right tweet.

A second treatment dimension in the US sample exposed half the participants in each of the three above treatments to an *attention prime* prior to the six tweets. Participants were shown an unrelated tweet followed by three questions about: the content, the timing of this tweet, and (importantly) the number of likes. We designed this manipulation to prime participants into paying careful attention to the subsequent six tweets and their endorsements, since absence of treatment effects can potentially be attributable to participants not noticing the metrics. We find that the prime did not reinforce the expected effect and in fact nullifies the effect for social media users.[7] However, this treatment also allows us to assert that the lack of an effect for non-social media users is not due to lack of attention.[8] Our analysis focuses on the non-primed group, N=779 (out of 1,519) in the US, and N=605 in Europe. All pre-registered analyses of the primed group are shown in the supplementary materials, where we also confirm that our main results are robust to its inclusion.

After the six tweets, we elicited participants' post-treatment attitudes using a different set of questions about COVID-19 policy responses. Participants stated their agreement on a 7-point Likert scale to a number of policies, including *"Prohibiting gatherings"*, *"Closing non-essential businesses (bars, stores that are not food or health related, etc.)"*, and *"Closing daycares, schools, colleges and universities"*, among others.[9] We use the first

[7] Perhaps due to these participants realizing that the metrics were manipulated (shown in Table A9).

[8] Passive/non social media users in the primed group were more likely to correctly answer a manipulation check post-treatment (Table A10).

[9] The EU sample included seven post-treatment policies, and the US sample included eight. See appendix section 5.2 for the full list of policies asked about.

principal component of these responses as an index measure of post-treatment policy attitudes, our main outcome variable. We again defined positive values as being more pro-economy.

After the post-treatment attitude questions, we conducted a factual manipulation check by asking participants which view had more likes in the six tweets shown earlier: "pro-economy", "pro-health", "neither (both had about the same number of likes)", or "don't know". Participants could not go back to the previous screen to check the number of "likes" on the tweets.[10]

Finally, we collected data on education, income, self-reported political ideology on a 0-10 left-right scale, party voted in the last election (or if they voted), experience of COVID-19, degree of stubbornness measured by the participants' resistance to change (Oreg, 2003), media consumption, trust in the media and the government, and we asked individuals about the frequency with which they discuss policy issues with family and friends (both on and outside of social media). We measured participants' social media use by asking about time spent per day on the social media platforms Facebook and Twitter and define active social media users as those who spend more than one hour daily on Facebook or Twitter combined.[11]

3 Empirical analysis and results

Key summary statistics are shown in appendix Table A1, for the whole sample and split by social media use. Notably, active users are younger, hold a more right-wing ideology, and tend to support more pro-economy policies pre-treatment. The proportion of active users is highest in Ireland (30%) and lowest in the US midwest (17%), in all regions Facebook use is more common than Twitter.

3.1 Main treatment effects

We estimate the effect of social media endorsements on participants' policy attitudes using OLS as follows:

[10] Though the question was not incentivized, we do not see strong reasons for participants to misreport (consistent with findings in Allcott et al., 2020*a*).

[11] Our results are robust to defining active social media users as those who use Facebook for at least 30 minutes a day as shown in appendix Table A8. Additionally, for the US sample we asked participants to measure their own level of social media activity on a scale from 0-100. Our main result also holds when using this alternative measure of social media use. In particular, the effect of endorsements is concentrated on those who report the highest levels of activity.

$$PostAttitudes_i = \beta_0 + \beta_1 Treatment_i + PreAttitudes_i' \lambda + X_i' \delta + \varepsilon_i \quad (1)$$

The dependent variable $PostAttitudes_i$ is the standardized first principal component of the responses to the post-treatment policy questions, with higher value representing a more pro-economy attitude. $Treatment_i$ represents the assigned treatment and equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. Hence, participants exposed to tweets where pro-economy views appear more popular are expected to show an increase in $PostAttitudes_i$, while participants exposed to tweets showing popular pro-health views are expected to show a decrease in $PostAttitudes_i$. $PreAttitudes_i$ includes the first principal component to the responses to the pre-treatment policy questions. X_i' is a vector of control variables including gender (coded as a dummy for male), age, region (census regions for the US, country for the EU sample), household income (coded as the log of the midpoint of the interval specified by the subject), education (coded as a dummy for whether the subject has at least a 2-year college degree) and political ideology (self-reported response on a 0-10 left-right scale).[12] We include country fixed effects for all of our pooled analyses below and we use robust standard errors in all specifications.[13]

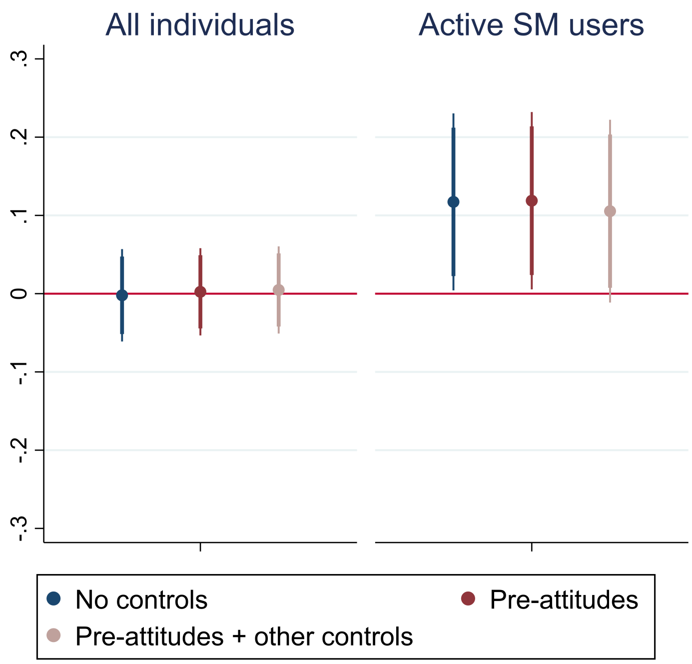
We estimate model 1 for i) the whole sample and ii) active social media users, which is the pre-registered subgroup of interest. In appendix Table A2 we additionally estimate the differential treatment effect for active social media users relative to non-active users, interacting treatment with $ActiveSMuser_i$, a dummy variable which equals 1 if the subject spends more than one hour a day on Facebook or Twitter (combined) and zero otherwise.

The coefficient of interest is β_1, the effect of perceived endorsements on policy attitudes. The results are shown in Figure 2. We observe no overall treatment effect on participants' policy attitudes. However, as hypothesised, we do find heterogeneity for

[12]Controlling for COVID-19 case numbers and a stringency index of government response yields similar results.

[13]Our specification differs from the pre-analysis plan in two ways. First, instead of including two separate coefficients and estimating their pooled average effect, we pool our treatments by defining a negative treatment for the pro-health group, a specification which is statistically equivalent to the one pre-registered (as shown in appendix Table A3) while being easier to interpret and implement. Second, because we are pooling our studies, we define region as the respondent's country for the Irish and Italian samples, and use the self-reported political ideology instead of vote for the Republican party in the last election. The analysis is robust to using a "right-wing" dummy, which equals 1 for US participants voting Republican and Italian participants voting Lega Nord in the last election, and 0 for all others (including Irish participants since there is no Irish right-wing party).

Figure 2: Main treatment effects



Notes: The figure shows the main treatment effects for all users (N=1,384) and active social media users (N=359) separately. Active social media users are defined as individuals who spend more than one hour daily on Facebook or Twitter combined. A fully interacted model that tests for differences between the two groups can be found in Table A2.

social media users, with a strong differential treatment effect for active users. Overall, the treatment shifts the attitudes of active social media users by about 0.12 standard deviations.

3.2 Heterogeneity by pre-treatment attitudes

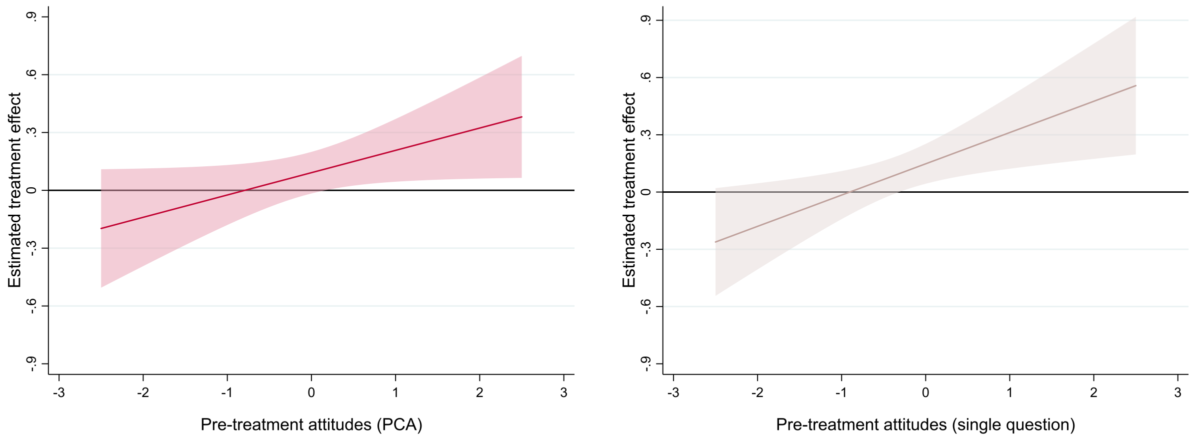
We estimate models of the following form:

$$PostAttitudes_i = \beta_0 + \beta_1 Treatment_i + \beta_2 PreAttitudes_i \quad (2)$$

$$+ \beta_3 Treatment_i \times PreAttitudes_i + X_i'\delta + \varepsilon_i \quad (3)$$

and do so specifically for active social media users.

Figure 3: Heterogeneity by pre-treatment attitudes



Notes: The figure shows estimated marginal treatment effects by pre-treatment attitudes. The results are also shown in table form in Table A4, including additional specifications.

We measure pre-treatment attitudes in two ways. First, we use the principal component of all the pre-treatment policy attitude questions. Second, we use the question that has the highest correlation with the post-treatment attitude index to represent participants' pre-treatment attitude. In the US sample, the question used is *"The government's highest priority should be saving as many lives as possible even if it means the economy will recover more slowly. What do you think of this statement?"*. In the European sample, the question used is *"Sweden's government has so far avoided implementing a lockdown in order to keep the economy going. What do you think of this policy?"*. Though this latter approach differs from the pre-registered specification, we find that the correlation between pre-treatment and post-treatment attitudes is substantially higher when using this measure, potentially better capturing the policy dimension of interest. We present results using both measures below.

The estimated marginal treatment effects from estimating model 2 are shown in Figure 3. We find that our treatment *polarizes* social media users. A one standard deviation increase in pre-treatment attitudes leads to a differential treatment effect of between 0.11 and 0.16 standard deviations (depending on how we measure pre-treatment attitudes). Put differently, individuals who held more polarized pre-treatment attitudes were more responsive to the treatment, and the treatment reinforced their pre-treatment attitudes. In the appendix we also present models in which we include triple-interactions of treatment, active social media use, and pre-treatment attitudes (Table A4).

3.3 Heterogeneity by manipulation check

A factual manipulation check allows us to identify individuals who paid conscious attention to the endorsement counts and to study the extent to which the treatment effects are driven by them (Kane and Barabas, 2019). We asked participants —after they had submitted their policy preferences— about the relative levels of "likes" in the tweets they had seen.[14] Below, we separate our sample by whether they answered this question correctly.[15] Importantly, this *post*-treatment attention check is endogenous (Montgomery, Nyhan and Torres, 2018), to the extent to which attention (or correct reporting) of the endorsement metrics is selective (see Iyengar et al., 2008; Wang, Morey and Srivastava, 2014, for evidence of motivated selective attention).[16]

The results are shown in Figure 4. We observe that the treatment effect on active social media users is concentrated on those who correctly answered the manipulation check. Because the coefficients appear robust to the addition of controls, selective attention is unlikely to explain these findings (Oster, 2019).[17] Instead, the results suggest that a relatively small sample of the population (about 10 percent) is very sensitive to the social cues provided by engagement metrics; the treatment shifts their policy views by about 0.38 standard deviations.

We also observe patterns suggestive of heterogeneous treatment effects for passive/non social media users. But these are not robust to the addition of controls, revealing instead that these users may pay selective attention to social cues which match their policy attitudes. In particular, passive/non social media users who correctly answered the manipulation check were more likely to hold views which aligned with the assigned treatment, while those who did not correctly answer the question were more likely to hold views which differed from their assigned treatment.

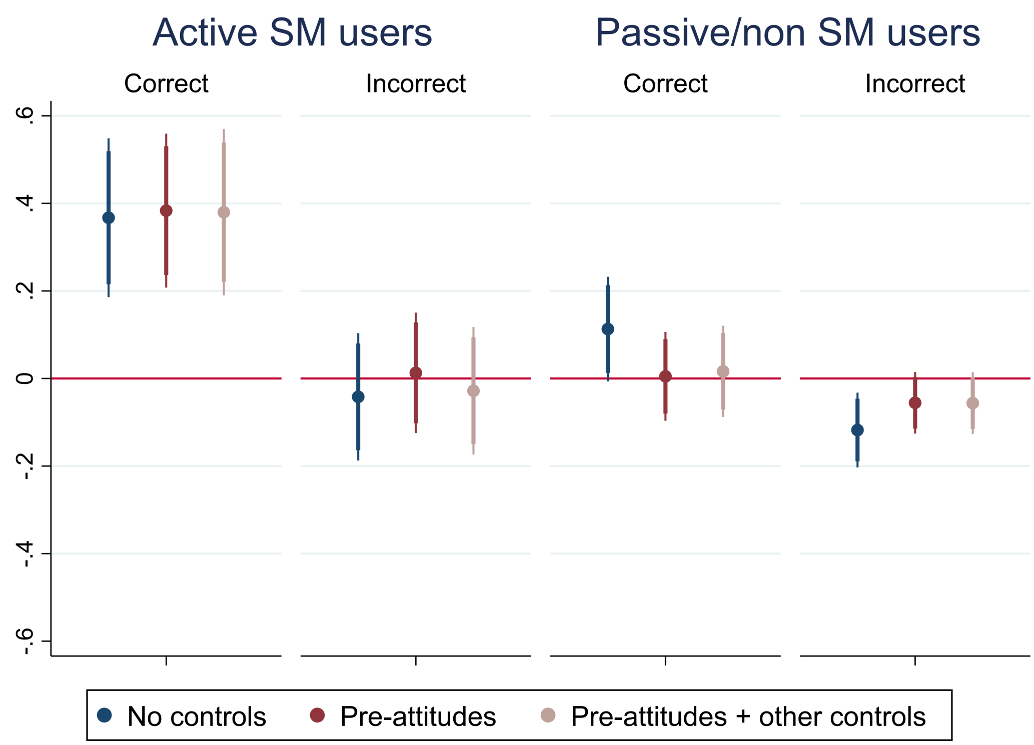
[14] We asked: "Views about COVID-19 policy response can be roughly split into two: (1) Pro-health: prioritise the elimination of COVID-19 over economic activities, for example by extending lockdown measures despite economic costs. (2) Pro-economy: prioritise economic activities over the elimination of COVID-19, for example by opening up the economy despite risks of a second wave. Which of these two views had more likes in the 6 tweets shown earlier? 'Pro-health', 'Pro-economy', 'Neither (both had about the same number of likes)', or 'Don't know'."

[15] Overall, 33.8 percent of participants answer the manipulation check correctly, with the rest answering incorrectly or "don't know". The proportion of correct responders is higher in the group of active social media users than passive/non-users (38.7 percent vs 32.1 percent, t-test, p=0.0225).

[16] Another possible explanation for incorrect reporting is consensus bias (Ross, Greene and House, 1977): respondents guess the answer in the direction of their own attitudes, thus failing the manipulation check if the treatment is not aligned with their views.

[17] Since the single question is highly correlated with post-treatment outcomes, it is likely to better capture unobservable selection as well. For this reason, we include both pre-treatment measures (PCA and single question) as controls.

Figure 4: Heterogeneity by manipulation check



Notes: The figure shows the main treatment effects separately depending on whether individuals correctly responded to the factual manipulation check question. In particular, we split our sample in four groups: active social media users who correctly answered the manipulation check (N=139), active social media users who incorrectly answered the manipulation check (N=220), passive/non SM users who correctly answered the manipulation check (N=329) and passive/non SM users who incorrectly answered the manipulation check (N=696). Estimates are also presented in Table A5 and a fully interacted model that tests for differences between the groups can be found in Table A6.

As further evidence of potential selection in these subsamples, regressing the treatment assignment on *pre-treatment* attitudes highlights that correctly answering the manipulation check is potentially endogenous (in appendix Table A7). The patterns appear particularly stark for passive/non social media users and suggest that they are more prone to selective attention. To evaluate the extent to which the heterogeneity in Figure 4 may be driven by selective attention, we test the sensitivity of our estimates to the addition of controls in a selection on unobservables framework (following Oster, 2019). Our results (shown in Figure A1) corroborate our reading of the results presented here, revealing that the estimates for passive/non social media users are sensitive to unobservable selection, while those for active social media users are not.

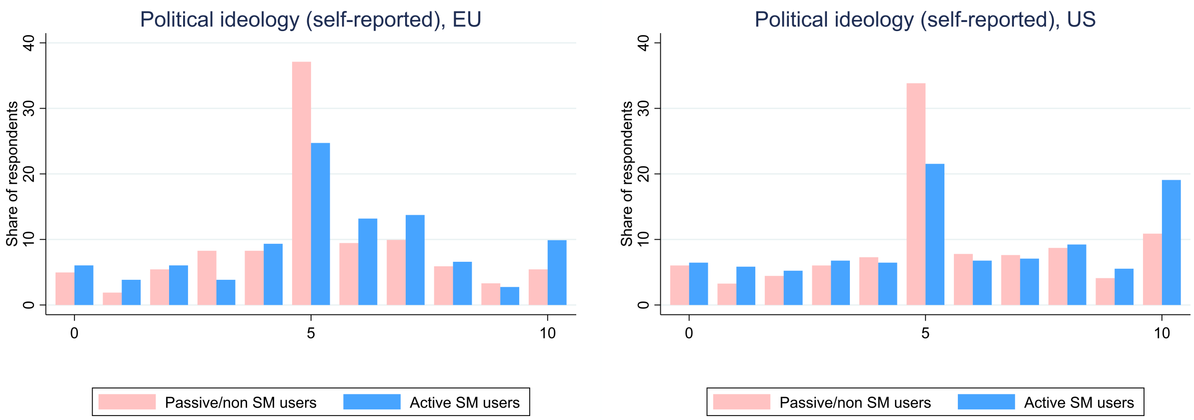
The analyses presented here suggest that the (relatively small) subset of active social media users who tend to pay conscious attention to endorsement metrics are heavily influenced by these social cues. On the other hand, passive/non social media users are more likely to notice endorsement metrics which reinforce their pre-existing attitudes, but they are on average not influenced by these metrics.

4 Discussion

Troubling trends in political polarization present serious challenges for the functioning of democracies (Przeworski, 2019). Many scholars have noted the concurrent rise of social media as a possible explanation for the increase in polarization. In concordance with these worries, active social media users in our survey were less likely to consider themselves politically moderate (Figure 5) and were less likely to hold (pre-treatment) moderate policy views with respect to COVID-19 (Figure A2). However, these patterns could well be the result of *selection* into social media use: individuals who hold more polar views tend to be more active on social media, perhaps as an outlet for their extreme opinions. Though recent work documents that deactivating Facebook can indeed reduce individuals' political polarization (Allcott et al., 2020*b*), the extent to —and the precise mechanisms through— which social media causes polarization remains debated.

To date, most work has emphasized ideological segregation or selective exposure (that is, the presence of echo chambers) as an important potential driver of polarization. Our controlled experiments in the US and Europe show that public endorsement metrics may also be a mechanism through which social media affects individuals' policy preferences and could also contribute to polarization. Importantly, substantial uncertainty surrounding the topic of COVID-19 is likely to make policy attitudes in this respect highly malleable. Future work should examine whether endorsements can shape policy

Figure 5: Self-reported political ideology and social media use



Notes: The figure shows the distribution of responses to the question "In political matters, people talk of 'the left' and 'the right'. How would you place your views on this scale, generally speaking?", separately for active and for passive/non social media users, and for our European (left, N=605) and US (right, N=1,519) samples.

attitudes in deeper entrenched topics in which views are likely to be more rigid. On the other hand, our results may underestimate the true effect of endorsement metrics on social media platforms where individuals are exposed to posts and endorsements by people they know.

Though the treatment effects we measured were large, they were concentrated on a small share of individuals: about 10 percent of participants, who are active social media users and paid attention to the endorsement metrics. That only a fraction of individuals were influenced in our experiment perhaps suggests that the broader effects of endorsement metrics on politics may be limited. However, social media dynamics could further propagate across society in different ways (Margetts et al., 2015; Tufekci, 2017). Social media engagement is also associated with other forms of political engagement, as such, these individuals could exert disproportionate influence in political processes (Vaccari et al., 2015; Barberá et al., 2019) and have a broad impact on public opinion (Centola et al., 2018). In our survey, active social media users are significantly more likely to (say they) have voted in the previous election. They also report more frequently discussing policy issues with friends or family members both on and outside of social media (Table A11).[18]

[18] These patterns are in line with findings in Guess (Forthcoming) of homogeneously partisan information consumption among only a minority of US citizens, but who nonetheless were on average more likely to vote.

We argue that micro-level studies such as ours can help disentangle the precise mechanisms through which social media affects users. Improved understanding of these mechanisms can inform social media platforms in the design of appropriate interventions to address issues of polarization, misinformation, and foreign influence in politics, among others (since platforms are unlikely to promote account deactivation, as in Allcott et al., 2020*b*). Finally, we hypothesize that social cues are likely to reinforce the effects of selective exposure. If individuals with stronger preferences are also more likely to "like" content, then not only will social media algorithms expose users to more polarized opinions (Levy, Forthcoming), but such content may also *appear* to have broader support. How these different features of social media interact to influence political views remains an important avenue for future work.

References

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational mobility and preferences for redistribution." *American Economic Review*, 108(2): 521–54.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social media and fake news in the 2016 election." *Journal of Economic Perspectives*, 31(2): 211–36.

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Y Yang. 2020*a*. "Polarization and public health: Partisan differences in social distancing during the Coronavirus pandemic." *Journal of Public Economics*, 191: 104254.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020*b*. "The welfare effects of social media." *American Economic Review*, 110(3): 629–76.

Bakshy, Eytan, Solomon Messing, and Lada A Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science*, 348(6239): 1130–1132.

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political Analysis*, 23(1): 76–91.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J Egan, Richard Bonneau, John T Jost, and Joshua A Tucker. 2019. "Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data." *American Political Science Review*, 113(4): 883–901.

Bond, Robert M, Jaime E Settle, Christopher J Fariss, Jason J Jones, and James H Fowler. 2017. "Social endorsement cues and political participation." *Political Communication*, 34(2): 261–281.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro. 2017. "Greater Internet use is not associated with faster growth in political polarization among US demographic groups." *Proceedings of the National Academy of Sciences*, 114(40): 10612–10617.

Bursztyn, Leonardo, and Robert Jensen. 2017. "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure." *Annual Review of Economics*, 9: 131–153.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2020. "From Extreme to Mainstream: The Erosion of Social Norms." *American Economic Review*, 110(11): 3522–48.

Carlson, Taylor N, and Jaime E Settle. 2016. "Political chameleons: An exploration of conformity in political discussions." *Political Behavior*, 38(4): 817–859.

Centola, Damon, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. 2018. "Experimental evidence for tipping points in social convention." *Science*, 360(6393): 1116–1119.

Cialdini, Robert B, and Noah J Goldstein. 2004. "Social influence: Compliance and conformity." *Annual Review of Psychology*, 55: 591–621.

Eady, Gregory, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A. Tucker. 2019. "How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data." *SAGE Open*, 9(1): 2158244019832705.

Egebark, Johan, and Mathias Ekström. 2018. "Liking what others "Like": using Facebook to identify determinants of conformity." *Experimental Economics*, 21(4): 793–814.

Gentzkow, Matthew, and Jesse M Shapiro. 2011. "Ideological segregation online and offline." *The Quarterly Journal of Economics*, 126(4): 1799–1839.

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal. 2020. "Does Information Change Attitudes Toward Immigrants?" *Demography*, 57(3): 1117–1143.

Guess, Andrew M. Forthcoming. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." *American Journal of Political Science*.

Guess, Andrew M, Brendan Nyhan, and Jason Reifler. 2020. "Exposure to untrustworthy websites in the 2016 US election." *Nature Human Behavior*, 4(5): 472–480.

Haaland, Ingar, and Christopher Roth. 2020. "Labor market concerns and support for immigration." *Journal of Public Economics*, 191: 104256.

Halberstam, Yosh, and Brian Knight. 2016. "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter." *Journal of Public Economics*, 143: 73–88.

Iyengar, Shanto, Kyu S Hahn, Jon A Krosnick, and John Walker. 2008. "Selective exposure to campaign communication: The role of anticipated agreement and issue public membership." *The Journal of Politics*, 70(1): 186–200.

Kane, John V, and Jason Barabas. 2019. "No harm in checking: Using factual manipulation checks to assess attentiveness in experiments." *American Journal of Political Science*, 63(1): 234–249.

Levy, Roee. Forthcoming. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review*.

Margetts, Helen, Peter John, Scott Hale, and Taha Yasseri. 2015. *Political turbulence: How social media shape collective action.* Princeton University Press.

Messing, Solomon, and Sean J Westwood. 2014. "Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online." *Communication Research*, 41(8): 1042–1063.

Montgomery, Jacob M, Brendan Nyhan, and Michelle Torres. 2018. "How conditioning on posttreatment variables can ruin your experiment and what to do about it." *American Journal of Political Science*, 62(3): 760–775.

Morales, Juan S. 2020. "Perceived Popularity and Online Political Dissent: Evidence from Twitter in Venezuela." *The International Journal of Press/Politics*, 25(1): 5–27.

Oreg, Shaul. 2003. "Resistance to change: Developing an individual differences measure." *Journal of Applied Psychology*, 88(4): 680.

Oster, Emily. 2019. "Unobservable selection and coefficient stability: Theory and evidence." *Journal of Business & Economic Statistics*, 37(2): 187–204.

Peterson, Erik, and Ali Kagalwala. 2021. "When unfamiliarity breeds contempt: How partisan selective exposure sustains oppositional media hostility." *American Political Science Review*, 1–14.

Przeworski, Adam. 2019. *Crises of democracy.* Cambridge University Press.

Ross, Lee, David Greene, and Pamela House. 1977. "The "false consensus effect": An egocentric bias in social perception and attribution processes." *Journal of Experimental Social Psychology*, 13(3): 279–301.

Scharkow, Michael, Frank Mangold, Sebastian Stier, and Johannes Breuer. 2020. "How social network sites and other online intermediaries increase exposure to news." *Proceedings of the National Academy of Sciences*, 117(6): 2761–2763.

Settle, Jaime E. 2018. *Frenemies: How social media polarizes America.* Cambridge University Press.

Sunstein, Cass R. 2018. *#Republic: Divided democracy in the age of social media.* Princeton University Press.

Tufekci, Zeynep. 2017. *Twitter and tear gas: The power and fragility of networked protest.* Yale University Press.

Vaccari, Cristian, Augusto Valeriani, Pablo Barberá, Rich Bonneau, John T Jost, Jonathan Nagler, and Joshua A Tucker. 2015. "Political expression and action on social media: Exploring the relationship between lower-and higher-threshold political activities among Twitter users in Italy." *Journal of Computer-Mediated Communication*, 20(2): 221–239.

Wang, Zheng, Alyssa C Morey, and Jatin Srivastava. 2014. "Motivated selective attention during political ad processing: The dynamic interplay between emotional ad content and candidate evaluation." *Communication Research*, 41(1): 119–156.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political effects of the internet and social media." *Annual Review of Economics*, 12: 415–438.

5 For Online Publication

5.1 Appendix Tables

Table A1: Summary statistics

	All individuals		Active SM users		Passive/non SM users		
	Mean	SD	Mean	SD	Mean	SD	Difference
Age	45.105	16.089	39.916	14.841	46.922	16.120	7.006***
Male	0.467	0.499	0.493	0.501	0.458	0.498	-0.035
Education	0.626	0.484	0.602	0.490	0.635	0.482	0.033
Income	6.700	2.670	6.806	2.411	6.663	2.756	-0.144
Political ideology	5.316	2.632	5.680	2.856	5.189	2.539	-0.490***
Ireland	0.220	0.415	0.256	0.437	0.208	0.406	-0.048*
Italy	0.217	0.412	0.251	0.434	0.205	0.404	-0.046*
USA West	0.212	0.409	0.271	0.446	0.194	0.396	-0.077**
USA Midwest	0.231	0.422	0.169	0.376	0.249	0.433	0.080**
USA Northeast	0.175	0.380	0.169	0.376	0.176	0.381	0.007
USA South	0.215	0.411	0.192	0.395	0.223	0.417	0.031
Active SM users	0.259	0.438	1.000	0.000	0.000	0.000	-1.000
Facebook use	1.256	1.043	2.532	0.727	0.809	0.717	-1.723***
Twitter use	0.719	0.960	1.663	1.151	0.388	0.599	-1.275***
Observations	1384		359		1025		1384

Notes: Summary statistics of age, gender (coded as a dummy for male), education (dummy for having at least a 2-year college degree), household income (log of the midpoint of the interval specified by the subject), political ideology (self-reported response on a 0-10 left-right scale), region (dummy for country for EU, census region for the US), ActiveSMuser (dummy for spending more than one hour a day on Facebook or Twitter combined), Facebook use (daily time spent on Facebook, 0 "never/no account", 1 "less than 30 minutes", 2 "from 30 minutes to 1 hour", 3 "more than 1 hour"), Twitter use (daily time spent on Twitter, 0 "never/no account", 1 "less than 30 minutes", 2 "from 30 minutes to 1 hour", 3 "more than 1 hour"). Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A2: Main Treatment Effects

	All individuals								Active SM users			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment	-0.002	0.002	0.005	0.010	-0.039	-0.032	-0.025	-0.036	0.117**	0.119**	0.105*	0.148***
	(0.030)	(0.028)	(0.028)	(0.026)	(0.035)	(0.031)	(0.031)	(0.028)	(0.058)	(0.058)	(0.059)	(0.056)
Treatment x Active SM user					0.161**	0.162**	0.145**	0.195***				
					(0.067)	(0.067)	(0.068)	(0.063)				
N	1384	1384	1384	1384	1384	1384	1384	1384	359	359	359	359
R-sq	0.003	0.115	0.126	0.285	0.019	0.143	0.156	0.303	0.027	0.044	0.075	0.177
Pre-attitudes (PCA)	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Other controls	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Pre-attitudes (Single Q)	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes

Notes: OLS estimates using post-treatment attitudes index (first principal component of the responses to the post-treatment policy questions) as outcome. The treatment variable equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. Pre-attitude controls include the first principal component of the pre-treatment policy questions and the single question with the highest correlation with post-treatment attitudes. Columns 5-8 present estimates of models of this form:

$$PostAttitudes_i = \beta_0 + \beta_1 Treatment_i + \beta_2 ActiveSMuser_i + \beta_3 Treatment_i \times ActiveSMuser_i + X_i'\delta + \varepsilon_i$$

ActiveSMuser is a dummy variable which equals 1 if the subject spends more than one hour a day on Facebook or Twitter (combined) and zero otherwise. Controls include age, gender, region (fixed effects), education, income and political position. All specifications include country fixed effects. Robust standard errors in parentheses. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A3: Main effects for pro-economy and pro-health treatments separately

	All individuals				Active SM users			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Econ	-0.007	0.002	-0.001	0.047	0.172	0.159	0.142	0.233
	(0.085)	(0.079)	(0.079)	(0.067)	(0.168)	(0.166)	(0.165)	(0.150)
Treatment Health	-0.003	-0.002	-0.010	0.026	-0.062	-0.079	-0.069	-0.063
	(0.084)	(0.079)	(0.079)	(0.067)	(0.159)	(0.157)	(0.155)	(0.142)
TE: $(\beta_1 - \beta_2)/2$	-0.002	0.002	0.005	0.011	0.117**	0.119**	0.105*	0.148***
	(0.030)	(0.028)	(0.028)	(0.026)	(0.058)	(0.058)	(0.060)	(0.056)
N	1384	1384	1384	1384	359	359	359	359
R-sq	0.003	0.115	0.126	0.285	0.027	0.044	0.074	0.178
Pre-attitudes (PCA)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Other controls	No	No	Yes	Yes	No	No	Yes	Yes
Pre-Attitudes (Single Q)	No	No	No	Yes	No	No	No	Yes

Notes: OLS regressions with the post-treatment attitudes index (first principal component of the responses to the post-treatment policy questions) as outcome. Treatment Econ (Health) equals 1 for the pro-economy (pro-health) treatment and 0 otherwise. TE equals the average treatment effect of the pro-economy and pro-health treatments, calculated as $(\beta_1 - \beta_2)/2$. Pre-attitude controls include the first principal component of the pre-treatment policy questions and the single question with the highest correlation with post-treatment attitudes. Controls include age, gender, region (fixed effects), education, income and political position. All specifications include country fixed effects. Robust standard errors in parentheses. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A4: Heterogeneity by Pre-treatment attitudes

	All individuals		Active SM users		All individuals		Active SM users	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.033	-0.026	0.092^{*}	0.079	-0.041	-0.038	0.148^{***}	0.141^{***}
	(0.032)	(0.032)	(0.055)	(0.057)	(0.029)	(0.029)	(0.053)	(0.054)
Treatment x Active SM user	0.148^{**}	0.130^{**}			0.205^{***}	0.198^{***}		
	(0.065)	(0.065)			(0.061)	(0.061)		
Pre-treatment attitudes	0.359^{***}	0.352^{***}	0.107^{*}	0.149^{**}	0.529^{***}	0.516^{***}	0.352^{***}	0.351^{***}
	(0.030)	(0.032)	(0.055)	(0.060)	(0.027)	(0.028)	(0.057)	(0.058)
Treatment x Pre-attitudes	-0.008	-0.014	0.116^{*}	0.114^{*}	-0.035	-0.042	0.164^{***}	0.162^{**}
	(0.040)	(0.040)	(0.060)	(0.059)	(0.034)	(0.034)	(0.062)	(0.063)
T x Active SM user x Pre-attitudes	0.076	0.087			0.185^{**}	0.197^{***}		
	(0.076)	(0.077)			(0.074)	(0.074)		
N	1384	1384	359	359	1384	1384	359	359
R-sq	0.144	0.158	0.057	0.087	0.298	0.307	0.183	0.199
Pre-attitudes measure	PCA	PCA	PCA	PCA	Single Q.	Single Q.	Single Q.	Single Q.
Other controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions with the post-treatment attitudes index (first principal component of the responses to the post-treatment policy questions) as outcome. The treatment variable equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. ActiveSMuser is a dummy variable which equals 1 if the subject spends more than one hour a day on Facebook or Twitter (combined) and zero otherwise. Controls include age, gender, region (fixed effects), education, income and political position. All specifications include country fixed effects. Robust standard errors in parentheses. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A5: Heterogeneity by manipulation check

	Active SM users				Passive/non SM users			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.367^{***}	0.380^{***}	-0.042	-0.028	0.113^{*}	0.016	-0.118^{***}	-0.056
	(0.092)	(0.096)	(0.074)	(0.074)	(0.061)	(0.053)	(0.043)	(0.036)
N	139	139	220	220	329	329	696	696
R-sq	0.129	0.273	0.021	0.216	0.013	0.335	0.012	0.389
Correct m. check	Yes	Yes	No	No	Yes	Yes	No	No
Pre-attitudes	No	Yes	No	Yes	No	Yes	No	Yes
Other controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions with the post-treatment attitudes index (first principal component of the responses to the post-treatment policy questions) as outcome. The treatment variable equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. Pre-attitude controls include the first principal component of the pre-treatment policy questions and the single question with the highest correlation with post-treatment attitudes. The sample is split both between Active SM users and Passive/non SM users and by those who correctly answered a post-treatment manipulation check asking participants which view had more likes in the six tweets shown. Controls include age, gender, region (fixed effects), education, income and political position. All specifications include country fixed effects. Robust standard errors in parentheses. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A6: Heterogeneity by manipulation check

	All individuals			Active SM users			Passive/non SM users		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.116***	-0.067*	-0.064*	-0.047	0.005	-0.014	-0.118***	-0.055	-0.055
	(0.043)	(0.036)	(0.035)	(0.073)	(0.069)	(0.071)	(0.043)	(0.035)	(0.035)
Treatment x Active SM user	0.062	0.091	0.084						
	(0.083)	(0.077)	(0.077)						
Treatment x Correct metrics	0.226***	0.082	0.087	0.417***	0.372***	0.396***	0.231***	0.055	0.068
	(0.074)	(0.062)	(0.061)	(0.114)	(0.110)	(0.113)	(0.074)	(0.062)	(0.062)
T x Active SM user x Correct metrics	0.215	0.267**	0.260**						
	(0.135)	(0.127)	(0.127)						
N	1384	1384	1384	359	359	359	1025	1025	1025
R-sq	0.037	0.304	0.311	0.061	0.189	0.208	0.016	0.354	0.367
Pre-attitudes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Other controls	No	No	Yes	No	No	Yes	No	No	Yes

Notes: OLS regressions with the post-treatment attitudes index (first principal component of the responses to the post-treatment policy questions) as outcome. The treatment variable equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. Pre-attitude controls include the first principal component of the pre-treatment policy questions and the single question with the highest correlation with post-treatment attitudes. Controls include age, gender, region (fixed effects), education, income and political position. All specifications include country fixed effects. Robust standard errors in parentheses. Significance levels shown below *p<0.10, ** p<0.05, ***p<0.01.

Table A7: Manipulation check and pre-treatment attitudes

	Active SM users				Passive/non SM users			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.004	-0.040	-0.027	-0.145*	0.122**	0.198***	-0.085**	-0.095**
	(0.100)	(0.094)	(0.081)	(0.076)	(0.055)	(0.058)	(0.042)	(0.042)
N	139	139	220	220	329	329	696	696
R-sq	0.007	0.072	0.006	0.037	0.057	0.068	0.012	0.012
Correct m. check	Yes	Yes	No	No	Yes	Yes	No	No
Pre-attitudes measure	PCA	Single Q.	PCA	Single Q.	PCA	Single Q.	PCA	Single Q.

Notes: OLS regressions with the **pre-treatment** attitudes as outcome. Pre-attitudes are measured both as the first principal component of the pre-treatment policy questions and the single question with the highest correlation with post-treatment attitudes. The treatment variable equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. The sample is split both between Active SM users and Passive/non SM users and by those who correctly answered a post-treatment manipulation check asking participants which view had more likes in the six tweets shown. Controls include age, gender, region (fixed effects), education, income and political position. All specifications include country fixed effects. Robust standard errors in parentheses. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A8: Defining Active SM users as those using Facebook for at least 30 minutes a day

	Active SM users				Active SM users		Passive/non SM users		Active SM users			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment	0.102^{**}	0.097^{**}	0.091^{*}	0.104^{**}	0.066	0.060	0.100^{**}	0.096^{**}	0.265^{***}	0.259^{***}	-0.002	0.008
	(0.049)	(0.049)	(0.050)	(0.046)	(0.046)	(0.047)	(0.044)	(0.045)	(0.077)	(0.078)	(0.065)	(0.060)
Pre-treatment attitudes					0.166^{***}	0.192^{***}	0.420^{***}	0.424^{***}				
					(0.052)	(0.056)	(0.049)	(0.050)				
Treatment x Pre-attitudes					0.141^{**}	0.143^{**}	0.136^{**}	0.133^{**}				
					(0.057)	(0.057)	(0.056)	(0.056)				
N	517	517	517	517	517	517	517	517	202	202	315	315
R-sq	0.016	0.050	0.064	0.205	0.066	0.080	0.211	0.218	0.064	0.204	0.012	0.261
Correct m. check									Yes	Yes	No	No
Pre-attitudes (PCA)	No	Yes	Yes	Yes					No	Yes	No	Yes
Other controls	No	No	Yes	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Pre-attitudes (Single Q)	No	No	No	Yes					No	Yes	No	Yes
Pre-attitudes					PCA	PCA	Single Q.	Single Q.				

Notes: OLS regressions with the post-treatment attitudes index (first principal component of the responses to the post-treatment policy questions) as outcome. The treatment variable equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. Pre-attitudes are measured both as the first principal component of the pre-treatment policy questions and/or the single question with the highest correlation with post-treatment attitudes, as indicated. Controls include age, gender, region (fixed effects), education, income and political position. All specifications include country fixed effects. Robust standard errors in parentheses. Significance levels shown below *p<0.10, ** p<0.05, ***p<0.01.

Table A9: Attention Prime Treatment

	Active SM users			Passive/non SM users		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.124	0.145^{**}	0.153^{**}	-0.011	-0.017	-0.012
	(0.084)	(0.064)	(0.063)	(0.052)	(0.035)	(0.035)
Attention prime	-0.083	-0.050	-0.057	0.000	-0.001	0.002
	(0.096)	(0.076)	(0.076)	(0.059)	(0.044)	(0.043)
Treatment x Attention prime	-0.265^{**}	-0.194^{**}	-0.183^{**}	0.053	0.024	0.017
	(0.118)	(0.087)	(0.085)	(0.072)	(0.052)	(0.052)
N	325	325	325	1194	1194	1194
R-sq	0.017	0.396	0.438	0.001	0.445	0.449
Pre-attitudes	No	Yes	Yes	No	Yes	Yes
Other controls	No	No	Yes	No	No	Yes

Notes: OLS regressions with the post-treatment attitudes index (first principal component of the responses to the post-treatment policy questions) as outcome. The treatment variable equals 1 for the pro-economy treatment, -1 for the pro-health treatment, and 0 otherwise. Pre-attitude controls include the first principal component of the pre-treatment policy questions and the single question with the highest correlation with post-treatment attitudes. The Attention Prime treatment showed participants a non-COVID related tweet and asked questions about this (including number of likes), before the treatment. Controls include age, gender, region (USA midwest, USA northeast, USA south, USA west), education, income and political position. Only the US sample was subject to this treatment. See the supplementary materials for more details. Robust standard errors in parentheses. Significance levels shown below *p<0.10, ** p<0.05, ***p<0.01.

Table A10: Correct manipulation check and the Attention Prime treatment

	Active SM users			Passive/non SM users		
	(1)	(2)	(3)	(4)	(5)	(6)
Attention prime	0.038	0.031	0.027	0.087^{***}	0.086^{***}	0.089^{***}
	(0.052)	(0.052)	(0.052)	(0.026)	(0.026)	(0.026)
N	325	325	325	1194	1194	1194
R-sq	0.002	0.022	0.043	0.009	0.011	0.028
Pre-attitudes	No	Yes	Yes	No	Yes	Yes
Other controls	No	No	Yes	No	No	Yes

Notes: OLS regressions with an indicator (1/0) for whether individuals correctly answered the manipulation check as outcome. Pre-attitude controls include the first principal component of the pre-treatment policy questions and the single question with the highest correlation with post-treatment attitudes. The Attention Prime treatment showed participants a non-COVID related tweet and asked questions about this (including number of likes), before the treatment. Controls include age, gender, region (USA midwest, USA northeast, USA south, USA west), education, income and political position. Only the US sample was subject to this treatment. See the supplementary materials for more details. Robust standard errors in parentheses. Significance levels shown below *p<0.10, ** p<0.05, ***p<0.01.

Table A11: Political engagement and social media use

	Voted				Discuss policy on SM		Discuss policy off SM	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Active SM user	0.069^{***}	0.086^{***}	0.044^{**}	0.059^{***}	1.237^{***}	1.039^{***}	0.424^{***}	0.370^{***}
	(0.025)	(0.024)	(0.020)	(0.020)	(0.081)	(0.078)	(0.074)	(0.072)
N	605	605	1519	1519	1519	1519	1519	1519
R-sq	0.010	0.166	0.003	0.060	0.139	0.235	0.023	0.084
Sample	EU	EU	USA	USA	USA	USA	USA	USA
Other controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The table shows the correlation between political engagement and social media use. The dependent variable in columns 1-4 is a dummy equal to 1 if the individual reports having voted in the previous elections. The dependent variable in columns 5-8 is a numerical value (0-4) to the question "How often do you discuss policy issues with your friends or family members on social media (columns 5-6) / outside of social media (columns 7-8)? [Never, Rarely, Sometimes, Often, Always]" This questions was only asked for the US sample. Controls include age, gender, region (fixed effects), education, income and political position. Robust standard errors in parentheses. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

5.2 COVID-19 policy questions

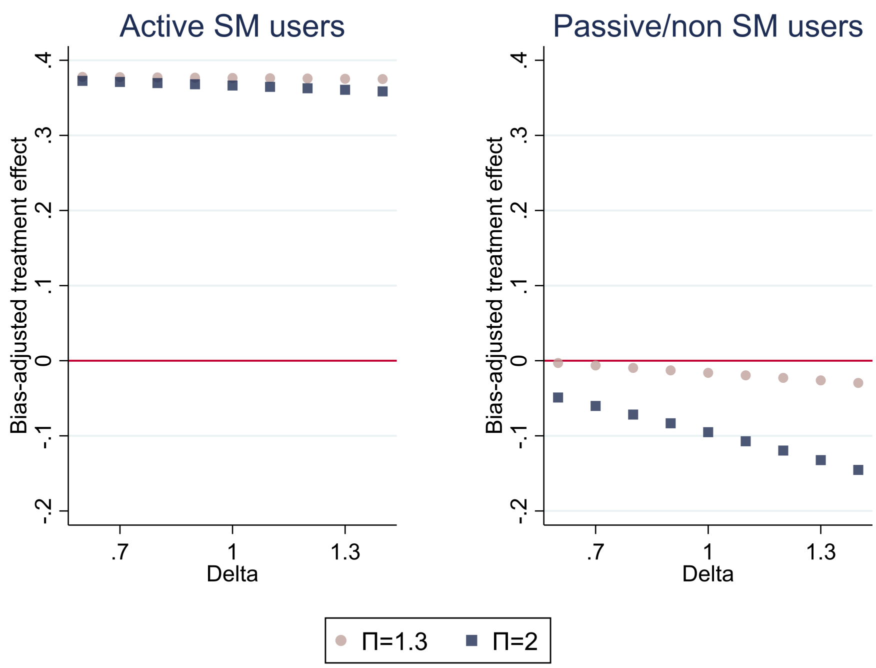
After the treatment, participants stated their agreement on a 7-point Likert scale to these policies:

- Closing the borders
- Prohibiting gatherings
- Prohibiting non-essential travels
- Closing daycares, schools, colleges and universities
- Closing non-essential businesses (bars, stores that are not food or health related, etc.)
- Handing out USD 1,000 fines to those who do not comply with social-distancing rules
- General lockdown of the population with a ban on leaving the home (except for medical reasons)
- Mandatory use of face-coverings in public places[19]

[19] This last question was added only to the US sample.

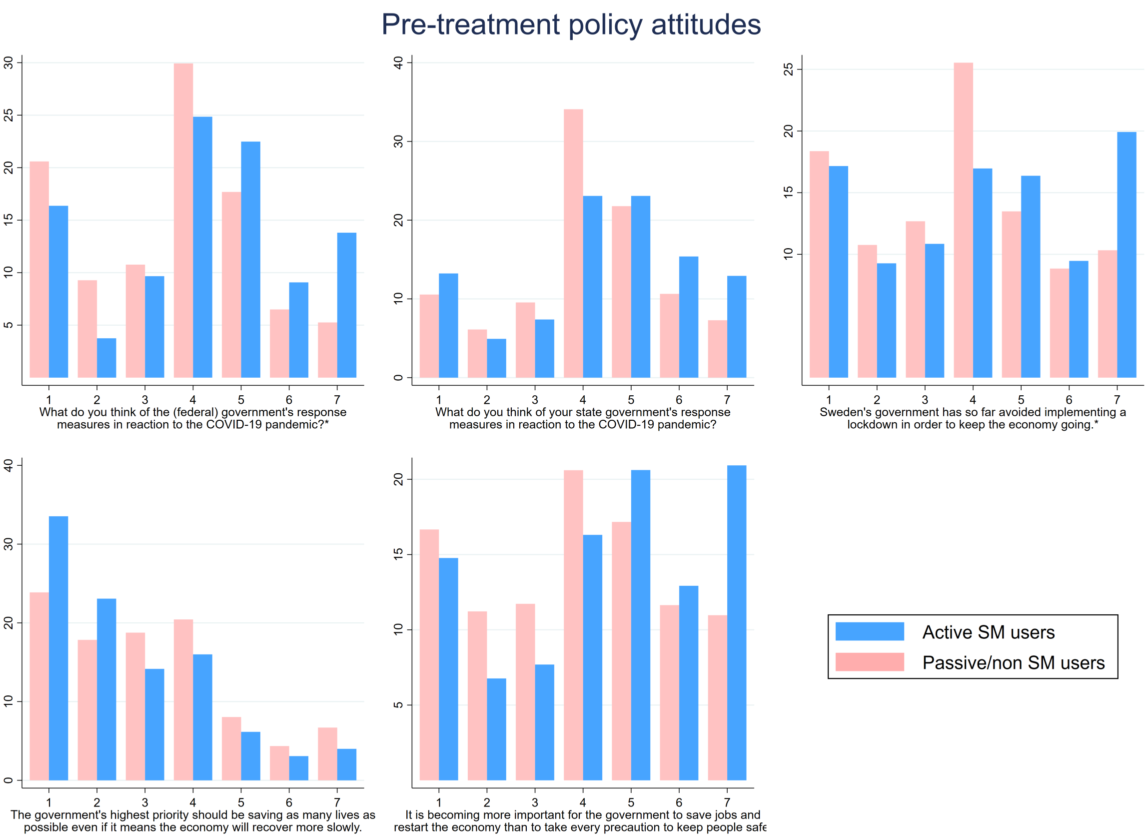
5.3 Appendix Figures

Figure A1: Selection-bias-adjusted treatment effects for participants with correct manipulation check



Notes: Following Oster (2019), the figure shows the estimated bias-adjusted treatment effects for a range of values of δ and two values of Π ($\Pi = 1.3$ is suggested, and $\Pi = 2$ is conservative). Controls include pre-treatment attitudes (both first principal component and single question) as well as age, gender, region (fixed effects), education, income and political position.

Figure A2: Pre-treatment policy attitudes and social media use



Notes: The figure shows the distribution of responses to the pre-treatment policy attitude questions, separately for active and for passive/non social media users. Questions marked with a * include participants in both Europe (N=605) and the US (N=1,519). All others include only US participants.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP20/11 Jonathan Briody: 'Parental Unemployment During the Great Recession and Childhood Adiposity' May 2020
WP20/12 Tensay Meles, Lisa Ryan, and Joe Wheatley: 'COVID-19 and EU Climate Targets: Going Further with Less?' May 2020
WP20/13 Xidong Guo and Sarah Parlane: 'Addressing Private Practice in Public Hospitals' May 2020
WP20/14 Patrick Massey and Moore McDowell: 'EU Competition Law: An Unaffordable Luxury in Times of Crisis?' May 2020
WP20/15 Judith M Delaney and Paul J Devereux: 'How Gender and Prior Disadvantage predict performance in College' May 2020
WP20/16 Morgan Kelly, Joel Mokyr and Cormac Ó Gráda: 'The Mechanics of the Industrial Revolution' June 2020
WP20/17 Morgan Kelly and Cormac Ó Gráda: 'Connecting the Scientific and Industrial Revolutions: The Role of Practical Mathematics' June 2020
WP20/18 Ronald B Davies, Dieter F Kogler and Ryan Hynes: 'Patent Boxes and the Success Rate of Applications' June 2020
WP20/19 Ronald B Davies, Yutao Han, Kate Hynes, and Yong Wang: 'Competition in Taxes and IPR' June 2020
WP20/20 David Madden: 'The Socioeconomic Gradient of Cognitive Test Scores: Evidence from Two Cohorts of Irish Children' June 2020
WP20/21 Deirdre Coy and Orla Doyle: 'Should Early Health Investments Work? Evidence from an RCT of a Home Visiting Programme' July 2020
WP20/22 Cormac Ó Gráda: 'On Plague and Ebola in a Time of COVID-19' August 2020
WP20/23 Morgan Kelly: 'Understanding Persistence' September 2020
WP20/24 Jeff Concannon and Benjamin Elsner: 'Immigration and Redistribution' September 2020
WP20/25 Diego Zambiasi: 'Drugs on the Web, Crime in the Streets - The Impact of Dark Web Marketplaces on Street Crime' September 2020
WP20/26 Paul Hayes, Séin Healy, Tensay Meles, Robert Mooney, Sanghamitra Mukherjee, Lisa Ryan and Lindsay Sharpe: 'Attitudes to Renewable Energy Technologies: Driving Change in Early Adopter Markets' September 2020
WP20/27 Sanghamitra Mukherjee: 'Boosting Renewable Energy Technology Uptake in Ireland: A Machine Learning Approach' September 2020
WP20/28 Ivan Petrov and Lisa Ryan: 'The Landlord-Tenant Problem and Energy Efficiency in the Residential Rental Market' October 2020
WP20/29 Neil Cummins and Cormac Ó Gráda: 'On the Structure of Wealth-holding in Pre-Famine Ireland' November 2020
WP20/30 Tensay Meles and Lisa Ryan: 'Adoption of Renewable Home Heating Systems: An Agent-Based Model of Heat Pump Systems in Ireland' December 2020
WP20/31 Bernardo Buarque, Ronald Davies, Ryan Hynes and Dieter Kogler: 'Hops, Skip & a Jump: The Regional Uniqueness of Beer Styles' December 2020
WP21/01 Kevin Devereux: 'Returns to Teamwork and Professional Networks: Evidence from Economic Research' January 2021
WP21/02 K Peren Arin, Kevin Devereux and Mieszko Mazur: 'Taxes and Firm Investment' January 2021
WP21/03 Judith M Delaney and Paul J Devereux: 'Gender and Educational Achievement: Stylized Facts and Causal Evidence' January 2021

UCD Centre for Economic Research
Email economics@ucd.ie