Papadopoulos, Georgios; Karatzas, Antonios; Martin, Thomas

Working Paper

A Reproduction of "Political Endorsement by Nature and Trust in Scientific Expertise During COVID-19" by Zhang (2023)

I4R Discussion Paper Series, No. 175

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Papadopoulos, Georgios; Karatzas, Antonios; Martin, Thomas (2024) : A Reproduction of "Political Endorsement by Nature and Trust in Scientific Expertise During COVID-19" by Zhang (2023), I4R Discussion Paper Series, No. 175, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/305223

INSTITUTE for REPLICATION

No. 175
I4R DISCUSSION PAPER SERIES

A Reproduction of "Political Endorsement by Nature and Trust in Scientific Expertise During COVID-19" by Zhang (2023)

Georgios Papadopoulos

Antonios Karatzas

Thomas Martin

October 2024

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 175

A Reproduction of "Political Endorsement by Nature and Trust in Scientific Expertise During COVID-19" by Zhang (2023)

Georgios Papadopoulos[1], Antonios Karatzas[1], Thomas Martin[2]

[1]University of East Anglia, Norwich/Great Britain
[2]Warwick University, Coventry/Great Britain

OCTOBER 2024

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

A reproduction of *"Political endorsement by Nature and trust in scientific expertise during COVID-19"* by Zhang (2023)

Georgios Papadopoulos (University of East Anglia), Antonios Karatzas (University of East Anglia), Thomas Martin (Warwick University)

Abstract

Zhang (2023) used an online, pre-registered, large-scale controlled experiment to test the effect of an endorsement of Joe Biden by the scientific journal *Nature* on several perceptual and behavioural outcomes. The main results of the paper were the following: the endorsement of Biden caused a large reduction in Trump supporters' trust in *Nature* and a considerably smaller reduction in their 'trust in US scientists'. The estimated effects are larger for individuals who, prior to the treatment, believed that *Nature* was unlikely to have endorsed a presidential candidate. The endorsement also made Trump supporters less likely to request COVID and vaccine related information from the endorsing journal. For Biden supporters, the respective estimated effects were generally positive, but small and insignificant. In his abstract, the author summarizes his key causal claim as follows: *"political endorsement by scientific journals can undermine and polarize public confidence in the endorsing journals and the scientific community"* (p.696). In this replication study, we computationally reproduced all results, with few and trivial exceptions. We then tested the robustness of those results that gave rise to Zhang's (2023) main causal claim. These tests include an alternative estimation method, an alternative way to capture support for the candidates, and a series of heterogeneity analyses by demographics. All test results support the author's findings but add interesting nuance. Some of our tests exploit variables from the raw data that were not included in the clean, published dataset, but the author willingly provided: a post-treatment 'manipulation check' that asked respondents to indicate the candidate that *Nature* actually endorsed, and data on requests for COVID related articles from other outlets besides *Nature*. We used these variables to conduct an Instrumental Variables (IV) procedure and test a 'causal mediation' model. Overall, and for Trump supporters in particular, our report corroborates the author's main finding of a strong negative effect of the endorsement on the overall perception of the endorser (*Nature*). However, the additional analysis provides weaker evidence for a reduction in trust in the scientific community more generally.

1. Introduction

Using a pre-registered, large online controlled experiment with approximately 4,000 subjects, this paper tested the impact of an endorsement of a presidential candidate (Joe Biden) by a high-status scientific publication (*Nature*) on various individual perceptual and behavioural outcomes. Through a randomised control trial, about half of survey participants read a short message summarising *Nature's* endorsement, while the control group read an irrelevant message about *Nature's* new design. The survey experiment was conducted about 10 months after the actual endorsement and about 8 months after the 2020 election. The author distinguishes between Trump and Biden supporters by eliciting respondents' *current* (i.e., at the time of the experiment) voting intention with a question inquiring them who they would vote for if they *"were to choose again"*.

The paper's main causal clam is *"political endorsement by scientific journals can undermine and polarize public confidence in the endorsing journals and the scientific community"* (p.696). The author makes this claim based of the following findings, which all relate to `treated' Trump supporters compared to 'control' Trump supporters: a) a large negative effect of the endorsement on trust in the journal *Nature* – its perceived informativeness (-0.85 standard deviation units with standard error 0.051) and unbiasedness (-0.63 standard deviation units with standard error 0.05); b) a small negative effect of the endorsement on general trust in scientists – their perceived informativeness (-0.13 standard deviation units with standard error 0.051) and unbiasedness (-0.16 standard deviation units with standard error 0.056); and c) a negative effect on the willingness to request COVID and vaccine related information from the endorsing journal *Nature* (-0.285 standard deviation unites with standard error 0.048). The author also examined the treatment effect on another five variables (capturing the subjects' perception of the competence of the two candidates and of climate change related issues), but the results are less conclusive, so the author barely discusses them. Because of this, and because the main causal claim does not hinge of these results, we do not refer to them in this report.

Another key point of the paper is a treatment heterogeneity analysis that the author runs based on the subjects' prior beliefs (specifically, how likely they considered *Nature* to have endorsed a political candidate in the run-up of the 2020 elections). The results, presented in Fig.4 of the manuscript, suggest that the treatment had a larger negative effect on Trump supporters who thought that *Nature* was unlikely to have endorsed a political candidate. The author then claims that this effect is driven by an *informational* mechanism, whereby subjects, resembling Bayesian agents, proceed to update their beliefs based on new information. A *contextual* explanation, such as priming, would have been supported if the treatment effect was relatively homogenous along the range of prior beliefs.

As part of the 'Norwich Replication Games' that took place in July 2023 at the University of East Anglia, we computationally reproduced all the results presented in the paper using the clean dataset and Stata code that were submitted by the author to the Harvard Dataverse. However, here we are primarily concerned with the *robustness reproducibility* of the author's main results, i.e., those results that give rise to the main causal claim (see points a, b, and c above). We note that for some robustness checks, we required access to the 'raw' data of the author that were not immediately available on the Harvard Dataverse. We contacted the author through the organizer of the replication games, and he was willing and happy to share the original raw dataset as directly downloaded from Qualtrics (the online platform on which the experiment was run). Our final dataset and accompanying Stata code can be found alongside this report on the Institute for Replication website.

In what follows (section 2), we begin with a discussion of the paper, focusing on three specific points of critique. However, the critical commentary does not cast doubt on the credibility and reliability of the author's results. Perhaps though, it challenges the author's level of confidence in the generalised

claims he makes. Our commentary is well-intended and, we hope, constructive. Although parts of the critique directly relate to some of the robustness tests that follow, we hope that, in its entirety, it will spark the interest of the reader and motivate further research into this very important contemporary topic.

Section 3 briefly reports that the author's work was fully reproduced computationally and adds a few minor remarks. We then proceed to a series of robustness tests and heterogeneity analyses (section 4) that generally support the author's baseline findings, adding interesting nuance is several ways. In section 5, we effectively extend the analysis by utilizing variables that were not included in the published, clean dataset, and consider an alternative analysis framework that we first motivate conceptually. These exercises are 'extensions', in the sense that they do not directly test the author's results but produce additional insight and nuance.

2. General discussion of the paper

The study is based on a treatment that is, in a way, retrospective. The endorsement of Joe Biden by *Nature* is an actual event, that took place 10 months *before* the experiment. The real-world effects of the endorsement, if any, had already materialized in the run-up to the election, and incorporated in the election's result (if the endorsement actually made 'exposed' voters to update their beliefs, and, in consequence, adjust their voting behaviour to reflect these updated beliefs). Of course, the author is not concerned with the election result, but with a change in the trust of subjects in *Nature* specifically, and US scientists more generally. Nevertheless, the fact that the experiment (and treatment) is not contemporaneous with the actual endorsement, poses some challenges to the interpretations of the results.

First, there is a possibility that subjects, especially Biden supporters, actually *knew* about the endorsement and had already updated their beliefs earlier in time (before the experiment). For such subjects, a 'null' treatment effect might not be a surprise. If pre-treatment actual knowledge of the endorsement (which is not observed in the data) is correlated with vote intention in the sense that Biden supporters being more likely to have known about the endorsement, the baseline results for Biden supporters might be downward biased. The author does not ask participants directly if they knew about the actual endorsement, but instead asks *"how likely do you think it is that Nature officially endorsed one of the candidates in 2020"*. With this question he aims at testing the explanatory value of two competing mechanisms at play – the "*information mechanism*" and "*priming"*. The author claims that his results support the informational explanation: for Trump supporters for whom the endorsement was more of a surprise, the negative effect on 'trust in *Nature*' was larger (due to, presumably, stronger update of beliefs). The statistical evidence of an informational mechanism for Biden supporters is very weak though, but the author attributes this to the smallness of the effect.

The author is probably justified to infer the existence of the informational mechanism, however, the question (in the manner it was phrased) cannot capture whether a respondent actually *knew* about the endorsement. Some indeed might have known, and others might be *guessing* (the latter being the evident intention of the question). Moreover, despite our reasonable assumption that Biden supporters might be more likely to have known about the actual endorsement, there is also another issue with the author's interpretation. It is not unlikely that some Trump supporters might have read about Nature's endorsement at the time and forgotten about it (indeed, the author mentions that *"these endorsements were widely reported by conservative media outlets"* (p.696) citing two politically charged, even smearing, news pieces from *Fox News* and the *National Review* from October 2020).

Such supporters might have answered that *Nature* was not likely to have endorsed a candidate, but then the treatment helps them *recall* the information, generating negative emotions. The possibility of an emotionally driven 'affective' response is an alternative explanation to a 'rational', belief updating informational explanation. The author does not seem to have considered this as a possibility, but, in our opinion, he would not be able to exclude it.

In any case, for an event that happened a whole 10 months before the treatment, where the event is the treatment itself (as in this paper), it is not that certain that this question differentiates between those for whom the treatment comprised new information and those for whom it did not.

Second, the author's main claim is that his study *"shows that electoral endorsements by Nature and potentially other scientific journals or organizations can undermine public trust in the endorser, particularly among supporters of the out-party candidate"*. Although at face value, this claim is empirically supported, we contend that it is not clear whether the results could generalise to other *"out-party candidates"* besides Donald Trump. Donald Trump is generally considered to be a unique phenomenon, a consequence of various causes, that galvanised and took on-board certain socio-economic groups, with the use of propaganda and affective communication[1] [2]. Crucially, Donald Trump has repeatedly himself made unscientific claims, attacked scientists, and undermined trust in scientific knowledge (e.g., Webb and Kurtz, 2022). Trump supporters might be fundamentally different to supporters of other out-party candidates abroad (or compared to supporters of past American presidential candidates), in the sense that they support someone who does not have the trust in science that is arguably essential for a policymaker. As such, with their responses to what effectively are questions about the status of science, Trump supporters might simply be emulating or 'parroting' their leader; the endorsement might simply be a 'trigger' for treated individuals to show more strongly (relatively to the control group) their allegiance to Donald Trump. We thus suspect, but have no way to prove, that the results of this study would not generalise widely. For example, would the effects of such an endorsement hold if instead of Donald Trump, the out-party candidate was John McCain or Mitt Romney, Barack Obama's Republican opposition in the 2008 and 2012 US general elections? What if in another developed country, the main candidates were both 'centrist' (one left- and one right-leaning)? Would the trust in *Nature* or science of the out-party centrist candidate's supporters change after a scientific publication's endorsement of the (other centrist) opponent? Our hypothesis is that treatment effects in such cases would be much weaker, if at all present. We believe that despite the excellent execution of this work and the interesting findings, given the peculiarity of Donald Trump and his supporters, researchers should consider replicating this work in other national settings.

Third, the author ran a manipulation check, asking respondents towards the end of the survey (but before requesting their demographical information) whether *"Nature made any explicit political statements in support of any candidates in the run up to the 2020 presidential election".* We were surprised to see that only 69% of the treated Biden supporters and 59% of the treated Trump supporters replied correctly that the journal endorsed Joe Biden. Even though these figures are much larger for treated versus control group participants, in absolute value they are surprisingly low, suggesting that a substantial fraction of the treated sample might not have understood what they read, or may have been inattentive. Although the author used this manipulation check to claim that the treatment was successful in terms of shifting actual knowledge of the endorsement, it was surprising to see that he did not question (or did something about) the quite high proportion of treated respondents (about 35%) who did not agree, post-treatment, that indeed *Nature* endorsed Biden.

[1] https://www.newyorker.com/magazine/2017/12/11/donald-trumps-fake-news-tactics
[2] https://www.bbc.co.uk/news/av/world-us-canada-46175024

3. **Reproducibility**

The published dataset and Stata scripts are well organized and easy to follow. All results are easily reproduced computationally, following the scripts. We did not identify any coding errors, but there are a few minor remarks that need to be made:

1. In the published paper, there is a series of typos in the last paragraph of the section *Trust in Nature* (last paragraph of the 1st column in p.698): all confidence intervals of the estimates are incorrect.
2. Table 1 (*Sample breakdown by demographics*) compares the study sample to the US adult population based on the 2020 American Community Survey (ACS). The code to reproduce it is not included in the do-file. The same holds for Figure 4 (*Treatment effect heterogeneity by prior belief*). However, we independently reproduced the latter, and all estimates match the ones reported in the author's Figure 4.
3. The published dataset is the author's 'clean' version, so it does not include some variables that the author used (to varying extent) in the supplementary analysis (e.g., the manipulation checks). As mentioned, upon request, the author kindly provided the raw dataset (i.e., as downloaded from Qualtrics) and some of the robustness checks reported in the next section employ some of the non-published data.
4. The author does not present any checks for whether the treatment and control groups are *balanced*, in the sense that characteristics of people in the treatment group are not, on average, statistically different from the characteristics of people in the control group. We conducted this balance check for all variables at the same time, using t-tests (for continuous variables) and chi-square tests (for categorical variables). We concluded that the two groups are balanced in terms of their observable characteristics, such as ideology, sex, education, age, race, area of living (urban/suburban/rural), interest in current events and interest in popular science. For brevity, the results of the statistical tests are not reported here.

As a final remark, we follow the author's analysis choice so in the entire replication exercise (with only exception being the robustness check in section 4.1) we use the 'standardised' outcome variables, which have a mean of 0 and standard deviation (SD) of 1. This is even though the original outcome variables are categorical and ordered in nature (commonly, in 5-point Likert scales). As a result, all reported coefficients in the original paper and this report represent differences in terms of standard deviations, even though such an interpretation may not be particularly useful or meaningful for ordinal variables (which can take only a limited number of values). We took this decision so as to enable direct comparisons of the treatment effects across models, with the author's baseline results. We do not intend this to be a criticism of the study, only a comment on a choice that was made probably out of convenience and to make interpretation easier.

For comparison purposes, Table 1 reproduces the point estimates of the treatment effect (with heteroskedasticity robust standard errors in brackets) for the main 4 outcome variables, separately for Trump and Biden supporters. Note that the 'Baseline dif.' represents the mean difference in the outcome variable between Trump and Biden supporters but for the control group.

Table 1: Baseline Results

	Nature informed	Nature unbiased	Scientists informed	Scientists unbiased
CATE Trump	-.854 (.052) [.000]	-.633 (.050) [.000]	- .130 (.053) [.014]	-.161 (.052) [.002]
CATE Biden	.108 (.031) [.000]	.045 (.031) [.167]	.048 (.033) [.146]	.016 (.031) [.606]
Baseline dif.	-.387 (.036) [.000]	-.655 (.040) [.000]	-.756 (.044) [.000]	-.937 (.042) [.000]
Sample Size	3885	3885	3885	3885

Heteroskedasticity robust standard errors in parentheses; *p*-values in square brackets.

Note that the author obtains these results by estimating the following linear model:

$$Y_i = \alpha + \beta D_i \times TS_i + \gamma D_i \times (1 - TS_i) + \delta TS_i + \epsilon_i, \quad (1)$$

where Y_i is the outcome variable for respondent i, D_i is a dummy variable taking the value of 1 if the respondent is in the treatment group and 0 if in the control group, TS_i is a dummy variable taking the value of 1 for Trump supporters and 0 for Biden supporters, and ϵ_i is a heteroskedastic error term. Here, β represents the treatment effect for Trump supporters, γ is the treatment effect for Biden supporters, and δ is the baseline difference. Instead, although equivalent, we preferred estimating the model as:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 TS_i + \beta_3 D_i \times TS_i + \epsilon_i. \quad (2)$$

In this model, β_1 is the treatment effect for Biden supporters, $\beta_1 + \beta_3$ is the treatment effect for Trump supporters, and β_2 is the baseline difference. We found this approach more convenient for the heterogeneous analysis we conducted below, where we add an additional 'triple' interaction term (for example, by prior belief, by demographics, etc.) and then obtain the treatment effects for each category of these additional moderating variables. For example, the model that investigates the moderating effect of sex (i.e. males or females), is as follows:

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 D_i + \beta_2 TS_i + \beta_3 D_i \times TS_i \\ & + \beta_4 M_i + \beta_5 D_i \times M_i + \beta_6 TS_i \times M_i + \beta_7 D_i \times TS_i \times M_i + \epsilon_i, \end{aligned} \quad (3)$$

where M_i is a dummy variable that is 1 for males and 0 for females. Here, the treatment effect, for example for Trump female supporters, is $\beta_1 + \beta_3$, and for Trump male supporters is $\beta_1 + \beta_3 + \beta_5 + \beta_7$. Please see the provided Stata do file for details of our estimation and postestimation approach.

4. Robustness, Sensitivity Analysis and Extensions

4.1 Alternative Estimation Method – Ordered logistic regression

In his entire analysis, the author assumes that the outcome variables, measured on an ordinal 5-point Likert scale, are of a cardinal scale, and consequently assumes that each of them is a linear function in the parameters. Therefore, he estimates all coefficients using linear models and the Ordinary Least Squares (OLS) estimator. As such, the first step in our replication exercise was to check whether the main results are driven, in anyway, by this decision, even though, in general, studies in social sciences have shown that in the case of ordinal outcome variables, results are in general not driven by the linearity in parameter assumption (Bloem 2022). To do that, we estimated non-linear regression

models, such as Ordinal Logit and Ordinal Probit models, that are more appropriate as they take the ordinal nature of the outcome variables into account. These models produced qualitatively similar results, both in terms of magnitudes of effects of the treatment as well as the statistical significance of these treatment effects. In addition, we tested whether there is evidence to reject the 'proportionality of odds' assumption imposed by Ordinal Logit/Probit models. Using the *omodel* Stata routine, we found strong evidence to reject the assumption in the models of all outcomes of interest. This suggests that the treatment effect varies for different levels of the outcome variables. Nevertheless, other models that relax the 'proportionality of odds' assumption, such as the Multinomial Logit or the Generalized Ordered Logit (user-written function *gologit2* in Stata), also produced results that are similar. However, inference from these models is not directly comparable to the OLS coefficients or the Ordinal Logit and Ordinal Probit ones. We decided not to explore this further, as it would divert the focus and add unnecessary complexity, and instead give room for the more interesting analyses of heterogeneous effects, reported in what follows.

4.2 Heterogeneity analysis

4.2.1 By prior belief

We reproduced the author's estimates of the treatment effect heterogeneity by prior belief of how likely it was *"for Nature to endorse a presidential candidate"*. The results are shown in in Fig.1A below, which are also presented in Fig.4 of the paper, but in different format. These figures present the treatment effects separately by Biden and Trump and for each category of the *prior belief* categorical variable (noting also that the figures also present the 95 Confidence Intervals for each treatment effect). As shown in Fig.1A, for Trump supporters, the negative effect of the treatment on 'trust in *Nature*' appears to get larger as this likelihood falls. The author finds no evidence that there is a moderating effect for Biden supporters. The author uses these results to claim support for an *"informational mechanism"* at play, rather than a *"priming effect"*: for subjects (especially Trump supporters) who did not expect *Nature* to have endorsed Joe Biden, this information is new and there a consequent stronger update of beliefs.

With all the caveats discussed in section 2, the author's claim sounds plausible. Following the same logic, one would expect to see that such prior beliefs would have a similar 'moderating' effect on the outcomes reflecting 'trust in US scientists'. However, the results in Fig.1B do not provide any clear evidence in support of this claim. There is no statistical evidence that the treatment effect becomes larger in magnitude as the likelihood of the belief that *Nature* endorsed a political candidate decreases; nevertheless, we still note that this effect is the largest for those subjects who thought that Nature endorsing a candidate was *"not likely at all"*.

However, the question about the likelihood of *Nature* endorsing *"a candidate"* is followed by a more specific question that the author does not consider in his analysis: *"Hypothetically, suppose Nature made an endorsement. Who would they endorse?"*. As mentioned in Section 2, one of our criticisms is that the author did not ask subjects directly if they actually knew about Biden's endorsement by *Nature*. This question, being conditional in nature, cannot directly capture whether a subject knew about the endorsement either. Nevertheless, it is an alternative way to capture 'prior belief' and test the explanatory value of the informational mechanism that the author espouses. We thus conducted the same analysis, but capturing prior belief with this second question. The results are graphically presented in Fig.2.

Fig.1A: Treatment effects by 'prior belief' – Trust in *Nature*

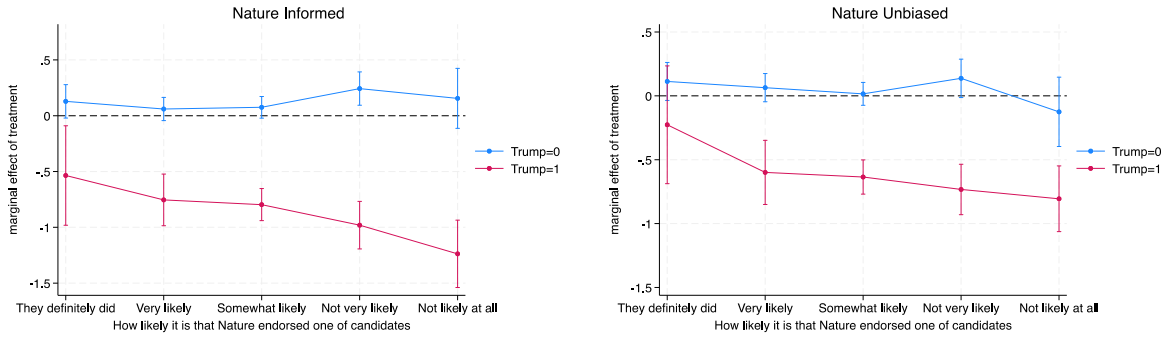


Fig.1B: Treatment effects by 'prior belief' – Trust in *Scientists* in general

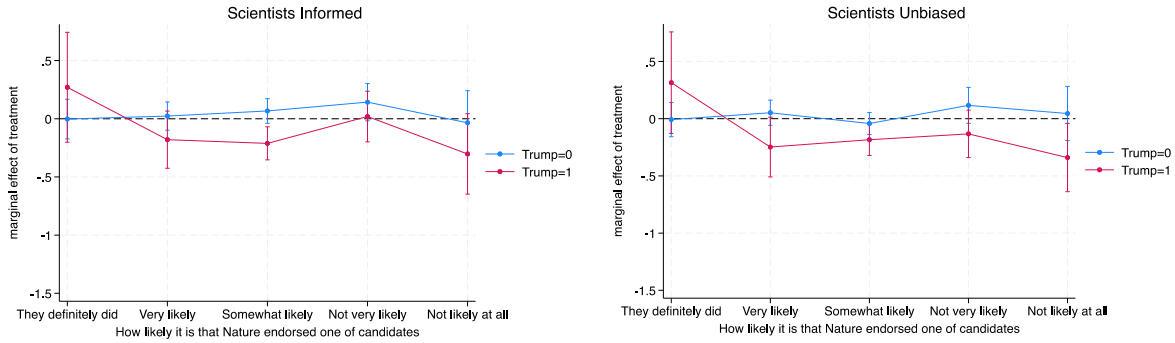
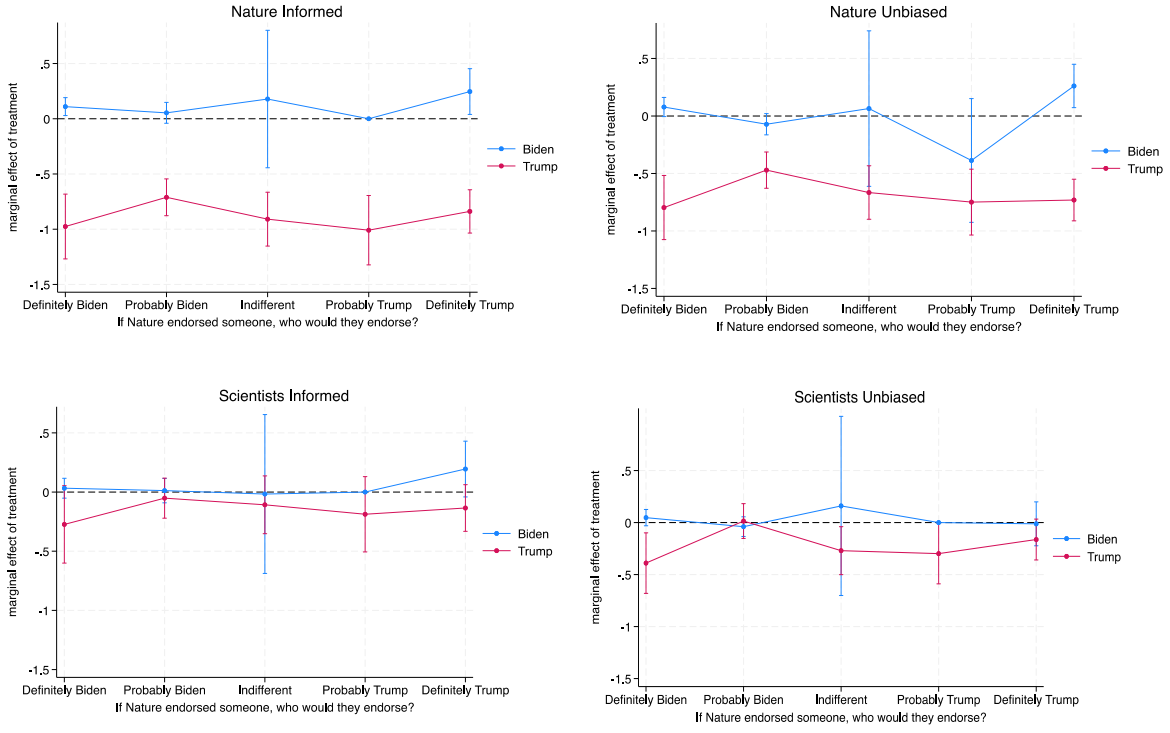


Fig.2: Treatment effects of prior belief captured using the alternative *prior belief* question



If the informational mechanism was at play, and in line with the previous results, one might expect that Trump supporters who thought that *Nature* was very likely to have endorsed their favourite candidate but then realized that it endorsed Joe Biden, would be forced to a larger updated of beliefs upon treatment, i.e., the treatment effect would have been stronger for them. We do not see this for any of the four outcomes of interest. In addition, in line with the baseline analysis, there is no clear pattern when it comes to Biden supporters.

This analysis casts some doubt on the "informational explanation", espoused by the author, which, in combination with the lack of a question to directly capture prior knowledge, makes this part of the paper (and related claims) seems a little bit rushed and underthought.

4.2.2 By demographics

The dataset provides rich information on demographic characteristics. The author, perhaps following his pre-registered plan, primarily used this information to check that the randomization of treatment worked, and for a regression adjustment using the LASSO technique (in the Supplementary file – which mainly contributes in terms of efficiency of the estimators) finding similar results.

However, we believe that this additional information provides an opportunity to assess whether the effects of the treatment differ depending on the demographic profile of individuals. Thus, as for *prior belief* above, we examine the moderating effect of some key demographic characteristics, one at a time.

Firstly, we looked for heterogeneous effects by sex. Table 2 shows that the treatment effects are similar for male and female Trump supporters (differences in marginal effects are not statistically different from zero), even though the effects appear to be slightly more negative for males.

However, the treatment effect for Biden supporters is stronger (and positive) for females for the *"Nature Informed"* and *"Scientists Informed"* outcomes. Although the seems like an interesting finding, we do not have a testable hypothesis behind, hence do not elaborate on it.

Table 2: Treatment Effects by Sex

	Nature informed	**Nature unbiased**	**Scientists informed**	**Scientists unbiased**
CATE Trump				
Males	-.874 (.076) [.000]	-.652 (.065) [.000]	- .135 (.073) [.065]	-.166 (.069) [.016]
Female	-.830 (.071) [.000]	-.608 (.076) [.000]	- .128 (.078) [.101]	-.154 (.078) [.048]
CATE Biden				
Males	.061 (.042) [.146]	.055 (.043) [.201]	.009 (.045) [.841]	.023 (.044) [.601]
Female	.150 (.044) [.001]	.039 (.045) [.386]	.092 (.048) [.055]	.012 (.043) [.780]
Sample Size	3868	3868	3868	3868

Heteroskedasticity robust standard errors in parentheses; *p*-values in square brackets.

Second, we looked at whether the effects differ by age group. Figure 3 illustrates the treatment effects for Biden and Trump supporters, by age group. For '*Nature* informed', the treatment effects are generally twice as large for older Trump supporters (>45 years). One can see a similar trend for the '*Nature* unbiased' outcome, while the effect is less heterogeneous in terms of age for the two variables

reflecting 'trust in US scientists'. When it comes to Biden supporters, the treatment effects on the two variables representing 'trust in *Nature'* are generally stronger in the 55-65 and >65 age groups. This suggests that the endorsement had a stronger 'polarizing' effect in older age groups.

Fig 3: Treatment Effects by Age

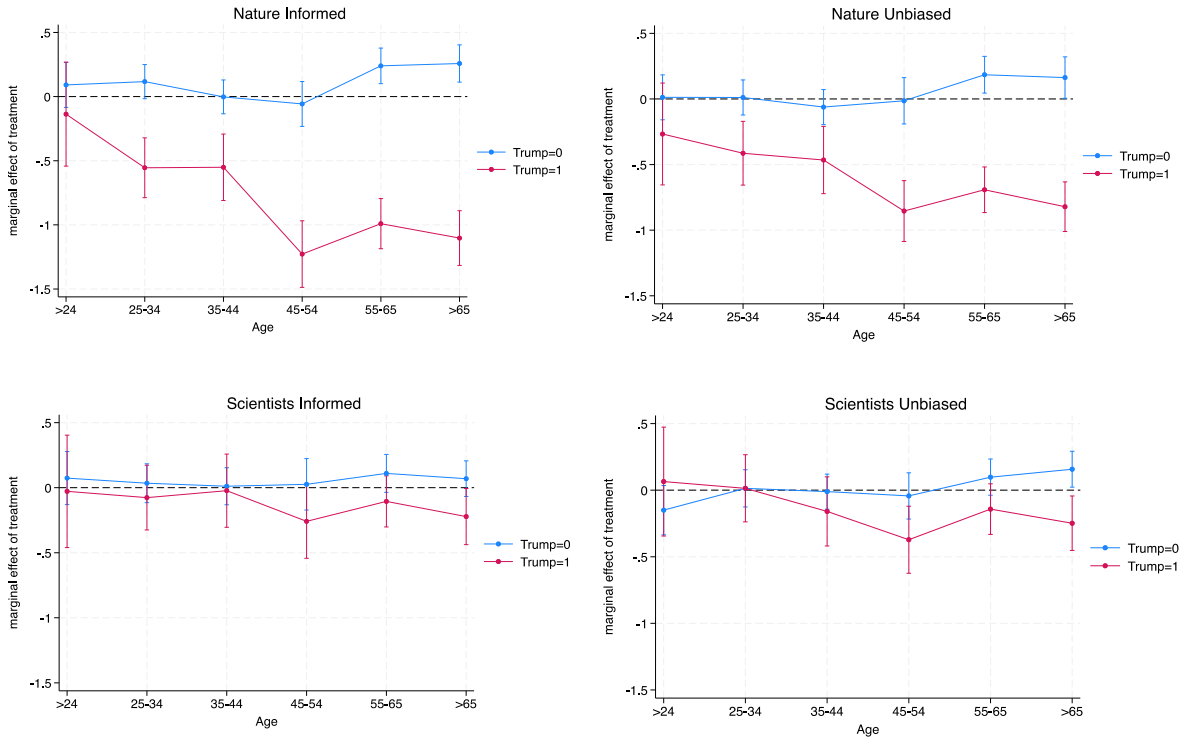
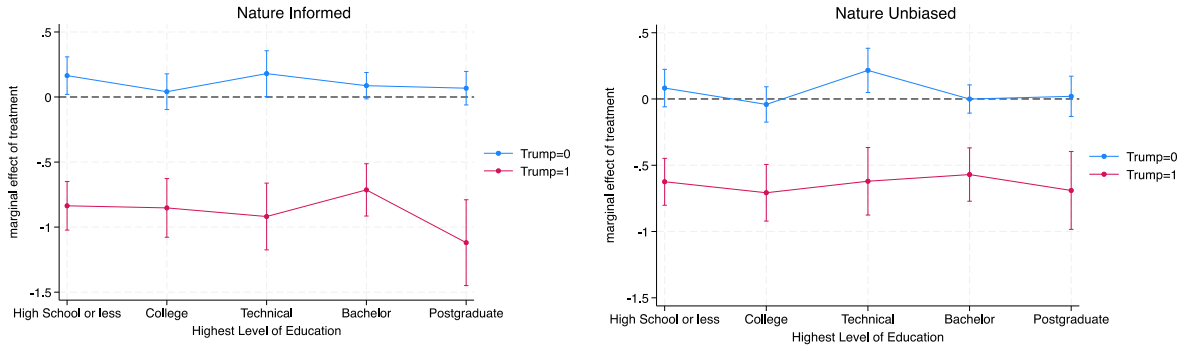
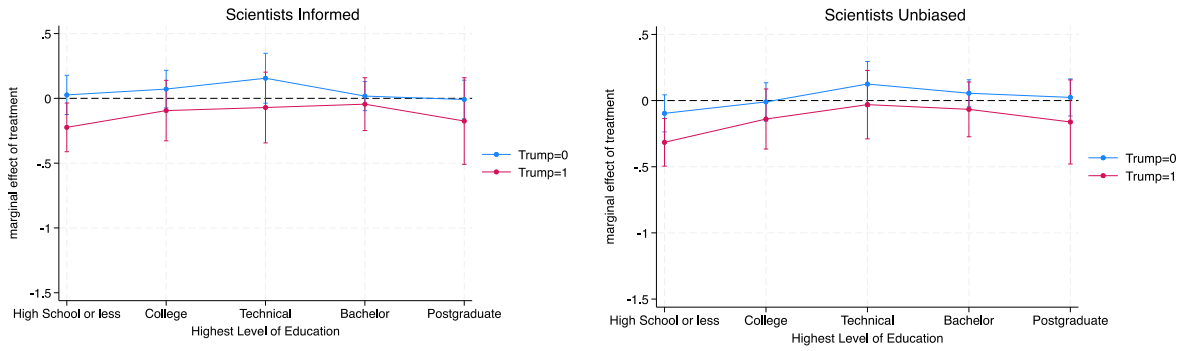


Fig.4 present the treatment effects for Trump and Biden supporters, by education level. Even though differences are relatively small, the treatment seems to have had the strongest negative effect on those Trump supporters with a postgraduate degree when it comes to the 'trust in *Nature'* variables. On the other hand, when it comes to 'trust in US scientists' it is the least educated Trump supporters (*"high school or less"*) who respond most negatively. This suggests that there might be two different mechanisms at play. We leave this for interested scholars to theorize upon and empirically investigate.

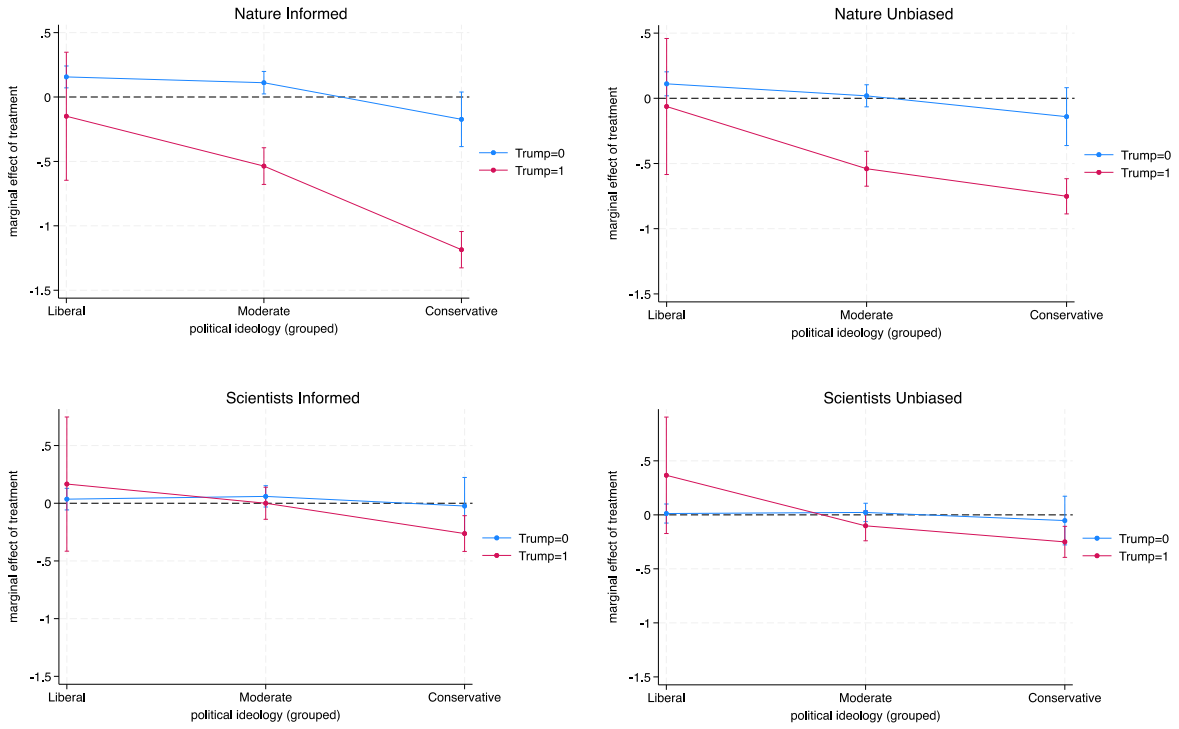
Fig 4: Treatment Effects by Level of Education





The author asked respondents to declare their ideology on a 7-point Likert scale, ranging from *"extremely liberal"* to *"extremely conservative"*. Due to some sub-samples being very small, we grouped the two left and right levels into *"liberal"* and *"conservative"*, and the three intermediate levels into *"moderate"*, and explored the heterogeneity of treatment effects by ideology for Biden and Trump supporters. Fig.5 illustrates that the negative treatment effects for Trump supporters on their 'trust in science' are partially driven by their conservativeness. In converse, for Biden supporters, the small but positive treatment effect concerns primarily liberal individuals (and moderates in the case of *Nature's* informativeness). For the two variables reflecting 'trust in science', ideology plays a smaller role, but it is still clear that the negative treatment effect for Trump supporters concerns almost entirely the conservative sub-sample.

Fig.5: Treatment Effects by Ideology



4.2.3 By using a propensity score

The analysis based on the subjects' demographics shows that the estimated treatment effects are, to some extent, driven by the characteristics of the respondents. This raises the question: What if there was a Trump supporter with a demographic profile that is similar to the profile of a 'typical' Biden

supporter, and vice versa? Would that similarity alter the magnitude of the treatment effect? To explore this, we use a propensity score (PS) approach. We first estimate the PS of being a Trump supporter, based on a Probit regression model with outcome variable being whether a person is a Trump or Biden supporter, and explanatory variables: *gender*, *age, education, race, urban status, religion, ideology,* and *interest in current events*. For brevity, the results of this stage are not presented here, but we note that all variables apart from the gender dummy have a statistically significant effect on the probability of being a Trump's supporter (i.e., they meaningfully differentiate between Trump and Biden supporters). The distribution of this estimated PS by vote intention (i.e. Trump vs Biden supporter) is presented in Fig.6A.

Fig 6A: Histogram of Propensity Score by Vote Intention

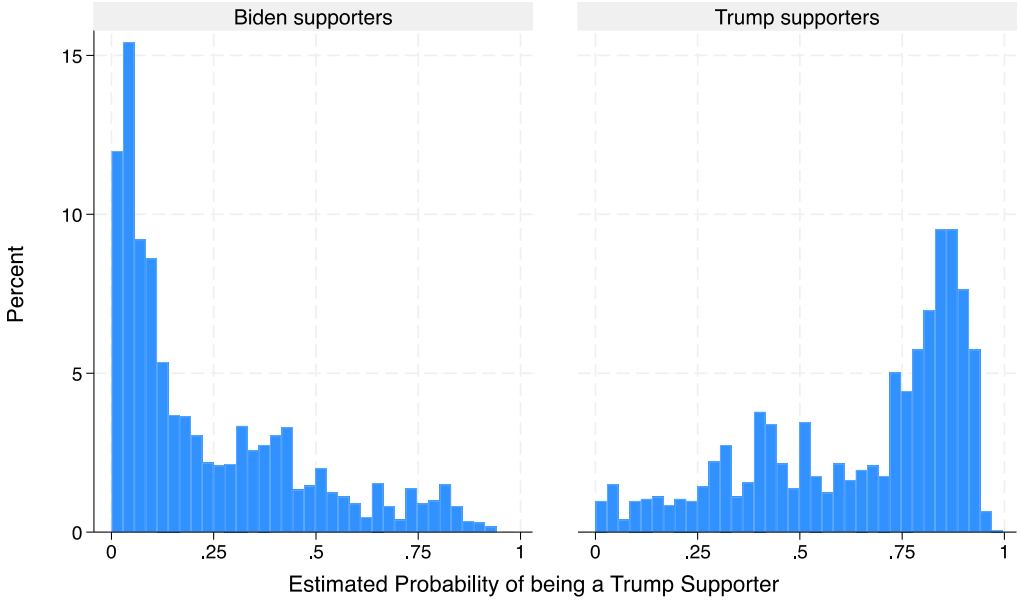
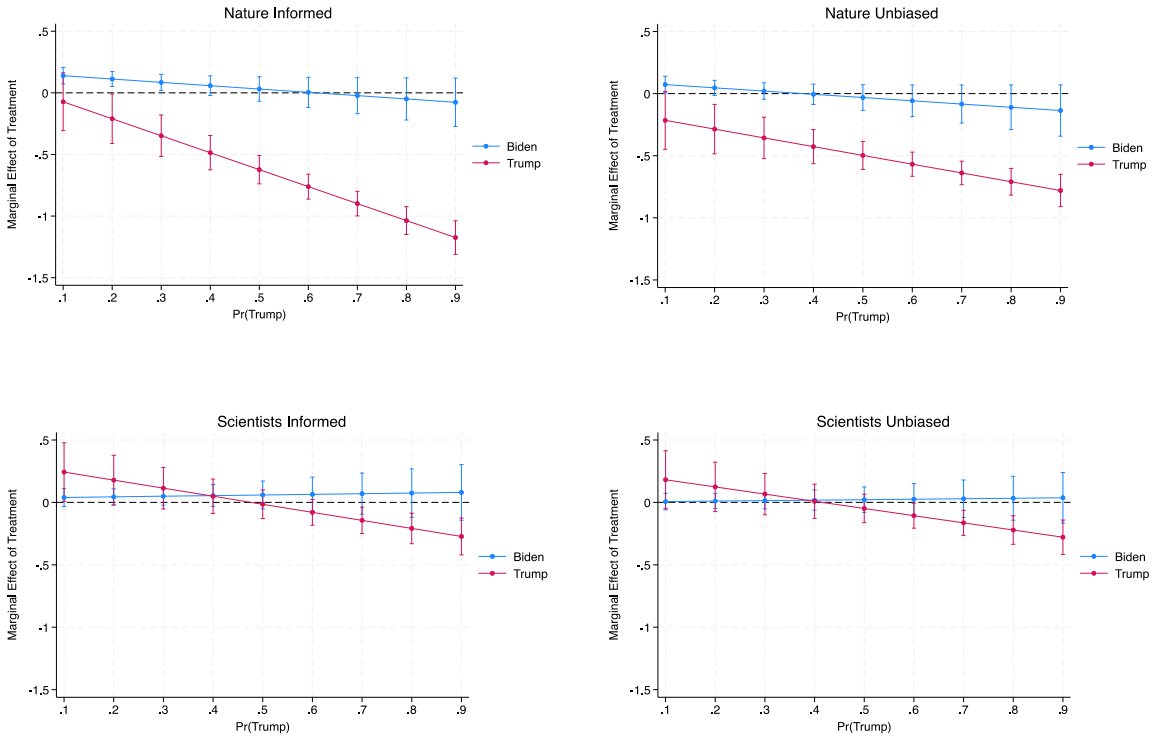


Fig.6B depicts the effects of the treatment by intention of voting as this PS increases from 0.1 (i.e., a person exhibiting characteristics that make them very unlikely to be a Trump supporter) to 0.9 (i.e., a person exhibiting characteristics that make them very likely to be a Trump supporter). For the 'trust in *Nature*' outcomes, we notice that indeed, to some extent, the effect of the treatment becomes more negative as this PS increases. For Biden supporters, the effect is positive and significantly different to zero only for those with PS near 0.1 to 0.3, while for Trump supporters the effect becomes significant for PS higher than 0.1. Moving to the more general 'trust in US scientists' questions, the moderating effect of PS is not statistically different from zero for Biden supporters, while for Trump supporters, the effect increases in magnitude as the PS gets higher. Trump supporters who resemble a Biden supporter's profile actually show a positive, but statistically insignificant, effect, while the effect becomes negative and significant for PS of about 0.6 or higher. We conclude that, to some extent, indeed, the treatment effects are driven by the demographic profile of the person, and it is likely that this identified effect could be even stronger in the presence of a richer set of demographics that would do an even better job distinguishing voting intention.

Fig.6B: Treatment Effects by Propensity Score of Being a Trump Supporter

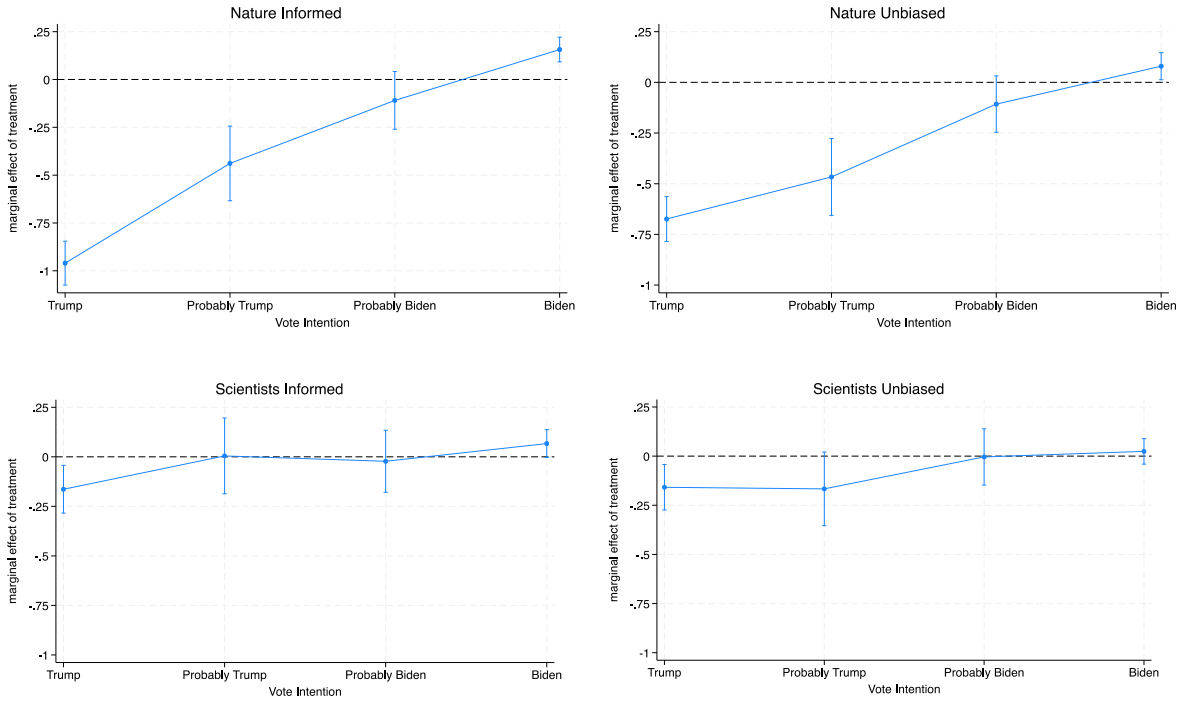


4.2.4 By type of 'voting intention'

4.2.4.1 By certainty of 'voting intention'

In this paper, the author uses the respondents' *current* vote intention, i.e., whether they would vote for Trump or Biden if there was an imminent election. However, the answer allowed respondents to express a *magnitude* of this intention, by stating whether they would *"definitely"* or *"probably"* vote for either candidate. Since the author grouped the probable and definite voters of each candidate to form his main variable of interest, it is possible that the effects that he finds are driven by the definite voters. Less certain voters, i.e., probable ones, especially those favouring Donald Trump, might be more lenient in their judgement of *Nature* and US scientists. As such, we explore whether the effects differ by this magnitude of voting certainty by keeping the variable at its original, ungrouped state. Indeed, as shown in Fig.6, the treatment effects are generally stronger for definite supporters, but not to the extent that would render the author's grouping problematic. We note though that the stronger the intention to vote for Trump or Biden, the more polarizing the effect of the treatment, especially when it comes to 'trust in Nature'.

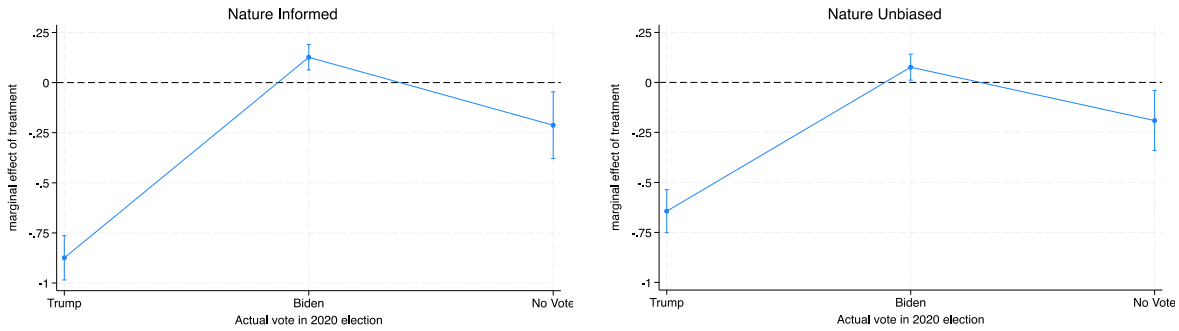
Fig.7: Treatment Effects by magnitude of voting intension

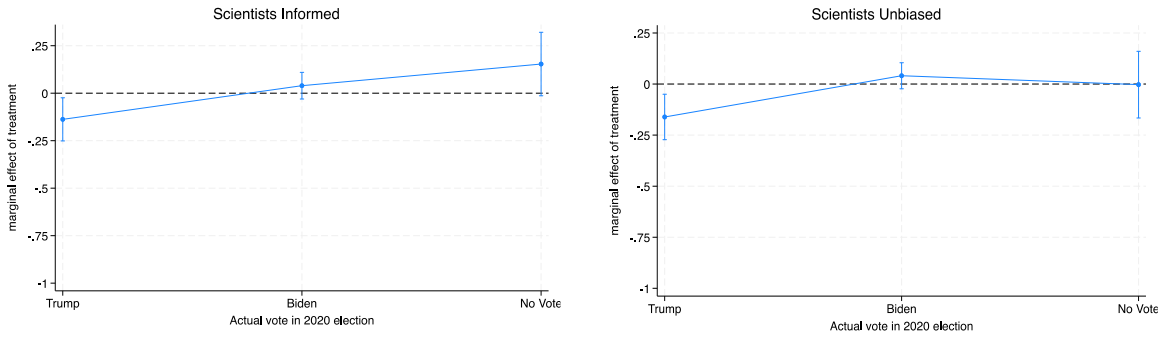


4.2.4.2 By actual 2020 elections vote

Besides the current voting intention, the author also asked respondents to declare whom they voted for in the 2020 elections. What if *actual* voting behaviour was used to capture support for Biden or Trump instead of voting *intention*? As shown in Fig.7A, the results are quite similar overall to baseline: the treatment effect is large and negative for Trump voters, and small and positive for Biden voters. Interestingly, the treatment effect for respondents who did not vote in the 2020 election is also negative for the 'trust in *Nature* questions'.

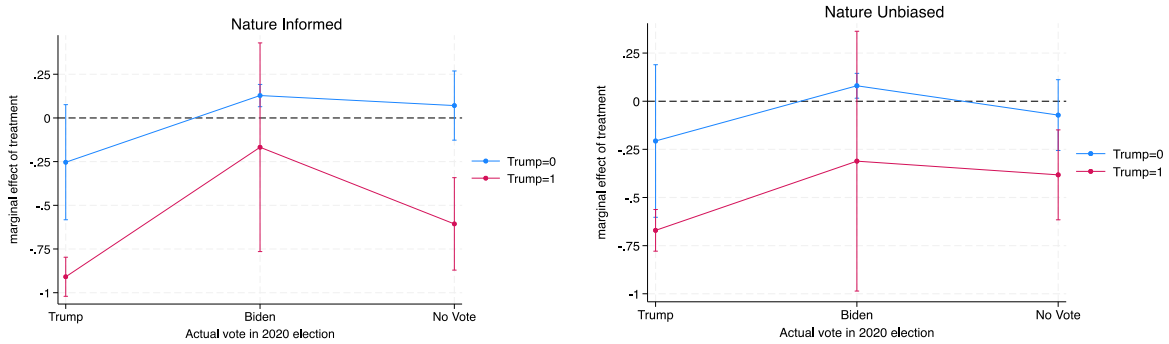
Fig.8A: Treatment Effects using actual vote in 2020 to capture support





Presumably, this effect for 'no voters' might be driven by those who intend to vote for Donald Trump. To explore this result, we further split the 'no voters' by their current vote intention. As shown in Figure 8B, the negative effect is driven by subjects who currently (at time of the experiment) intend to vote for Trump.

Fig.8B: Treatment effects for 'no voters' by current voting intention



4.3 Extensions

As already mentioned, a few of the ideas we had for checking the robustness of the paper's findings depended on data that was not immediately available in the published, 'clean' dataset. We reached out to the author, who willingly shared with us the original Qualtrics 'raw' dataset. We thus consider the three following exercises as extensions.

4.3.1 Request for information

In the section *"Demand for Information"*, the author reports the finding that the endorsement statistically significantly reduced the proportion of Trump supporters requested COVID related information from the journal *Nature* (column 1 of Table 3 below). This analysis is based on a question that asked respondents to indicate whether they would be happy to receive links for easy-to-read articles from different sources, including *Nature*, *Mayo Clinic*, other media websites, or not read at all. The author allowed respondents to pick all four options if they wished, so, using the raw data and for each analysis presented below, we split respondents on the basis of whether they pick the option of interest or not (leading to four dichotomous outcomes that were analysed using a linear probability model).

In the paper, the author only considers requests for information from Nature. However, the author did not examine whether this reflects a general growing dismissal of scientific outlets, which, for policymakers, is what would really matter. If the endorsement led to a general dismissal, we would expect to see a similar negative effect for demand for articles from *Mayo Clinic*. Our results do not support this story. We find that for Trump supporters, the treatment effect on the probability of

requesting *Mayo Clinic* articles is not statistically different to 0. Interestingly, for Biden supporters, the probability of requesting *Mayo clinic* articles reduced significantly (p-value < 0.01), suggesting a possible substitution effect (from *Mayo clinic* to *Nature*). Moreover, the treatment does not seem to influence the probability of respondents selecting information from 'other' outlets.

Finally, when it comes to the option of 'no information', the treatment had no effect on Trump supporters, according to which, the endorsement did not make Trump supporters more negative towards COVID and vaccine related information. Overall, the negative effects seem to be limited to the specific outlet (*Nature*), which, from a policy perspective, is a somewhat positive conclusion.

We note that we also applied a logistic regression model instead of a linear probability one to test the validity of these findings, and the statistical results were almost identical.

Table 3: Request for information

	Requested Nature	**Requested Mayo**	**Selected other**	**No vaccine info**
CATE Trump	-.285 (.046) [.000]	.023 (.051) [.652]	-.003 (.047) [.949]	.064 (.056) [.253]
CATE Biden	.048 (.041) [.242]	-.112 (.041) [.006]	-.033 (.043) [.443]	.064 (.037) [.084]
Baseline dif.	-.386 (.045) [.000]	-.151 (.046) [.001]	-.277 (.045) [.000]	.330 (.047) [.000]
Sample Size	3885	3885	3885	3885

Heteroskedasticity robust standard errors in parentheses; *p*-values in square brackets.

4.3.2 Using the manipulation check to capture the 'Truly Treated' respondents

As mentioned in section 2, the author conducted a 'manipulation check' by including a question towards the end of his questionnaire, asking subjects to provide their view as to whether *Nature* made any explicit political statement about either candidate before the 2020 election. Although considerably more treated subjects answered correctly compared to the control group, we felt that the figures of 68.7% and 58.5% (for treated Biden and Trump voters respectively) were surprisingly low. As such, using the additional data sourced from the author, we excluded from the analysis those treated subjects who answered mistakenly, under the assumption that these individuals might have been careless or inattentive throughout the survey, thus unreliable. Those treated subjects that remain in the sample can be considered to have been 'truly treated'. Running the baseline OLS regression models in the reduced sample produced the results in Table 4.

Table 4: Treatment effects for 'truly treated' respondents

	Nature informed	**Nature unbiased**	**Scientists informed**	**Scientists unbiased**
CATE Trump	-.895 (.061) [.000]	-.818 (.055) [.000]	- .184 (.061) [.003]	-.263 (.059) [.000]
CATE Biden	.200 (.032) [.000]	.131 (.034) [.000]	.150 (.035) [.000]	.106 (.033) [.001]
Baseline dif.	-.387 (.036) [.000]	-.655 (.040) [.000]	-.756 (.044) [.000]	-.937 (.042) [.000]
Sample Size	3199	3199	3199	3199

Heteroskedasticity robust standard errors in parentheses; *p*-values in square brackets.

As expected, all treatment effects are stronger; for the 'truly treated' subjects, i.e., for those individuals who we can confidently say that they fully understood that *Nature* endorsed Joe Biden, the effect of the endorsement was more 'polarizing'. However, even though the effects for Trump supporters are larger, they are only marginally so. In contrast, all effects for Biden supporters are at least double in magnitude compared to the full-sample estimates.

We note that because it is likely that the reduced sample is not *balanced* anymore, we checked whether the estimated effects change once we control for the demographic variables available in the dataset. That is, we control for potential differences in respondents' characteristics, to make sure that the results are not driven by demographical differences between treatment and control subjects. Including these variables does not change the estimates, which remain qualitatively the same.

4.3.3 Using an Instrumental Variable (IV) approach to estimate the impact of actual knowledge on trust

As explained earlier, the idea of the paper is that trust in *Nature,* and the scientific community in general, are affected by the fact that the treatment provides participants with knowledge about the endorsement. However, as discussed above, the 'manipulation check' at the end of the survey revealed that many treated individuals failed to recognise that *Nature* endorsed Biden (some said *"don't know"* while some others surprisingly said *"No"* or that *Nature "endorsed Trump"*). Here, we use this 'manipulation check' as the explanatory variable of interest to ask *"how does gaining knowledge about the endorsement (i.e. switching from "Don't know" to "Yes, Biden") affect the outcome variables?"*. So, the regression model of interest is of the form:

$$Trust = \alpha_0 + \alpha_1 \, Knowledge \; + \; u, \tag{4}$$

where $Trust$ is either the '*Nature* Informativeness' or the '*Nature* Unbiasedness' outcome variable, $Knowledge$ is a dummy variable taking value 1 if the individual knows (by answering correctly to the 'manipulation check') *Nature* endorsed Biden and 0 if they responded with *"don't know"* (those who answered *"No"* or *"Yes, Trump"* are dropped), and u is an error term. We could estimate this model, separately for Trump and Biden supporters, to assess the impact of gaining knowledge about this endorsement on trust. However, such a regression would not identify the causal effect of 'gaining knowledge' on trust, as it suffers from omitted variable bias. For example, people with higher innate intelligence are more likely to trust *Nature*, and, at same time, it is reasonable to assume that they are more attentive as well, and therefore more likely to say that they know about the endorsement. In other words, this regression would only tell us that those who have knowledge of the endorsement have α_1 standard deviations of higher/lower trust than those who answered *"don't know"*; this does not provide us with the causal effect of gaining this knowledge.

Instead, to answer the question and overcome the endogeneity issues of this model, here we suggest using the treatment dummy variable as an instrument for 'gaining knowledge' and run a standard Two-Stage Least Squares Instrumental Variable (IV) procedure. According to this IV procedure, for the treatment to be a valid instrument it needs to be: i) *relevant* – i.e., to be strongly statistically related to the endogenous variable (with *F*-statistic higher than 10), and ii) exogenous – i.e., to influence trust in *Nature* only through its impact on gaining knowledge.

Regarding *relevance*, we note that for Trump supporters, 65.5% of those in the treatment said *"Yes Biden"* compared to 16.5% in the control group, about 49 percentage points lower (the difference is about 45 percentage points for Biden supporters). This clearly shows the shift in knowledge created

by the treatment, and it is this exogenous shift in knowledge that we exploit in our IV strategy. To test for the statistical significance of this difference, we obtain the same results through the following *first stage* regression:

$$Knowledge = \delta_0 + \delta_1 Treatment + e \quad (5)$$

Results are presented in Table 5. Clearly, the treatment had a very strong and positive statistically significant impact on $Knowledge$ for both Trump (by 49 percentage points) and Biden (by 45 percentage points) supporters, corresponding to *F*-statistics higher than 400 (in this case, the *F*-stat is just the square of the *t*-stat), comfortably fulfilling the instrument relevance condition.

Then, running a regression of $Trust$ on the OLS fitted values from (5) leads to the second-stage estimates of the IV procedure, noting also that an adjustment on the standard errors is needed. If the exogeneity condition holds, then these fitted values are not correlated with the error term in (5), and therefore, estimates of the 2nd stage reflect the causal effect of $Knowledge$ on $Trust$. Following the logic of this study, we believe that the exogeneity condition is likely to hold too; the main point of the treatment was to make respondents aware that *Nature* endorsed Biden and we agree that this is the most logical way the treatment is acting on people's perceptions.

Nevertheless, it is also important to note that, as it is known in the IV literature (see for example, Angrist and Pischke, 2009), IV estimation provides us with the causal effect of the explanatory variable of interest on the outcome, but only for those who are influenced by the instrument; i.e., it is a Local Average Treatment Effect (LATE) on the treated instead of the general Average Treatment Effect. This is the case here as well. First, those in the control group cannot be gaining knowledge due to the treatment; instead, here we assume that the participants who answered this manipulation check correctly, actually knew about the endorsement in advance, which is a reasonable assumption as the experiment was conducted 10 months after the actual endorsement. Similarly, it is reasonable to assume that, as the treatment was randomized, a similar proportion of those in the treatment group would know about the endorsement in advance as well, even in the absence of the treatment. Second, there are people in the treatment group who did not notice the treatment, and therefore were not 'truly treated' either. The IV method adjusts for both types of 'non-compliers', and, under *relevance* and *exogeneity*, the IV estimates are interpreted as the effect of gaining knowledge about the endorsement on the 'trust in *Nature'* only for those whose knowledge was affected by the treatment, which is actually the main interest of this work.

The second stage results are also presented in Table 5. If the exogeneity condition indeed holds, the second stage results, for example for Trump supporters and the '*Nature* informed' outcome, indicate that gaining knowledge that *Nature* endorsed Biden through the treatment reduces perceived informativeness of *Nature* by 1.84 standard deviations, an effect that is also strongly statistically significant.

Table 5: Results of the IV estimation (outcome standardised informativeness of Nature)

	First Stage	**Second stage**		**Sample Size**
		Nature Informed	**Nature Unbiased**	
Trump supporters	.490 (.023) [.000]	-1.841 (.145) [.000]	-1.357 (.121) [.000]	1320
Biden supporters	.451 (.019) [.000]	.233 (.072) [.001]	.100 (.073) [.171]	2098

Heteroskedasticity robust standard errors in parentheses; *p*-values in square brackets.

For both Trump and Biden supporters, the estimated coefficients of this second stage are in line with, but more than double, the effects of the treatment estimated in the paper. This is not a surprise, given that here we purely capture those who switched knowledge status from 'not-knowing' to 'knowing' about the endorsement *due to the treatment*, while the results of the paper are confounded by people who already knew about the treatment and those who were exposed to treatment but actually did not notice it. We finally note that the results hardly change if we control for any demographic differences across participants, and this was to be expected as well, as the treatment was randomised and therefore not expected to be correlated with unobservables either in the first or the second stage equations.

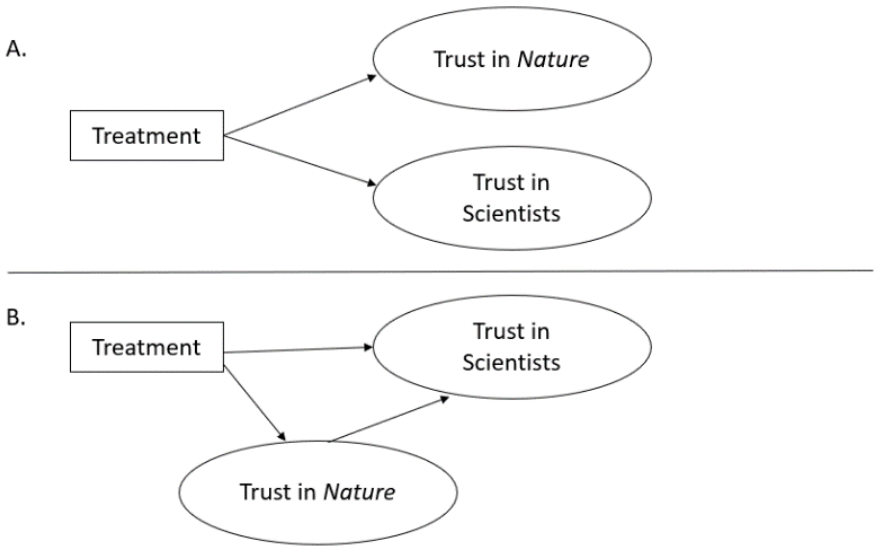
4.3.4. Mediation

Interestingly, the author runs separate models for 'trust in *Nature*' and 'trust in US scientists', effectively assuming that the responses that subjects give to the questions about trust in US scientists are not dependent on their responses to the questions about trust in *Nature*, a major scientific publication. However, because the questions about US scientists follow in order the ones about *Nature*, one can reasonably argue that, to some extent, the first are driven by the second; the set of questions about the specific outlet (*Nature*) might have primed the respondents to answer in a similar manner to the second set of more general questions about scientists. This could be investigated if the order of these questions had been randomized across survey respondents, and the author had provided information about which question came first for each participant. However, in this instance, our understanding is that the order of the questions was not randomized. Moreover, and independent of the aforementioned, the response of someone who is asked if they believe that US scientists are unbiased, might depend on their belief as to whether a major scientific publication (that is co-edited by Americans and where American scientists contribute significantly) is unbiased; the two outcomes are, at least to some extent, conceptually dependent. As such, one can hypothesize a causal relationship between them, in the sense that a respondent's 'trust in *Nature*' partially causes 'trust in US scientists'. If this is the case, only part of the treatment effect on 'trust in US scientists' will be *direct*, while the underlying assumption of the baseline analysis by the author is that the *total* effect is direct.

Fig.9A illustrates, using a causal diagram, what the author effectively assumes and tests; 'trust in *Nature*' and 'trust in science' being independently affected by the treatment. Fig.9B presents a different causal 'story'; 'trust in science' being affected both directly and indirectly (through 'trust in *Nature*') by the treatment. In other words, part of the treatment effect on 'trust in science' is *mediated* through 'trust in *Nature*'. In practical terms, one can think of this dynamic in the following manner: some subjects might change their levels of 'trust in science' directly due to *Nature's* endorsement of Joe Biden (the direct effect of the treatment), while other subjects might decrease or increase their 'trust in science' *because of* (or after) increasing or decreasing their 'trust in Nature' (the *indirect* effect of the treatment). As such, the total treatment effect on 'trust in science' is a combination of the direct and indirect effects. In sum, in what follows, we propose and test a causal mediation model. Causal mediation has a long history in psychology and political science (e.g., Baron and Kenny, 1986; Imai et al., 2010), and has now been formally incorporated in Stata through the 'mediate' routine. The approach to mediation applied here is based on the potential-outcomes framework (Ngyuen et al., 2021). What follows is primarily based on the very informative Stata handbook[3] that details the theory and practice of causal mediation.

[3] https://www.stata.com/manuals/causalmediate.pdf

Fig.9: Causal diagrams



Mediation analysis effectively comprises two linear regressions: one to estimate the effect of the treatment on the mediator ('trust in *Nature*') – what we call the 'mediator-treatment equation' – and one to estimate the effect of the mediator on the outcome ('trust in science') holding the treatment fixed, what we call the 'outcome-mediator equation'. The indirect effect is generally calculated by multiplying the two estimated effects, and the direct one by subtracting the indirect from the total treatment effect. In the causal mediation framework, the outcome-mediator equation also includes the interaction between the treatment and the mediator, thus allowing for the effect of 'trust in *Nature*' on 'trust in science' to differ between the treatment and the control groups. Our analysis also includes appropriate 'control' variables, consisting of all available demographic characteristics in both the outcome and mediator equations, and the manipulation check variable in the outcome equations only (explaining why later in this section), so our estimates of the total effects will not match the author's reported estimates. Because in this exercise we are simply interested in exploring the 'breakdown' of the total effect of the endorsement (the treatment) on 'trust in science' (the outcome) into the direct effect and indirect effect (through 'trust in *Nature*' the mediator) we follow 'best practice' and decompose the total effect (or *Average Treatment Effect* – ATE) in two ways.

Decomposition 1 separates the direct effect under the untreated mediator condition from the total indirect effect (see Nguyen et al., 2021). We are interested in this decomposition because we expect that the endorsement of Joe Biden has a direct effect on 'trust in science' but want to determine whether a portion of the total effect can be attributed to a change in 'trust in *Nature*', and if so, how much of the total effect is due to mediation. Under this decomposition, the *Natural Direct Effect* (NDE) is the average direct effect of the endorsement on 'trust in science' when 'trust in *Nature*' is held at its value associated with the control group (i.e., those subjects who did not see the endorsement). It is calculated as Y [1, M (0)] − Y [0, M (0)], where Y denotes 'trust in science' and M denotes 'trust in Nature' (please refer to Table 6 for definitions of this notation). The *Natural Indirect Effect* (NIE) on the other hand, estimates the average indirect effect of the endorsement through 'trust in *Nature*'; that is Y [1, M (1)] − Y [1, M (0)]. The ATE, NIE and NDE for the two outcomes reflecting 'trust in science' are presented in the Decomposition 1 column of Table 7, separately for Trump and Biden supporters.

Decomposition 2 separates the indirect effect under the untreated condition from the total direct effect (see Nguyen et al., 2021). We are interested in this decomposition because one might argue that the total effect of the endorsement on 'trust in science' is indirect (through 'trust in *Nature*') because the subjects have effectively been primed through the research design (due to the order of the questions) to 'trust' or 'distrust' US scientists. Credibly, an interlocutor might question this argument and claim that there must also be a 'natural' direct effect. Under this decomposition, the *Pure Natural Indirect Effect* (PNIE) is the average indirect effect of 'trust in *Nature*' under the control condition (i.e., not being exposed Biden's endorsement). It is the difference $Y[0, M(1)] - Y[0, M(0)]$. The *Total Natural Direct Effect* (TNDE) on the other hand, estimates the average direct effect of the endorsement when 'trust in *Nature*' is held at its value associated with being exposed to the endorsement; that is $Y[1, M(1)] - Y[0, M(1)]$. The ATE, PNIE and TNDE for the two outcomes reflecting 'trust in science' are presented in the Decomposition 2 column of Table 7, again separately for Biden and Trump supporters.

Table 6: Notation for causal mediation

$Y[1, M(1)]$	*"The population-average value of the outcome that would be expected if everyone was treated"*, i.e., the population-average 'trust in science' if everyone had seen Biden's endorsement.
$Y[0, M(0)]$	*"The population-average value of the outcome that would be expected if nobody was treated"*, i.e., the population-average 'trust in science' if nobody had seen Biden's endorsement.
$Y[1, M(0)]$	*"The expected value of the outcome when everyone is treated but counterfactually experiences the value of the mediator associated with being untreated"*, i.e., the expected value of 'trust in science' had everyone seen Biden's endorsement but counterfactually demonstrated the level of 'trust in Nature' associated with not having seen the endorsement.
$Y[0, M(1)]$	*"The expected value of the outcome when everyone is untreated but counterfactually experiences the value of the mediator associated with being treated"*, i.e., the expected value of 'trust in science' had nobody seen Biden's endorsement but counterfactually demonstrating the level of 'trust in Nature' associated with having seen the endorsement.

Note: Definitions are quotes verbatim from the entry for *mediate* in the Stata manual

Table 7: Decompositions of the total treatment effect on 'trust in US scientists'

		Decomposition 1		**Decomposition 2**	
	ATE	**NIE**	**NDE**	**PNIE**	**TNDE**
			Scientists Informed		
Trump	-.063 (.058) [.278]	-.360 (.034) [.000]	.297 (.056) [.000]	-.491 (.045) [.000]	.428 (.063) [.000]
Biden	.066 (.034) [.052]	.058 (.015) [.000]	.008 (.032) [.803]	.049 (.013) [.000]	.017 (.031) [.583]
			Scientists Unbiased		
Trump	-.157 (.052) [.003]	-.356 (.033) [.000]	.199 (.046) [.000]	-.407 (.036) [.000]	.250 (.048) [.000]
Biden	.023 (.031) [.458]	.028 (.014) [.046]	-.005 (.028) [.858]	.026 (.013) [.046]	-.003 (.027) [.912]

Robust standard errors in parentheses; *p*-values in square brackets.

In this commentary, we present the results using causal language (i.e., attributing a causal interpretation to the mediation model), and assess the credibility of the causal claims afterwards, by evaluating the assumptions of a causal mediation model.

For Trump supporters, the results suggest that the indirect effect of the treatment on 'trust in science' (through changes it causes on 'trust in *Nature*') is negative and strongly statistically significant. Evidently, the indirect effect is negative because a negative coefficient (effect of treatment on '*Nature* informed') is, in the background, multiplied with a positive one (effect of '*Nature* informed' on 'Scientists informed'). Then, as the total treatment effect is close to zero[4], the resulting direct effect is positive and strongly statistically significant. A similar situation holds for perception of the unbiasedness of US scientists, with an equally large indirect effect, but a slightly smaller direct effect (as the total effect in this case is a bit larger and statistically significant). The positive direct effects suggest that some treated Trump supporters, presumably those who might be more open-minded to new information from appropriate experts (such as the journal *Nature*), increase their perception that US scientists are informed and unbiased.

For Biden supporters, the total effect 'scientists informed' is similar in magnitude to Trump's but positive, and due to the higher precision of the estimate (Biden supporters are the largest sub-sample) it ends up being significant at the 10% significance level. Importantly, we also note that most of this effect is through the indirect path: Biden supporters, on average, increase their trust in *Nature*, which then leads them to increase their trust in science. There is no evidence of a direct effect on 'scientists informed'. A similar result holds for 'scientists unbiased' but with an indirect effect that is smaller in magnitude.

However, to attribute a causal interpretation to the estimates of the causal mediation model, we need to evaluate the assumptions that need to hold. Since the treatment assignment is randomized, we do not worry about unobserved confounding in the treatment-outcome and treatment-mediator relationships. What we need to worry about though is confounding in the mediator-outcome relationship (i.e. the regressions of 'trust in science' on 'trust in *Nature*'); it is reasonable to believe that individual characteristics, besides the treatment, are correlated with both 'trust in *Nature*' and 'trust in science' and therefore affect the estimated relationship between the two. For example, education might be correlated with both 'trust in *Nature*' and 'trust in science', as more educated subjects are more likely to exhibit higher levels of trust in experts. For this reason, and as mentioned, we included all observed covariates (the ones that were used to calculate the propensity score in section 4.2.3) in the regression of the outcome: gender, age, education, race, urban status, religion, ideology, and interest in current events. Despite all this, even though inclusion of such characteristics reduces some of the bias, there may still be unobserved characteristics associated with both 'trust in *Nature*' and 'trust in science', i.e., the second 'leg' of the indirect effect, suggesting that the relationship between the two is still 'correlational' and cannot be given a causal interpretation. For example, subjects with high innate intelligence (an unobservable) are more likely to trust both *Nature* and US scientists, generating a positive correlation between 'trust in *Nature'* and 'trust in science', even holding all demographics fixed. Crucially, if due to such unobservable characteristics, this estimated effect is considerably upward biased, the indirect effect will be considerably inflated in terms of magnitude (i.e., more negative). This, in consequence, would suggest that the direct effect would also be considerably positively biased (since it is the difference between the total effect and the indirect effect), resulting in both effects being statistical artifacts. Although we cannot preclude this

[4] Note that this differs to the baseline result of the author (-.130 standard deviation units) entirely because of the inclusion of the covariates.

possibility, we argue that the 'true' relationship between trust in *Nature* and trust in science is strictly positive, so the respective coefficient should always be positive, even if all relevant unobserved confounders were somehow observed. Because the effect of the randomly assigned treatment on 'trust in *Nature*' is unbiased and negative, the indirect effect of the treatment on 'trust in science' should be strictly negative. Consequently, the direct effect of the treatment on 'trust in science' for Trump supporters will always be smaller in magnitude than the total effect (i.e., less negative), or more likely, positive like in this exercise. This would reflect a fraction of Trump supporters who are swayed by the endorsement towards a more positive (or less negative) perception of US scientists.

The last assumption is that there are no confounders in the mediator–outcome relationship that are *caused* by the treatment. Are there any such variables? First, drawing from earlier discussion, most treated subjects get to actually *know* that *Nature* endorsed Joe Biden. This knowledge that a major scientific publication considers Biden to be a more appropriate president (and Trump less appropriate) might prompt a 'rational' Bayesian updating of beliefs, which leads to correlated responses to the 'trust in *Nature*' and 'trust in science' questions. Second, seeing the endorsement might have generated positive or negative emotions to subjects, which affect their responses to all subsequent questions. For example, for Trump supporters, the endorsement of a rival by *Nature* might generate negative emotions towards the 'liberal scientific elite' and a consequent 'irrational' reaction against both the journal itself, and scientists in general.

To address the first concern, as a proxy for actual knowledge, we include in the outcome-mediator equation the answer to the manipulation check, which asked respondents to indicate whom they thought Nature endorsed in the run-up of the elections. When it comes to the second concern, we argue that, to a considerable extent, the emotional response will depend on which candidate the treated subject supports. Since we conduct the mediation analysis separately for Trump and Biden supporters, this concern is partly addressed. In addition, the emotional response of the subject might also depend, to some extent, on the demographic characteristics that we observe. However, it goes without saying that these attempts do not preclude the possibility that the treatment has indeed caused an unobserved confounder, undermining the causal nature of the model.

In conclusion, although we cannot argue that we solved the problem through the mediation analysis, this exercise shows that it is likely that the effects presented for 'trust in US scientists' of Trump supporters might primarily be indirect, and possibly, partly driven by the fact that this set of questions followed from the main question about 'trust in *Nature*'. We do not intend this to undermine the author's analysis or claims; rather, to ameliorate somewhat the assertion that a political endorsement by a scientific publication can undermine the public's trust in science in general, and to call for further research in the area with alternative research designs. If political endorsements by scientists, or scientific journals, have indeed a robust effect on the perceived trust of a large fraction of the public in one of the fundamental pillars of modern society, i.e., science, then this has important implications for how scientists should communicate their political beliefs, and how policymakers should treat such communications. We hope researchers in the relevant fields of study will take up this task.

5. Conclusion

In this report, we reproduced and replicated the results of the paper *"Political endorsement by Nature and trust in scientific expertise during COVID-19"* by Floyd Jiuyun Zhang published in *Nature Human Behaviour.* In summary, this work corroborates the author's analysis and supports his main causal claim that electoral endorsements by scientific publications can undermine public trust in the endorser, and

potentially in science in general, among supporters of the out-party candidate (in this case, Donald Trump). In fact, some of the exercises we undertook, strengthen the paper's main results regarding the treatment effects on the subjects' 'trust in *Nature*'. Furthermore, through our heterogeneity analysis we add interesting nuance to the findings. However, we believe that some of our tests show that the effects of the endorsement were predominantly 'local', in the sense that they affected the subjects' overall perception of *Nature*, the specific publication that made the endorsement. The evidence that the endorsement's effect 'transfers' to a general distrust in other scientific publications and the scientific community in general is somewhat weaker. This does not undermine the author's main causal claim, but, we think, somewhat ameliorates the strength of the second part of the claim. We hope that the replication exercises, critical remarks, and open questions, discussed in this report, spark further research in this very important topic.

Acknowledgements

We would like to thank the organizers of the 'Norwich Replication Games' and the Institute for Replication for giving us the opportunity to engage with this paper and challenge ourselves. We would also like to express our gratitude to Floyd Zhang for his responsiveness and willingness to share his raw data. All errors in this analysis, code and report are our responsibility.

References

Angrist, J.D. and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baron, R.M., and D.A. Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51: 1173–1182.

Bloem, J.R. 2022. How much does the cardinal treatment of ordinal variables matter? An empirical investigation. Political Analysis 30(2): 197-213.

Imai, K., L. Keele, and D. Tingley. 2010. A general approach to causal mediation analysis. Psychological Methods 15: 309–334.

MacKinnon D.P., J.L. Krull and C.M. Lockwood. 2000. Equivalence of the mediation, confounding, and suppression effect. Prevention Science 1:173–81.

Nguyen, T.Q., I. Schmid, and E.A. Stuart. 2021. Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. Psychological Methods 26: 255–271.

Webb R.M. and L. Kurtz. 2022. Politics v. science: How President Trump's war on science impacted public health and environmental regulation. Progress in Molecular Biology and Translational Science 188(1):65-80.

Zhang, F.J. 2023. Political endorsement by *Nature* and trust in scientific expertise during COVID-19. Nature Human Behaviour 7:696–706.

Appendix A: A response to Zhang's comment on this report

We thank Dr Floyd Zhang for reading our replication report and providing constructive comments (published alongside this report). We are grateful to him for giving us credit for this work. In what follows, we provide a response to the main concerns he raised.

Dr Zhang's entire concern seems to be with the mediation analysis (section 4.3.4 of the report). His argument is clear and consists of three interconnected claims:

1. First, he explains that the original analysis (and results) in his *Nature* paper, when it comes to the effect of the treatment on the subjects' 'trust in scientists' is *"consistent with it [the effect] being mediated by the effect on 'trust in Nature'"*. Relatedly, he claims that *"the causal model behind this claim implies this mediation"*. The rationale he provides for this is *"the idea that participants make generalizations. If the treatment increases (decreases) one's trust in Nature as the result of learning about the endorsement, one infers that other scientists are also more (less) likely to be trustworthy, as Nature is considered exemplar of the broader scientific community"*. As he mentions, behavioral models assuming either a Bayesian agent considering *Nature* as a representative outlet providing information about the population of scientists, or a non-Bayesian agent with a cognitive tendency to stereotype, are consistent with this.
2. He then claims that he can conceive of *"no plausible behavioral model that would rationalize a large direct effect on trust in scientists in general, un-mediated by 'trust in Nature'"*, and as a consequence, the mediation analysis does not serve in discriminating between the implied model and any alternative one.
3. The ultimate implication, according to Zhang, is that the mediation analysis *"is not suggestive that the endorsement effect is 'local' to the endorser nor does the result alleviate the policy concern that such endorsement can affect the credibility of the broader scientific community"*.

Overall, our impression is that our disagreement primarily concerns semantics, and not the substance of his and our analysis. In what follows, we provide a detailed response, including some additional descriptive results after 'digging' in the data again. We thank again Dr Zhang for giving us the opportunity to revisit our work and further engage with his dataset.

When it comes to Dr Zhang's first claim above, we have no reason to disagree. Sure, this is exactly what should be happening: the subject adjusts their belief on the trustworthiness of scientists after adjusting their belief on a prestigious and representative scientific journal (i.e., *Nature*, the endorser). However, Zhang's analysis *"implies"* this process, while our causal mediation framework (with its inherent limitations that we have discussed) *explicitly* tests it. Zhang's implied model and analysis, and our causal mediation framework with its results then agree: the originally presented effects of the endorsement on the trustworthiness of scientists, for both Trump and Biden supporters, are 'indirect' (through 'trust in *Nature*').

Before we move to evaluate Dr Zhang's second claim above, let us note that in the mediation framework, a 'direct' effect does not mean that there is absolutely no other variable in the respective causal path, i.e., between the treatment and 'trust in scientists'. It only means that we do not theorize on, and/or observe any such variable in that causal path. So, we might be calling it 'direct' in this framework, but there *could* exist other mediating variables (other than 'trust in Nature') that could explain the effect. It is also worth noting Zhang's second claim relates only to Trump supporters; for Biden supporters, only the indirect effects are statistically significantly different to zero (and positive), accounting for the Average Treatment Effect (ATE) entirely (the direct effect is practically zero).

Taking a step back, part of the rationale for conducting the mediation analysis was the interesting (but maybe not adequately discussed) side-finding of the original study that, particularly for Trump supporters, the negative effect of the treatment on 'trust in *Nature*' does not *fully* 'transfer' to 'trust in scientists'. This is what both Zhang's and our analyses suggest. If this effect fully 'transferred', we would observe a comparable negative effect on the two outcomes (perceived informativeness and unbiasedness) for both *Nature* and scientists in general. Instead, the observed effect for the *Nature* related outcomes is orders of magnitude larger (i.e., more negative) than for scientists in general; -0.854 versus -0.130 standard deviation units for informativeness, and -0.633 versus -0.161 for unbiasedness. What explains this discrepancy *empirically*?

An obvious reason would be that many Trump supporters might consider it a bit 'far-fetched' to radically adjust (downwards) their perception of the trustworthiness of scientists in general, just because a single scientific outlet endorsed Trump's opponent. Hence, and compared to the control group, treated Trump supporters might give a very low score ('1' or '2', i.e., "not informed" or "not unbiased") when answering the questions about *Nature*, followed by 'milder', i.e., less negative responses to the two questions about scientists in general.

But can this fully explain the observed 'asymmetry' in the two sets of effects (*Nature* versus scientists)? The mediation analysis suggests otherwise (i.e., the 'direct' effect of the treatment is of the opposite sign). So, is it possible that a set of treated Trump supporters (granted, driven by certain and possibly shared unobserved characteristics) adjust their perception of trustworthiness of Nature downwards, *but their perception of the trustworthiness of scientists upwards*? In other words, is there a set of Trump supporters who, due to Nature's endorsement of Biden, *lost their trust in the endorser* but *increase their trust in scientists* more generally? Equivalently, can we conceive of a reason why there might exist a positive direct effect of the endorsement on 'trust in science' for Trump supporters? Dr Zhang mentions that he *"cannot conceive of a behavioural model*" to explain this. Maybe he is right and there is no such behavioural model. Or maybe there is a behavioural model which does not assume a rational agent applying Bayesian updating to new information, etc. We are not experts in this literature, so we do not draw from theory. But we can draw from our intuition and some descriptive findings from the dataset.

Intuitively, we think that one can easily conceive of such subjects. It is those Trump supporters who might equate 'scientists' with 'other scientists apart from the editors of *Nature*' (scientists potentially backing Donald Trump and his policies, scientists appearing in conservative news outlets, etc.). Basically, subjects who, due to the endorsement, emotionally respond with an attitude "I do not trust *Nature* because it has been captured by the 'liberal elite', but I do trust 'other' scientists, those honest, independent ones". These subjects might see this endorsement as an opportunity to express their disapproval to a mainstream scientific outlet, but simultaneously, reaffirm and reassert their self-image of a rational, science-abiding American (just not the science of the scientists in *Nature*).

Interestingly, basic descriptive analysis supports our intuition: First, among the 770 treated Trump supporters, 82 of them answered "not informed" (i.e., 1 or 2) in the '*Nature* informed' question, yet still answered "informed" (i.e., 4 or 5) in the 'scientists informed' question. In comparison, among the 766 untreated Trump supporters, only 10 answered similarly (and in fact, all of them answered a 4). We observe similar, though somewhat weaker, patterns for the 'unbiased'-related questions. That is, of the treated Trump supporters, 55 answered "not unbiased" in the '*Nature* unbiased' question but still answered ''unbiased" in the 'scientists unbiased' question, while only 16 untreated supporters answered similarly.

Overall, intuition and some descriptive findings suggest that there is a set of Trump supporters who, due to our inability to control for other unobserved characteristics, make the positive 'direct' effect of the endorsement on Trump supporters' 'trust in scientists' seem entirely plausible (but see caveat about 'direct' effect earlier in this section). Whether there is a behavioural model to explain this is beyond the scope of this report but constitutes an interesting further research direction.

Finally, Dr Zhang's third claim suggests that he disagrees with our conclusion that *"the endorsement remained predominantly 'local', in the sense that it primarily affected the subjects' overall perception of Nature"*. It is not clear whether this disagreement is solely due to his objections to our causal mediation analysis. If he disagrees with the statement altogether, we would like to emphasize that this conclusion was based on other results as well. First, as mentioned earlier, the treatment effects are much smaller in magnitude on the 'trust in scientists' measures than the 'trust in *Nature*' measures (6.5 times for informativeness and almost 4 times for unbiasedness). Second, in section 4.3.1, we show that the endorsement did not make Trump supporter less reluctant to request articles from Mayo clinic, request articles from other outlets, or request information regarding the COVID-19 vaccine generally. We thus stand by this statement, and have not removed it from the Conclusion section, where its premises were clearer. Nevertheless, toward reconciliation, and considering that many readers might solely rely on the abstract to evaluate the success of this replication exercise, we have adjusted the statement that Dr Zhang quoted in his response. It now reads as follows: "*Overall, and for Trump supporters in particular, our report corroborates the author's main finding of a strong negative effect of the endorsement on the overall perception of the endorser (Nature). However, the additional analysis provides weaker evidence for a reduction in trust in the scientific community more generally*". We hope that this can be considered a fair summary of our results.