How 'Free' is Free Speech in Academia? Effects on Researchers and their Research

Kris Gulati [a] and Lorenzo Palladini [b]

[a] Berkeley Haas and University of California: Merced
[b] University of Luxembourg

June 1, 2024

ABSTRACT

Freedom of speech plays a crucial role in science and the pursuit of knowledge. However, recent evidence indicates a *decline* in such freedom in the academy with rising calls for sanctions against scholars who make statements or express their opinions about matters of public interest that are deemed controversial. Using a novel dataset, we examine the effect of these incidents using a series of difference-in-difference designs. We find that the affected scholar's body of work, published prior to the incident, receives 4 percent fewer citations after the incident. Affected scholars also become less productive after the incidents, publishing 20 percent fewer works and receiving 14.5 percent fewer citations than scholars with similar characteristics. Considering that scholars at the same institution as the affected scholars and scholars within their coauthorship network tend to cite the affected scholars' work less, our results seem to be motivated by scholars seeking to distance themselves. Leveraging a large language model (GPT4) to assist with the classification of the rich qualitative data, we find that the citation penalty remains for both incidents involving subjects within the scholar's academic field and outside, and whether or not the incidents are based on statements classified as hate speech or not. Our results oppose the Mertonian norm of universalism, which suggests that scientific notions must be evaluated independently from the personal opinions of the scholars proposing them.

KEYWORDS: Political Economy, Politics of Science, Innovation, Economics of Science

*Kris Gulati thanks Emergent Ventures for financial support. We both thank FIRE for their financial support. We are deeply grateful to Komi Frey, Katrin Hussinger, Andrew Johnston, Abhishek Nagaraj, Michael Rose, Rajiv Sethi, Matteo Tranchero, and Joshua Graff Zivin for their helpful comments and suggestions. We are grateful to the audiences at The Max Planck Institute, University of Bordeaux, and The University of California: Merced for their feedback. We thank Christian Venezuela for his outstanding research assistance. Corresponding author: please email kgulati@berkeley.edu

"Freedom of speech is one of science's most important norms. According to [John Stuart] Mill, social and scientific progress occurs through vigorous debate involving opposing points of view. To generate different points of view, people must have freedom of thought and speech. Progress cannot occur if the majority uses its power to suppress minority viewpoints." Resnik (2008, p.31).

1. Introduction

In 2019, University of California Davis mathematics Professor, Abigail Thompson, published an op-ed in the *Wall Street Journal* comparing mandatory diversity statements for job applicants to mandatory loyalty oaths that her university required seventy years earlier. The reaction was immediate. Hundreds of colleagues from across the country made or signed statements calling her view "dangerous" and petitioning her employer to investigate whether her behavior was consistent with their institutional policy (Smith 2019; Soucek, 2021).

In 2020, another mathematics Professor, Andrea Bertozzi (UCLA), faced calls for sanction for her work on "predictive policing". An invited lecture to honor Bertozzi's "contribution to the mathematical sciences" was cancelled because the organizers "do not believe that mathematicians should be collaborating with police departments" (Castelvecchi, 2020).

Social sanctions for both 'extramural' speech, like that of Thompson, and speech constituting part of 'academic freedom', like that of Bertozzi, are increasingly common.[1] Frey and Stevens (2023) find that the annual number of attempts to suppress or punish scholars' speech in the U.S. has increased dramatically over time and that 90 of the top 100 universities in the US have experienced events such as these. A recent survey of university faculty showed that 40% of liberal faculty, 56% of moderate faculty, and 72% of conservative faculty fear losing their jobs or reputations for something they say aloud or post online (Honeycutt et al., 2022). Similarly, another survey of academics found that 26% reported they were not free to engage in the research of their choice, suggesting tangible consequences for the rate and direction of scientific inquiry (F. S. Union Tech. Rep., 2023).

The increasing number of cases and changes to the academic climate have prompted vigorous debate and action within universities and by policy makers. For example, in 2023 the United Kingdom passed the

[1] Extramural speech refers to scholars expressing their opinions on subjects outside their field of expertise (for example, a biology scholar expressing her views on political topics); academic freedom, instead, refers to scholars expressing their views within their academic context (for example, a political science scholar expressing her opinion on political matters that might arise from her scholarly activity).

Freedom of Speech Bill putting onus on universities to protect the freedom of speech of academics. Professor Arif Ahmed, Director for Freedom of Speech and Academic Freedom of the Office for Students, a non-departmental public body for the Department for Education, explained the rationale for passing the bill: "Freedom of speech and academic freedom are fundamental to higher education. The core mission of universities and colleges is the pursuit of knowledge, and the principles of free speech and academic freedom are fundamental to this purpose. They provide a necessary context for advancing new ideas, encouraging productive debate and challenging conventional wisdom." (Office for Students, 2023).

Despite the increasing prevalence of this phenomenon, the active discussion by policymakers, and the relevance of free speech for the rate and direction of science, there has been little empirical work on how these incidents affect scientists and their research. This paper seeks to investigate whether the scientific output of scholars involved in such incidents is affected.[2]

Ex ante, it is unclear whether researchers' professional outcomes would be affected at all, and if they are affected — the outcome could plausibly be positive or negative. Affected scholars, for instance, may benefit from additional public attention. Media coverage, a condition for inclusion in our dataset on incidents, has been positively associated with scholars' professional success (Philips et al., 1991; Azoulay et al., 2019). Likewise, the premature death of superstar scientists leads to an increase in citations to their papers because it mobilizes a 'salesforce' of academics. Finally, negative reviews can increase sales of products (Berger et al., 2010), which may have parallels to academics receiving criticism, which is still linked to additional media attention.

On the other hand, researchers could face professional penalties and ostracism as the result of their involvement in such an incident (see Azoulay et al., 2015; Widmann et al., 2023). Such an action would be in stark contrast with some of the key principles in science known as 'Mertonian norms' (Merton, 1973), which we draw upon as a theoretical foundation for this paper. Of particular importance to this paper, the norm of universalism suggests that scientific notions must be evaluated independently from the personal opinions of the scholars proposing them. If such norm is not respected, it might lead to bias and unfairness

[2] Notice that the presence of such incidents is rooted in a conflict of opinions, typically emerging when a scholar articulates her stance on matters of public interest that contradicts widely held beliefs. "Matters of public interest" refer to issues, topics, or subjects that are deemed relevant to the general public. These are often areas of concern that impact society at large, and public interest is typically high due to the potential consequences associated with these matters. They typically include debating on politics, climate, foreign affairs, gender, immigration, among others. By definition, therefore, matters of public interest attract debate and conflict of opinions, and are hence the reason why such incidents arise. A typical defense for maintaining freedom of speech especially in academia is that it allows for the inclusion of diverse perspectives, opinions, and ideas in public discourse which enriches debates and ensures that a wide range of viewpoints is considered, contributing to a more comprehensive understanding of complex issues.

in the evaluation of scientific work as research could be judged based on the personal opinions of the researchers rather than the merit of their contributions, which then undermines the objectivity of scientific inquiry and by extension its credibility. If a penalization in terms of a reduction of citations is observed following an incident, then this would provide evidence of the subjectivity and bias in scientific inquiry shading doubts over the credibility, objectivity, and inclusivity of the scientific research process.

To the best of our knowledge, this is the first paper evaluating whether researchers and their research are affected by their speech. Clark et al., (2023, p.1) suggest this may be due to a paucity of data on the topic as "[it] is difficult to detect and measure, [hence] it is rarely empirically studied". We overcome this challenge by leveraging a new and rich dataset, 'Scholars Under Fire' (German and Stevens, 2022), collected by the Foundation for Individual Rights and Expression (FIRE). FIRE is a non-partisan, non-profit organization that defends and promotes free expression in the United States, particularly on university campuses. FIRE assembled a large research team to gather information, through news reports, on incidents involving scholars in the U.S. who have received calls for sanction because of their speech.

Using a difference-in-difference (DiD) approach, we compare the citations of the work published by the affected scholars *before the incidents* to a control group of papers published in the same journal and in the same year as the affected author's work. Using this design, we find that the work of affected scholars receives approximately 4 percent fewer citations as a result of the targeting incidents. In addition, we find that, compared to a control group of scholars with similar pre-incident characteristics (matched on similar publications, citations, field of study, and location), affected researcher's publications decrease by 20 percent after an incident and receive 14.5 percent fewer citations, demonstrating a direct loss to research produced.

Furthermore, we find that there are several factors that mitigate the publication and citation penalty. First, we find that institutions who declare support for the scholar suffer a smaller decline in their productivity. Second, we find that tenured scholars suffer a smaller penalty than untenured scholars. Third, we find that institutions that adopt the 'Chicago principles', a set of guidelines demonstrated to a commitment to the freedom of speech on campus, mitigates the decline in the penalties. This suggests that there are policy levers that universities and governments can draw upon to protect scholars.

The drop in the number of citations comes from researchers wanting to distance themselves from the focal scholar. To validate this mechanism we, first, compare the citation patterns of scholars in the same institution as the affected scholars vis-à-vis scholars in other institutions and, second, compare the citations

coming from scholars who have coauthored at least once before the incident with the affected scholars vis-à-vis scholars that are unconnected, at least in the coauthorship network, to the affected scholars. Both analyses rely on the assumption that, compared to more 'distant' scholars, scholars that are closer to the affected scholar not only have a higher probability of being aware of the affected scholars' work, a fundamental prerequisite for citing a paper, but are also more likely to be informed about any incidents involving the affected scholars. This decrease in citations primarily originates from 'close' scholars rather than 'distant' scholars suggesting scholars may want to distance themselves from the focal scholar.

In addition, we leverage GPT4 to explore the rich qualitative data by classifying the data in different dimensions to explore heterogeneity in our results and further elucidate the mechanism. First, using a legal definition of hate speech, we classify each statement as hate speech (or not) as a way of testing whether only extreme statements are penalized. We find that the effect is not solely driven by these extreme statements. Second, we classify each statement as inside or outside the affected scholar's domain of scholarly expertise, i.e. academic freedom and extramural speech, respectively. We find that the effect holds in both instances. Third, we observe that the citation penalty increases with the degree of extremeness of the incident. We interpret this as evidence that the scientific community distances itself from the affected scholars.[3]

We contribute to the literature in several ways.

First, the political science and legal-rights literature have long histories of studying free speech in the academy (see, for example, Norris, 2023a; Norris, 2023b; Wood, 2022; Alexander, 2006; Wight, 2021; Whittington, 2018) and more recently this topic has received lots of attention in popular books (Lukianoff and Haidt, 2018; Lukianoff and Schlott 2023). Researchers studying science and innovation are interested in how speech and controversy may affect or distort the scientific enterprise. To date, existing work has largely relied on surveys. We make use of modern applied economists' toolkits and a novel dataset to provide a rigorous empirical analysis of this area of research looking at tangible and important career outcomes for researchers: publications and citations. To the best of our knowledge, we are the first empirical paper studying the role of speech in science.

Second, we contribute to the innovation literature shedding light on how scholars decide to cite the work of colleagues. Understanding these motivations is central to the work of science, as citations play a pivotal role in connecting knowledge and rewarding knowledge creators. Several papers have contributed to this

[3] These results are yet to be included in this draft.

literature. For instance, Bao and Teplitskiy (2023) shed light on 'rhetorical citations', where researchers cite works that have not significantly influenced their own research. Similarly, Teplitskiy et al. (2022) examine how the status of authors influences the likelihood of their work being cited, suggesting an underlying bias towards more established researchers. Rubin and Rubin (2021) demonstrate the strategic nature of citations, showing that academics likely choose to cite as strategic signals rather than to document intellectual origins and influences. Koffi (2023) uses machine learning showing there is a gender bias in citations in some fields. Our paper's results are consistent with scholars sanctioning other scholars for opinions they have expressed in the public domain, providing a new motive as to why scholars decide to cite others. To the best of our knowledge, we are the first paper to document this citation strategy empirically.

Third, we provide empirical evidence testing the notion of 'Mertonian norms' (Merton, 1973). Prior work on retractions (Azoulay et al., 2015; Azoulay et al., 2017) and sexual harassment (Widmann et al., 2022) find results consistent with the hypothesis of Mertonian norms not holding empirically. However, retractions and sexual harassment are both clear examples of professional misconduct and illegal actions, respectively. Our empirical setting allows us to test the validity of these norms in a context far from the area of professional misconduct. This paper poses the question, could it really be the case that scholars are sanctioned for their public opinions? Our setting allows us to test this theoretical framework under a much more ambiguous environment, where an effect on citations is ex-ante less certain.

Finally, we contribute to the recent literature on the role of politics and social forces shaping the transmission of information. For example, Braghieri (2024) find that college students are more likely to censor their opinions on sensitive political topics in public. Morales and Samahita (2023) conduct a series of lab experiments to study how public opinion may be shaped by social norms. Huang and Ho (2023) study the effect of increasing the salience of silence in public discourse and find that it can exacerbate self-censorship. Finally, Djourelova (2023) finds that slanted language in the media can influence public opinion. Our paper shows that the provision of credit and the production of knowledge in science may be affected by a scholar's engagement in public discourse, which has impacts on the transmission of information in research.

2. Theoretical Background

2.1 Mertonian Norms

In 1942, sociologist Robert Merton published his article "Science and Society in a Democratic Order", where he described science as an "ethos", held together by four ethical "norms": communalism, universalism, disinterest, and organized skepticism. Of particular relevance to this paper is the norm of universalism, which emphasizes that the "acceptance or rejection of scientific claims […] is not to depend on the personal or social attributes of their protagonist […]" (Merton, 1973, p. 270). Throughout the years, these norms have become "the rules of the game of doing science" (Ziman, 1999, p. 721) and Anderson et al. (2010) suggest that modern scientists seem to adhere to Merton's norms and, despite controversies, these norms remain the "communal property of science".

According to the universalism norm, the merit of a scientific assertion is dependent on the strength of the empirical support and logical reasoning rather than the identity of the researcher making the claim. Therefore, scientific notions must be evaluated independently from the scholars proposing them. By extension, scientific notions should be independent of the scholars' opinions and personal views. This norm fosters a culture of open and unbiased inquiry, reinforcing the notion that scientific progress thrives when ideas are assessed on their intrinsic merit rather than the private characteristics of those who propose them.

If such a norm is not upheld, it might lead to bias and unfairness in the evaluation of scientific work as research could be judged based on the researcher's personal opinions rather than the merit of its contributions. This may then undermine the objectivity of scientific inquiry and by extension its credibility. If a penalization in terms of a reduction of citations is observed following an incident, then this would provide evidence of the subjectivity and bias in scientific inquiry casting doubts over the credibility, objectivity, and inclusivity of the scientific research process.

2.2 Competing Empirical Predictions

Ex ante, it is not clear whether we would observe any effect on scholar's research outcomes. Even if there was an effect, it is plausible that the affected scholars in our dataset could experience positive or negative consequences.

Inclusion in our dataset is predicated upon some form of media coverage: local or national. The positive effects of potentially even negative media coverage is often expressed colloquially in the phrase, 'all publicity is good publicity'. In the context of researchers, this could translate to more citations and publications through increased awareness and media coverage of the scholars and their work.
The relationship between media coverage and a scientific premium has been documented. Philips et al., (1991) analyze a twelve-week strike at the *Times* in 1978, where the editions were still produced but not distributed. They find that academic work covered in the media (and distributed) received 72.8% more citations than the control group (Philips et al., 1991). It could also be the case that negative media coverage increases a scholar's visibility and so their research outcomes. For example, negative reviews can increase the sales of products in some markets (Berger et al., 2010). Beyond media coverage, researchers form clusters of communities and groups in the form of various formal and informal networks. For example, the premature death of superstar scientists leads to more citations to their papers than similar control papers because it mobilizes a 'salesforce' that promotes the deceased's work (Azoulay et al., 2019). Likewise, affected scholars could receive additional support from their networks leading to more support and positive career and research outcomes.

On the other hand, it is plausible that researchers experience a negative impact. Prior research has shown that scientific misconduct such as scientific retractions (Azoulay et al, 2015; Azoulay et al., 2017) and scientists committing sexual harassment (Widmann et al., 2023) lead to penalties for those involved. These are clear and prominent examples of professional negligence and misconduct. In these settings, penalization is arguably expected. Our setting provides a much more ambiguous setting to test the professional effects.

2.3 Freedom of Speech

In an academic environment, freedom of speech can be divided into 'extra-mural freedom of speech' and 'academic freedom of speech'. Both play pivotal roles within universities and the scientific community, contributing to the richness of intellectual discourse and the pursuit of knowledge.

'Academic freedom of speech' safeguards the autonomy of scholars within their academic field, ensuring they can explore, question, and disseminate ideas without fear of censorship or reprisal. 'Extra-mural freedom of speech' acknowledges the right of individuals, including academics, to express their views beyond the confines of their institutional roles or field of expertise. Together, these freedoms create an environment where diverse perspectives can flourish, fostering innovation, critical thinking, and the open exchange of ideas. In the realm of science, these freedoms are particularly relevant, as they allow

researchers not only to explore unconventional theories but also to communicate their findings transparently to the public, contributing to the dissemination of knowledge and the advancement of society as a whole.

While it is straightforward how academic freedom of speech can have an effect on science, the case of extramural freedom is less direct. Extra-mural freedom of speech acknowledges the right of individuals, including academics, to express their views beyond the confines of their institutional roles or field of expertise. This freedom is crucial for scholars to actively engage in public discourse fostering a vibrant and intellectually diverse academic environment. Allowing academics the freedom to express opinions on a wide range of issues encourages a cross-pollination of ideas and perspectives, which is particularly useful as novel ideas in science and innovation are often linked to combining ideas from disparate fields (Fontana et al., 2020), and interdisciplinary and multidisciplinary work is increasingly common (Porter and Rafols, 2009). This not only enriches public debates but also demonstrates that intellectual inquiry transcends disciplinary boundaries. Extramural freedom of speech empowers academics to share their insights on societal, political, and cultural matters, contributing to a more informed and engaged citizenry. By defending this freedom, we safeguard the principles of academic autonomy, free speech, and the broader societal impact of intellectual pursuits, recognizing that the expertise of academics extends beyond their specific disciplines.

This may have tangible consequences for science and innovation. For example, Audretsch et al. (2023, p.1), analysis shows, "academic freedom has a causal impact on innovation. Based on the estimates, the global decline in academic freedom that occurred in the last decade has resulted in a global loss quantifiable in the range of 4.0 to 6.7% fewer patents filed and 5.9 to 23.5% fewer patent citations". Thus, the decline of free speech, or even the fear of social sanctions in academia, may have downstream consequences which suppress the rate and direction of science and innovation.

3. Method and Data

3.1. Method

To analyze the effect of the incident on the citation trajectories of the affected scholars' previous body of work, we employ a difference-in-difference design. Our treatment group consists of works published by affected scholars before the incident. The control group consists of up to 10 randomly selected works from the pool of papers published in the same journal-year as the treated work. This approach controls for the

citation trajectories of the works and assumes that, absent a treatment, works published in the same journal and year will follow, on average, similar citation trajectories and be of comparable quality (see Azoulay et al. 2015, or Furman and Stern, 2011, for example, with a similar journal-volume-issue approach).

We estimate an equation of the form:

$$\text{Citations}_{it} = f(\beta_1 \text{Treat}_i * \text{Post}_t + \delta\Gamma_i + \varphi_t + \eta_k) + \varepsilon_{it} \qquad (1)$$

where $Citations_{it}$ represents the number of citations a work *i* has received in year *t*. As the dependent variable tends to follow a count distribution, we estimate the model as a Poisson model. The variable $Treat_i$ is a binary variable that indicates whether a work is authored by a scholar involved in an incident (1) or whether a work is published in the same journal and in the same year as a treated work but is not authored by a scholar involved in an incident (0). Note that the affiliation with the treatment or control group ($Treat_i$) is time-invariant and, hence, included in the work's fixed effect (Γ_i). Γ_i controls for inherent differences between works caused by unobservable factors in the form of work fixed effects. The variable $Post_t$ is a binary variable that takes the value one from the year after the scholar authoring the work has been targeted.[4] φ_t captures common time trends through a set of year dummies. η_k captures journal-year specific characteristics. More precisely, we interact the year of publication and the journal fixed effects to control for the time varying journal specific characteristics. In addition, to account for the correlation or dependence among observations belonging to (1) the same work, (2) the same affected scholar, and (3) to the same sub-group consisting of one treated work and the respective control works (hereafter, treatment group), we cluster the standard errors at the work, affected scholar, and treatment group.

The main result of the model is provided by the coefficient β_1, which captures the average difference in the change of citations between treatment and control works after the shock. If works in the treatment group were to receive fewer citations after their author has been involved in an incident, while works in the control group do not, β_1 shows a negative and significant effect.

3.2 Data

We primarily make use of a previously unexploited, rich secondary dataset, Scholars Under Fire (Frey

[4] Similarly it takes value 1 also for the control works after the scholar who authored the corresponding treated work is involved in the incident.

and Stevens, 2023), which is collected by FIRE, an apolitical, non-partisan, and non-profit organization financed through individual donations and foundation grants that, among other objectives, advocates for free speech on college campuses (https://www.thefire.org/about-us). FIRE has become the nation's leading defender of fundamental rights on college campuses through its unique mix of programming, including student and faculty outreach, public education campaigns, individual case advocacy, and policy reform efforts. Its mission is to defend and sustain the individual rights of all Americans to free speech and free thought by educating Americans about the importance of these rights, promoting a culture of respect for these rights, and providing the means to preserve them.[5]

As one of its main initiatives, FIRE gathers incidents involving scholars in the US who have voiced their constitutionally protected opinions about matters of public interest in a public domain and subsequently have suffered calls for sanction at any public and/or private American higher education institution from 2000 to the present.[6] FIRE gathers data mainly from news reports from campus, local, and national news outlets. It then compares its search results with other existing sources tracking similar incidents to identify additional cases that did not emerge during their own search. An incident is defined by FIRE as a "campus controversy involving efforts to investigate, penalize or otherwise professionally sanction a scholar for engaging in constitutionally protected forms of speech".

FIRE classifies incidents into several categories related to the topics on which the scholar has expressed her opinion (number of cases in parenthesis)[7]: abortion (14), climate (18), court trial (25), COVID-19 (50), elections (48), foreign affairs (34), freedom of speech (48), gender (164), immigration (28), institutional conflict (219), Israel/Palestine (65), law enforcement protests (83), mental or physical health (41), partisanship (194), race or racial issues (351)[8], religion (103), Russia/Ukraine (2), gun rights (17),

[5] As FIRE states, "this cornerstone [freedom of speech] of our free society is under serious threat. Far too many of us fear sharing our views or challenging those that seem to dominate. Nearly 6-in-10 Americans believe our nation's democracy is threatened because people are afraid to voice their opinions".

[6] Notice that the annual number of incidents has increased dramatically over time, from 4 in 2000 to 145 in 2022, which is in line with the growing concerns on freedom of speech in academia. Figure 1 shows the yearly number of incidents as provided by FIRE and as present in our final dataset.

[7] Notice that a scholar might express one or more opinions at once or one opinion involving multiple topics.

[8] We acknowledge that the number of race-related cases is particularly high. Nevertheless, for 231 of these cases, the incident involved also other topics, hence such cases are not only directly related to racial issues. As reported by FIRE, race-related expressions include expressions regarding racial inequality, historical racism, race-specific DEI efforts, the Black Lives Matter movement, and the use of racial slurs. Also notice that the use of racial slurs is classified as 'hate speech' and, in a robustness check we show that these cases are not driving our results.

economic inequality (50), sexuality (74), and terrorism (50).[9] To give a flavor of our dataset, here are five examples:

1. In 2013, Christopher Rollston, Professor of Theology at Milligan University, wrote an opinion article for The Huffington Post's religion section about the marginal status of women in the Bible. He resigned.
2. In 2017, Bruce Gilley, Professor Political Science at Portland State University, was subject to a petition demanding the retraction of his peer-reviewed paper on the supposed benefits of Western colonialism. The petition gathered 10,956 signatures.
3. In 2021, Edward Livingston, Professor of Medicine at University of California, Los Angeles, was forced to resign as deputy editor of The Journal of American Medical Association due to outcry over a podcast where he questioned structural racism.
4. In 2005, Hans-Herman Hoppe, Professor of Economics at University of Nevada, Las Vegas, tried to illustrate the concept of time preferences by citing homosexuals as a group of people who tend to spend more readily because they are not thinking about starting a family and so do not feel they must save money to raise children or buy homes. The initial backlash resolved in no sanction towards Prof. Hoppe.
5. In 2012, David Shorter, Professor of Anthropology at University of California, Los Angeles, came under scrutiny for posting a link to a site advocating for the cultural and academic boycott of Israel.

For each incident, FIRE provides a comprehensive set of information, including "what was being expressed (*topic*); who was being addressed or talked about (*subject*); the reason for the expression (*intent*), and where the scholar's speech took place (*context*). Additionally, [FIRE] identified those who initiated the […] incident (*source*); how they want the scholar sanctioned (*demands*). [FIRE] also included how the scholar reacted to the […] incident (*response*); how the institution or administration reacted (*administrative response*); and the outcome of the […] incident (*outcome*)", among many others.[10,11]

[9] If in any category there are statements involving 'hate speech', we drop them in a robustness check. As mentioned above, our results hold the exclusion of such cases.

[10] Refer to the following link for a complete list of all available information on the scholar and on the incident: https://www.thefire.org/research-learn/scholars-under-fire-variable-codebook

[11] FIRE also provides a detailed example of how one of the incidents is coded into the different categories: "Sandra Sellers, a former adjunct professor of law at Georgetown University, was unwittingly recorded telling her colleague, "I end up having this angst every semester that a lot of my lower ones [students] are Blacks," the topic was categorized as "race"; subjects were "Black people," "graduate students," and; the intent was categorized as both "personal view/opinion" and "unintentional/accidental"; the context was "direct interaction"; the source was both "undergraduate students" and "graduate students"; the demands included a "list," "termination," and "policy change"; the scholar's response was to "express regret" and "leave"; the institution's response was to "apologize" for and "terminate" the scholar; and the result was that the scholar was "terminated."

In total, our preliminary dataset comprises approximately 844 scholars.[12] After careful inspection of each incident and the scholars involved, our sample contains 392 scholars. This reduction is due to several reasons. Firstly, we manually check (1) whether the incident involves a scholar's expression of her own opinions,[13] (2) whether the incident gathers some level of media attention at the campus, local, and/or national level,[14] and (3) whether the incident involves illegal actions.[15] In this step, 349 scholars do not get through one or more of these filters. For the remaining 495 scholars, we gather bibliometric data using OpenAlex (Priem, Piwowar, and Orr, 2022). OpenAlex is an open-source dataset enabling us to obtain publication and citation information for the vast majority of the scholars in our list. To ensure accuracy, we manually search the OpenAlex API for each scholar's data by (1) searching for their name and institution, as provided by FIRE, (2) identifying one of their publications and saving the publication ID as provided by OpenAlex, (3) identifying, through the publication, the OpenAlex ID related to the scholar, and (4) using the OpenAlex scholar ID in the API to extract all their associated publications.

Secondly, out of the initial 495 scholars, only 392 resulted in at least 1 publication; in fact, it should be noted that our dataset comprises also people in the academic environment with titles such as specialists, lecturers, and adjuncts, among others, which might not be actively involved in research. Note, also, that some scholars in OpenAlex have multiple IDs. In these cases, we manually check whether the OpenAlex IDs are correct and if so, we include all the relevant publications. At this point, 41,303 works have been found. We proceed to clean the dataset by excluding all non-English works (39,574 works remain in our dataset). To ensure comparability between pre- and post-incident, we include in our sample only scholars that have at least one publication before and after the incident so that their performances before and after the 'shock' can be meaningfully compared.[16] In other words, we exclude scholars who might have left academia for reasons possibly related to the incident; in fact, if we were to include scholars that, for example, abandoned the academic environment in the aftermath of the incident in which they were involved, we would surely include a negative bias in our estimates. This conservative approach decreases our sample to 301 affected scholars with 36,569 works. While surely the drop in the number of scholars is

[12] Notice that some scholars are included in the dataset multiple times as they were involved in multiple incidents. In these cases, the first incident in chronological order that gets through our filters is the one we consider.
[13] For example, Bright Sheng, Professor of music at the University of Michigan, was forced out of class for showing a 1965 movie of Othello in which actor Laurence Olivier darkened his skin to play Othello. Students later called for sanction. We deem examples like this outside the scope of the paper as they are not a clear expression of the scholars opinion.
[14] In order to test whether an affected scholar's work is impacted by her opinion once it is shared in a public domain, we need to assume that the event has reached the scientific community. If that is not the case, then surely no effect can be expected.
[15] To better fulfill our purposes of analyzing whether the affected scholar's work is impacted by the affected scholar's opinion when it is shared in a public domain, we try to avoid including cases of misconduct which as studied by Widmann et al. (2022), for the case of scholars involved in sexual harassment, for example, have already been proved to lead to a penalization.
[16] In a robustness check, we only focus on affected scholars with at least 5 publications before and after the incidents.

large, it is worth mentioning that the percentage drop in the number of works is significantly lower pointing to the fact that in this step we have probably dealt with those scholars with a lower number of publications hence less actively involved in research.

Each of the affected scholars enters our sample with her first publication and exits with her last. Still, to pinpoint the effect of the incident we must focus on the affected scholars' body of work published before the incident. Therefore, we must drop those works that are published after the scholar is treated. 8,544 works are now dropped, i.e., on average approximately 28 works per affected scholar.

As mentioned above, following the approach proposed by Furman and Stern (2011), we compare the citations of the works published by the affected scholars before the incidents to a control group of works in the same journal, volume, and issue. This approach controls for the citation trajectories of the works and assumes that, absent a treatment, works published in the same journal and issue will follow, on average, similar citation trajectories (see Azoulay et al., 2015, among others). However, replicating this exact approach reveals problematic in our dataset as 11,863 and 13,267 works out of the total 28,025 works left in our dataset contain missing values in their volume and issue, respectively. For this reason, to extract the control group works we take advantage of the OpenAlex API and select up to 50 works that have been published in the same journal and in the same year as the treated works.[17] We believe this modified approach can still yield a high-quality control group and the assumption of similar citation trajectories detailed above still holds. Some of the treated works (1,627) do not yield any comparable works and so are removed. Notice also that information from the journal is missing for 4,506 works which are also not included in our final sample. Hence, 21,892 observations on treated works are left in the dataset from 283 affected scholars. After some careful inspection of both treated and control works, we find that some of them have a publication date that suggest their dates are incorrect. We drop all the works in both groups with a publication year earlier than 1970 (15,330 observations are dropped), works that are not published in English language (27,391 observations are dropped), and works that have been retracted (49 observations are dropped), resulting in 21,381 treated works from 283 affected scholars and 798,266 control works.

We include in our sample scholars involved in incidents until 2021 and works published up until 2020. Among the pool of possible control works, we randomly select up to 10 control works per treated work. After carefully cleaning the data and downloading information on the number of yearly citations received

[17] In a robustness check (still to be included in this draft), we only on the papers for which volume and issue information are available.

by each work, we are left with 4,582,071 observations in which 4,163,956 belong to the control group and 418,115 belong to the treated group. All together our preliminary dataset comprises 201,533 and 20,822 control and treated works respectively (9.67 control works per treated work) observed from their publication year until 2023. The treated works belong to 283 affected scholars.

Before going into our analysis, we account for outliers in the dependent variable, number of yearly citations, by dropping the top 1% of the observations. In addition, after careful consideration, we exclude works from scholars whose cases are followed by 'Turning point'.[18] To conclude we focus our analysis on a time window that includes four periods before and after the treatment year. Our final dataset therefore comprises 1,619,462 observations on 197,805 works divided between 1,472,152 and 147,310 observations in the control and treated group respectively and 247 affected scholars. As described later in the methodology section, using work level fixed effects reduces the number of observations in our sample to 759,244 as 104,465 works have zero citations and are hence automatically dropped. This leaves us with a final dataset of 246 affected scholars with 9,568 treated works and 84,772 control works. Figure 1 below shows the number of scholars that are affected in each year and compares the ones in the raw data to the ones used in our analysis.

INSERT FIGURE 1 ABOUT HERE

3.3 Descriptive statistics

Table 1 provides some descriptive statistics of the number of citations divided by treated and control group, pre and post. On average both groups receive less citations in the post period compared to the pre one, which is line with the fact that over time works become less relevant as new or updated insights are generated. Nevertheless, the drop in citations seems to be larger, in magnitude, for the treated group. Similar conclusions are found in Figure 2 where we compare the average citations per years received by treated and control works over time. On the top left, we report all the works in our regression, which as mentioned before, by construction excludes the works that have never received any citations; on the top right, we show that the above-mentioned movement in both groups is still present when we also include the works that have never received any citation. To conclude in the bottom section, we show that our

[18] Turning point often targets scholars in a certain year for events that have taken place in the past. Considering these cases might cause issues for our analysis, 332,054 observations are dropped on 19,395 works divided into 1,786 and 17,609 treated and control works respectively with treatment years taking place mostly in 2021 and affecting 21 affected scholars in total.

results are not due only to recent trends and when excluding works of scholars treated in 2021, we still find the same relations. Figure 3 provides an overview of the distribution of the number of citations per year for all works including the ones that have never received any citations.

INSERT TABLE 1 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

INSERT FIGURE 3 ABOUT HERE

Table 2 provides some key information on the affected scholars in our treated group. On average, scholars involved in an incident tend to be relatively older people within the academic community (see for example the mean academic age at treatment, i.e., the difference between the year of the incident and the year of their first publication, is 33.5). This is in line with our expectations as usually scholars who have received tenure are more likely to expose themselves more publicly. It is important to notice that on average the affected scholars in our sample do continue their research activities also after their involvement in an incident (almost all the scholars involved in the incidents have more than five publications after the shock and are hence considered as active researchers). This is helpful for our analysis because it could be argued that, while control works might receive more citations after the shock because their authors keep publishing hence also their prior work gains visibility and perhaps the affected scholars might publish less after being involved in such an incident meaning that their prior work would suffer from a decrease of visibility, this does not seem the case as on average the affected scholars seem to even improve their productivity in the post shock period. Furthermore, the loss in citations to scholars' prior work is seen after two years. This further suggests that the citation penalty is not because of their visibility decreasing, which may become more noticeable over a longer period of time.

INSERT TABLE 2 ABOUT HERE

To conclude Table 3 below shows parallel movement of the dependent variable in the pre-treatment period between treated and control group. The specification in Table 3 extends equation (1) in that we interact the year dummies (φ_t) with the treatment indicator ($Treat_i$). From column 1 to 3 we add fixed effects starting from work level fixed effect in column 1, adding the year fixed effect in column 2, and then also the journal *x* publication year fixed effect to account for the changing journal quality. In addition in column 4 we provide a subsample check in which we exclude the scholars treated in 2021. Our results are robust in all these specifications. Notice that we cluster the standard errors at the work, affected scholar, and treatment group level (the treatment group includes the treated work and the respective control works).

INSERT TABLE 3 ABOUT HERE

---

4. Results

4.1. Citation penalty on the researcher's prior work

Our initial analysis examines the effect of an incident on the citations received by the affected scholar's earlier work. Table 4 reports the results of Equation 1. As mentioned before, given that $Treat_i$ is time invariant, it is collinear with the work fixed effect hence it is excluded from our regression. The coefficient for the interaction of $Treat_i * Post_t$ is negative and significant ($p < 0.01$) in column 1, 2, and 3 providing initial evidence of a decrease in the citation trajectories for the treated group; more specifically, treated works receive about 4.2% less citations after their authors are involved in an incident. Notice here we include fixed effects in the same fashion as in Table 3. Column 3 is our preferred specification. Notice that the moderate magnitude of our results is consistent with earlier work focusing on retractions (Azoulay et al., 2015) and sexual harassment (Widmann et al., 2022). [19]

INSERT TABLE 4 ABOUT HERE

---

[19] Results are robust to the use of the DiD estimator of Callaway and Sant'anna (2021) and to OLS estimations with and without *log + 1* transformation of the dependent variable.

4.2. Heterogeneity

We carefully investigated each of the incidents and categorized them. Table 5 below reports the percentages of scholars in our sample that fall into each category. As described above, we exclude the scholars that have been targeted by Turning Point as this association often targets scholars for incidents that took place in the past. Including these scholars might therefore generate noise in our results in terms of treatment year. In our original sample, 9% of the scholars are targeted by turning point. 10% of the scholars are involved in extreme incidents. Our categorization of extremeness includes scholars who expressed racism; strong, vulgar, or inappropriate words or actions in the classroom. We want to isolate these incidents as they are not necessarily related to free speech but are more closely related to 'hate speech' which is not the focus of this paper.[20] In addition, we want to also isolate those scholars that have been fired, terminated, or have resigned (14%) as they might bias our results and show a drop in their citations not necessarily related to a penalization coming from their peers. Table 6, column 1, below show that our results are robust when we exclude both extreme and fired, terminated, or resigned scholars. In addition, we also group scholars that are often controversial. As mentioned above, in case a scholar is involved in multiple incidents or if the scholar had a history of controversy but didn't occur in the dataset more than once, we consider the year of the first incident as the treatment year; nevertheless, these scholars, given their controversial nature, might have been already suffering a penalization as their peers might be already aware of their private opinions about controversial arguments. Table 6, column 2, shows that our results are robust when we exclude also these scholars from our analysis (notice here we keep excluding also the fired, terminated, or resigned, and extreme scholars to reach a cleaner dataset).

We also distinguish between scholars involved in extramural freedom of speech and academic freedom of speech. For example, Prof. Jodi O'Brien is involved in an incident because some of her academic writings were at odds with the church. This is an example of a scholar expressing her opinion within her own field of study, i.e. academic freedom of speech. On the other hand, Prof. Bert Chapman is involved in an incident because he posted an article on a blog named "Conservative Librarian" entitled "An Economic Case Against Homosexuality" which argued that "the cost for AIDS research and treatment should factor into the national debate over the acceptance of gays and lesbians," and made other statements and arguments reflecting his opinions and his religious views about homosexuality. Given that Professor Chapman's field is library science, he is clearly expressing opinions on matters of public interests outside of his field, i.e. extramural freedom of speech. As shown in columns 3 and 4 of Table 6 where we include

[20] In a robustness test, we ask ChatGPT to identify the incidents involving hate speech and flag them as extreme. We also hired some Ras to conduct a similar classification.

only scholars expressing opinion outside their fields of expertise and scholars expressing their opinions within their field, respectively, a penalization takes place for both groups, but is stronger for scholars expressing opinion outside their fields.[21]

Analyzing cases of 'academic freedom of speech' and 'extramural freedom of speech' separately provides further insights into the mechanism as it allows us to mitigate against the concern that our results might be driven by 'Bayesian discounting'. Bayesian discounting, in the context of this study, is when scholars lower their opinion of an affected scholars' work quality based on the content of her speech. For instance, one example in our dataset is Prof. Jason Hill, who is an expert on political philosophy and American foreign policy. He expressed pro-Israeli views in an op-ed and received calls for sanction with a petition signed by 3,581 people. In this case, other scholars may adjust their beliefs on Professor Hill's work based on his speech. Bayesian discounting should not occur when scholars express their opinions outside of their domain of research expertise, since the statement is unrelated to their research. Our results show a citation penalty also for cases of extramural speech where the scholars' statements are orthogonal to their research. As our results hold across both types of speech, we can rule out Bayesian discounting as the sole factor driving the citation penalty.

In columns 5, 6, and 7, we use GPT4 to classify incidents involving hate speech using the following definition "public speech that expresses hate or encourages violence towards a person or group based on something such as race, religion, sex, or sexual orientation", incidents involving extramural freedom of speech, and incidents involving academic freedom of speech, respectively. In line with the results of our and our RAs' classifications (columns 1, 3, and 4, respectively), also when GPT4 classifies these incidents, our results hold.

To conclude, in columns 8 and 9 we distinguish between incidents involving scholars affiliated to institutions that have signed and not signed the Chicago principles, respectively. The Chicago Principles articulate the importance of free expression as an essential feature of the university so we expect scholars affiliated to non-signing institutions to be more negatively affected. The results are in line with our expectations and affected scholars affiliated to signing institutions do not seem to be affected at all.

[21] In addition, we find that male scholars are strongly penalized while female scholars are not. Still, we refrain from making any claims as only 13% of the dataset includes female scholars.

INSERT TABLE 5 ABOUT HERE

INSERT TABLE 6 ABOUT HERE

In Table 7, we continue our heterogeneity analysis. In column 1, we include only scholars who are more actively involved in research activities both before and after their involvement in an incident. We identify such cases by subsetting only to the cases in which the scholars have at least 5 publications before and 5 after the incident (166 scholars are included in this sub-sample). Our results hold. In columns 2 and 3, we divide the sample between scholars who have already received the tenure as of the incident year and scholars who have not. While both groups are negatively affected, the magnitude of the coefficient for non-tenured affected scholars is larger. To conclude, we split the sample into scholars who received and did not receive support following their involvement in an incident from their own institutions in columns 4 and 5, respectively. In line with our expectations, scholar receiving support from their institutions are not affected by the incident while unsupported scholars absorb all the negative effect.

INSERT TABLE 7 ABOUT HERE

4.3. Mechanism

This paper proposes that affected scholars are penalized after the incidents as fellow researchers distance themselves by citing their body of work less. Testing the validity of this mechanism implies identifying a strategy for which we can isolate and analyze the exogenous distancing choice of other researchers.[22] To this end, still focusing on citations patterns, we identify the number of citations to the affected scholars' works coming from (1) scholars that are affiliated to the same institutions as of the year of the treatment (close scholars) and (2) scholars that are affiliated elsewhere as of the year of the treatment (distant scholars). As opposed to distant scholars, close scholars not only are more likely to know about the focus scholars' work, an obvious prerequisite to cite a paper, thanks to the well-established informal interactions

[22] While looking at whether the number of coauthors decreases after the shock, which could also point to a penalization and distancing effect, might seem a valid approach, we argue that it is in fact possible that a drop in the number of coauthors might be endogenous as the choice of collaborating with other researchers depends also on the affected scholars themselves.

that take place within the same institutions, such as internal seminars, research lunches, interfaculty meetings, as well as simple chat in the university's corridors, but also might have an incentive to cite more of their fellow scholars work because of some University's policies. Above all, close scholars are also the most likely to know about the affected scholars' incidents. If the pre-established drop in the number of citations comes mainly from close scholars vis-à-vis distant scholars, then it is more likely that we are indeed observing a distancing effect resulting from the shock.

To carry out this analysis, we, first, identify the main affiliations of the affected scholars at the time of the treatment[23] and second, starting from the same set of works published by the affected scholars used in the main analysis, we classify the yearly citations to their works between close scholars' citations, i.e. citations coming from papers in which at least one of the co-authors is affiliated to the same institution of the affected scholar, and distant scholars' citations, i.e. citations coming from papers published by anyone else. This process allows us to calculate the average number of citations coming from close vis-à-vis distant scholars to the affected scholars' body of work for each period of observations and plot their time trends. As shown in Figure 4 below, the percentage drop in the close scholars' citations is larger than the one found for the distant scholars' citations; in fact, while in the treatment year (t = 0), the average number of close scholars' citations is 0.09, it quickly drops to 0.069 in period 2, 0.054 in period 3, and 0.037 in period 4 resulting in a percentage drop of nearly 24%, 40%, and 59% in period 2, 3, and 4, respectively, compared to the treatment year. The reduction in distant scholars' citations, although obviously greater in magnitude as the group of citing scholars includes all scholars outside of the affected scholar institution[24], consists of 11%, 22%, and 20% in period 2, 3, and 4, respectively, compared to the treatment year. As the percentage drop is larger for the group of citations coming from researchers that are closer to the affected scholars, this evidence points to a distancing effect.

INSERT FIGURE 4 ABOUT HERE

To further validate our proposed mechanism, we run a similar analysis in which we define close and distant scholars' citations in a different fashion, following Widmann, Rose and Chugunova (2022) and leveraging the affected scholars' coauthorship network. Here, we assume that scholars who have co-

[23] If a scholar has more than one affiliation, we consider the one that is (1) located in the US and (2) is most used in her works published between two years before the treatment and the treatment year.

[24] Note that in the Figure 4 the number of distant scholar citations are scaled by a factor of 10 to make them visually comparable to the close scholar citations.

authored at least once before the incidents with the affected scholars are again more likely to (1) know the affected scholar work and (2) know about the incidents in which they are involved. If the number of citations to the affected scholars' body of work coming from direct coauthors (close scholars) drops more (in percentage) than the number of citations from other scholars (distant scholars), this might also be indicative of a distancing effect and provide further support for our mechanism.

As shown in Figure 5 below[25], this is exactly what we observe. Indeed, while the average number of citations from close scholars at the time of the treatment equal 0.33, it quickly drops to 0.26, 0.19, and 0.15 resulting in a percentage drop of about 21%, 43% and 55% in period 2, 3 and 4, respectively compared to the treatment year. On the other hand, the percentage drop in distant scholar citations is about 6%, 17%, and 10% % in period 2, 3 and 4, respectively compared to the treatment year, indicating a significantly lower percentage drop in the citations.

We interpret both these findings as indicative of a distancing effect which validates our proposed mechanism.

INSERT FIGURE 5 ABOUT HERE

4.4 Future Productivity

To complement our work on the citation penalty to researcher's prior work, we also look at the affected scholars' future productivity and impact, i.e. number of publications and future citations, respectively. To do so, we match affected scholars based on their observable characteristics as of the year before they were involved in an incident. We consider the following set of matching criteria: (1) main topics covered in their work[26], (2) location[27], (3) stock of publications and citations[28], (4) first publication year, and (5) number of coauthors.

[25] Note that also here the magnitude of the citations from distant scholars is larger than the ones from close scholars as the latter considers all citations coming from scholars who are not in the coauthor networks of the affected scholars as of the year of the treatment. Hence, also here the number of citations from distant scholars is scaled by a factor of 5.

[26] In OpenAlex this information is given by the variable 'concept'. We extract the top 3 most common concepts among all the works published before the incident.

[27] Country of main affiliation, i.e. U.S.

[28] For the stock of publications we allow for a margin of 10% while for the stock of citations we allow for plus/minus 100 citations.

From the pool of comparable control scholars, we randomly select 10 scholars. Nevertheless, for some of our affected scholars we are unable to get 10 control scholars. Our sample of affected scholars is 260 while the control group is composed of 2,341 scholars hence, we have on average 9 control scholars per affected scholar.

As shown in Table 7, columns 1 and 2, parallel trends hold. Results from columns 3 and 4 instead show the results for the main regression using a model similar to Equation 1 for both the number of yearly citations and the number of publications. Consistently with our results from the paper-level analysis above, the effect of the penalization kicks in 2 years after the incident for the number of citations, but the magnitude of the effects is much larger, i.e., as opposed to the control group, affected scholars experience an overall drop in citations of 14.5%. When looking at the number of publications, we find that the effect realizes one year after the shock and corresponds to a drop in the number of publications of 20%. Notice that also here scholars targeted by Turning Point have been dropped (38 scholars). Our results are nevertheless robust to their inclusion in the analysis and to the exclusion of scholars who have been eventually fired, terminated, or have resigned, as well as scholars involved in extreme incidents and that are often controversial.

INSERT TABLE 8 ABOUT HERE

4.4.1 Mechanism - Publications

We further analyse the mechanism behind the drop in publications and pose that if an affected scholar has the support of her institution, aside from the clearly sub-optimal psychological conditions in which she might find herself, which are surely not conducive to high research output, she will have to spend a significant portion of her time into meetings (such as ad-hoc committees' meetings created following the incident) which will mechanically lead to less time spent on research and therefore, most likely, to a lower scientific output in terms of publications. In addition, it could be the case that if her institution decides to not show support and therefore, most likely 'attack' the scholar, she will receive less support in the form of grants and funding. Following this argument, observing a larger drop in publications for affected scholars who were not supported after the incident by their institutions vis-à-vis supported scholars, might be indicative of the fact that it is indeed the effect of the incident that is leading to a loss of publications. The preliminary evidence shown in Table 9 below where we split our sample into supported and unsupported scholars, in columns 1 and 2, respectively, provide initial support for this mechanism. In fact, aside from

the smaller coefficient for the supported scholars, we also observe a very low significance level in column 1 (p-value > 0.1).[29]

INSERT TABLE 9 ABOUT HERE

---

5. Discussion

5.1 Contributions and Implications

Freedom of speech is important for universities, and more broadly science and innovation. Yet recent evidence shows declining trends in scholars' perceptions of their freedom to express their opinions and an increasing fear of losing their jobs or reputations due to their speech (Honeycutt et al., 2022), or to share differing perspectives or argue against the consensus among their colleagues (F. S. Union Tech. Rep., 2023). While the literature on the costs and benefits of free speech has been largely analyzed by social scientists, legal scholars, and academics in the humanities, there is little-to-no evidence derived from a purely data-driven approach. This paper is the first to empirically analyze whether scholars are penalized for exercising their right to free speech about matters of public interest.

While ex-ante it is not clear whether researchers' professional career outcomes would suffer or benefit from their involvement into such incidents, our results clearly show a penalization effect. In fact, while affected scholars could benefit from their involvement in such incidents thanks to the additional public attention they gain (Phillips et al., 1991) and the plausible support they could receive from other academics (Azoulay et al., 2019), our results show a penalization consisting of a 4% reduction in yearly citations to the affected scholars' prior body of work. This is consistent with previous literature establishing that researchers face professional sanctions for professional misconduct (see for example the work on retractions (Azoulay et al., 2015; Azoulay et al., 2017)) or the work from Widmann et al. (2022) on sexual harassment. Differently from the other papers, we offer an interesting case of social sanction in which scholars are not penalized due to misbehavior in their work but instead because of their personal opinions on topics that may be perceived as controversial.

[29] As these are only preliminary evidence, at this stage, we refrain from making any claims and do not discuss these findings in the discussion section.

Our paper therefore provides an interesting test for the validity of the Mertonian norms (Merton, 1973). According to the universalism norm, the merit of a scientific assertion is dependent on the strength of its empirical support and logical reasoning rather than the identity of the researcher making the claim. Therefore, scientific notions must be evaluated independently from the scholars proposing them. By extension, scientific notions should be independent of the scholars' opinions and personal views. Academia is often upheld as a strongly objective institution, characterized by unbiasedness and fairness in the evaluation of scientific work. Our results clearly show a potentially worrisome subjectivity in the evaluation of science as research seems to be judged based on the researchers' personal opinions rather than the merit of its contributions, which then undermines the objectivity of scientific inquiry and by extension its credibility. This provides evidence of the subjectivity and bias in scientific inquiry casting doubts over the credibility, objectivity, and inclusivity of the scientific research process.

Thirdly, we contribute to the science and innovation literature on the underlying reasons scholars decide to cite other works. Academics strive to improve the quantitative metrics by which they are evaluated (Franzoni et al., 2011). Our paper shows that scholars may be sanctioned for their speech, which has implications for the likelihood of scholars to engage in public discourse and research. This may have a deleterious impact on science and innovation.

Beyond the scope of this paper is its implications on trust in science. Recently, evidence has pointed to a strong decline in trust in science as an institution (Gauchat, 2012; Kennedy and Tyson 2023). During the COVID-19 pandemic, *Nature* endorsed Joe Biden. This endorsement lowered stated trust in the journal among Trump supporters, as well as lowered the demand for COVID-related information provided by *Nature* (Zhang, 2023). Our paper suggests that scientific citations may be shaped by speech, which may add to the public's distrust in science. We hope that future research can study this.

5.2 Policy Implications

Governments, civil society, and academic institutions have often discussed the role of free speech in academia. However, much of the existing debates lack empirical support. Recently, many universities have adopted policies on protecting the free speech of their academics at their institution. We hope that our paper provides university administrators with additional evidence to guide their decision on how the academics' professional careers may be affected.

Furthermore, many governmental organizations are talking about or have adopted legislation to protect free speech. For example, the United Kingdom recently adopted the "University Freedom of Speech Bill",

which is legislation created to ensure universities protect and promote freedom of speech on campus. Likewise, Quebec recently passed a bill to protect academic freedom. We hope that our study guides policy makers on whether to protect academic free speech and how strong those protections should be.

5.3 Limitations

The vast majority of our data comes from scholars in the United States. This is a limitation because we are unsure of how our results may generalize to other countries. Calls to sanction free speech are increasingly a global phenomenon (Scholars at Risk, 2021) and so we hope we (or other researchers) can collect data on scholars outside of the US and extend our analysis to a global level.

Secondly, we only focus on cases that receive some sort of media attention. This means that we may not be detecting cases that the wider academic community is not aware of. However, finding instances of penalization where there is little attention from the academic community inherently makes it difficult to collect data.

Thirdly, as these types of incidents are increasing in number only in more recent years, we are not able to analyze their long-term effects on the affected scholars yet.

Finally, we only capture the most visible incidents. For example, if an academic says something privately, then this is something we cannot capture. Ultimately, we think that this is unavoidable because situations like these are hard to track with data.

6. Conclusion

In conclusion, our study sheds light on a previously unexplored topic: are academics sanctioned for exercising their right to free speech? Our results suggest so. Affected scholars in our dataset do face a citation penalty to their prior work and, additionally, they are less productive after the incidents, publishing fewer papers, and receiving fewer citations. These findings are consistent with the notion that academics do not seem to separate the researcher from their research, in contrast to the Mertonian norm of universalism which might cast doubts on the objective evaluation of scientific work and could risks undermining the credibility, objectivity, and inclusivity of research.

7. References

"Academic freedom survey" (F. S. Union Tech. Rep., 2023)

Alexander, L., 2006. Academic freedom. *U. Colo. L. Rev.*, 77, 883.

Anderson, M. S., Ronning, E. A., Vries, R. D., & Martinson, B. C. (2010). Extending the Mertonian norms: Scientists' subscription to norms of research. *The Journal of higher education*, *81*(3), 366-393.

Audretsch, D., Fisch, C., Franzoni, C., Momtaz, P. P., & Vismara, S. (2023). Academic Freedom and Innovation: A Research Note. *arXiv preprint arXiv:2303.06097*.

Azoulay, P., Fons-Rosen, C. and Zivin, J.S.G., 2019. Does science advance one funeral at a time?. *American Economic Review*, *109*(8), pp.2889-2920.

Azoulay, P., Furman, J.L. and Murray, F., 2015. Retractions. *Review of Economics and Statistics*, *97*(5), pp.1118-1136.

Azoulay, P., Wahlen, J.M. and Zuckerman Sivan, E.W., 2019. Death of the salesman but not the sales force: how interested promotion skews scientific valuation. *American Journal of Sociology*, *125*(3), pp.786-845.

Bao, H. and Teplitskiy, M., 2023. Do" bad" citations have" good" effects?. *arXiv preprint arXiv:2304.06190*.

Berger, J., Sorensen, A.T. and Rasmussen, S.J., 2010. Positive effects of negative publicity: When negative reviews increase sales. *Marketing science*, *29*(5), pp.815-827.

Bhagwat, A., and Weinstein J., 2021. 'Freedom of Expression and Democracy' in Adrienne Stone, and Frederick Schauer (eds), *The Oxford Handbook of Freedom of Speech*, Oxford Handbooks

Callaway, B. and Sant'Anna, P.H., 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), pp.200-230.

Castelvecchi, Davide. 2020. Mathematicians urge colleagues to boycott police work in wake of killings. *Nature News*, June 19

Clark, C.J., Jussim, L., Frey, K., Stevens, S.T., Al-Gharbi, M., Aquino, K., Bailey, J.M., Barbaro, N., Baumeister, R.F., Bleske-Rechek, A. and Buss, D., 2023. Prosocial motives underlie scientific censorship by scientists: A perspective and research agenda. P*roceedings of the National Academy of Sciences*, *120*(48), p.e2301642120.

Emerson, T.I., 1976. Legal foundations of the right to know. *Wash. ULQ*, p.1.

Fontana, M., Iori, M., Montobbio, F. and Sinatra, R., 2020. New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy*, *49*(7), p.104063.

Franzoni, C., Scellato, G. and Stephan, P., 2011. Changing incentives to publish. *Science*, *333*(6043), pp.702-703.

Frey, K. & Stevens, S.T. (2023). Scholars under fire: Attempts to sanction scholars from 2000 to 2022. The Foundation for Individual Rights and Expression.

Furman, J.L. and Stern, S., 2011. Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review, 101*(5), pp.1933-1963.

Gauchat, G., 2012. Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American sociological review, 77*(2), pp.167-187.

German, K. & Stevens, S.T. (2022). Scholars under fire: 2021 year in review. Available online at: https://www.thefire.org/research/publications/miscellaneous-publications/scholars-under fire/scholars-under-fire-2021-year-in-review-full-text/

Haidt, J. and Lukianoff, G., 2018. The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure. Penguin UK.

Hemel, D. J. (2019). Economic Perspectives on Free Speech. *Economic Perspectives on Free Speech (November 25, 2019). In Frederick Schauer and Adrienne Stone (eds), Oxford Handbook of Freedom of Speech (Oxford University Press, Forthcoming), University of Chicago Coase-Sandor Institute for Law & Economics Research Paper*, (898).

Honeycutt, N., Stevens, S.T. and Kaufmann, E., 2023. The Academic Mind in 2022: What Faculty Think About Free Expression and Academic Freedom on Campus.

Kennedy, B., and Tyson A., 2023, "American's Trust in Scientists, Positive Views of Science Continue to Decline," (accessed November 20, 2023), [available at: https://www. pewresearch.org/science/2023/11/14/americans-trust-in-scientistspositive-views-of-science-continue-to-decline/].

Koffi, M., 2023. Innovative ideas and gender inequality (No. 35). Working Paper Series.

Lukianoff, G. and Schlott, R., 2023. The Canceling of the American Mind: Cancel Culture Undermines Trust and Threatens Us All—But There Is a Solution. Simon and Schuster.

Macfarlane, B., 2023. The DECAY of Merton's scientific norms and the new academic ethos. *Oxford Review of Education,* 1-16.

Merton, Robert K., 1973. The Sociology of Science: Theoretical and Empirical Investigations. Chicago: *University of Chicago Press.*

Norris, P., 2023a. Cancel culture: Heterodox self-censorship or the curious case of the dog-which-didn't-bark.

Norris, P., 2023b. "Cancel culture: Myth or reality?" *Political Studies, 71*(1): 145–174.

Phillips, D.P., Kanter, E.J., Bednarczyk, B. and Tastad, P.L., 1991. Importance of the lay press in the transmission of medical knowledge to the scientific community. *New England journal of medicine, 325*(16), pp.1180-1183.

Porter, A. and Rafols, I., 2009. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81*(3), pp.719-745.

Posner, R.A., 2014. Economic analysis of law. *Aspen Publishing.*

Priem, J., Piwowar, H., Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv. https://arxiv.org/abs/2205.01833*

Redish, M.H., 2018. The value of free speech. In Freedom of speech (pp. 153-207). Routledge.

Resnik, D.B., 2008. Freedom of speech in government science. Issues in science and technology, 24(2), p.31

Reston, J.J., 2005. Galileo: A life. Beard Books.

Rubin, A. and Rubin, E., 2021. Systematic bias in the progress of research. Journal of Political Economy, 129(9), pp.2666-2719.

Scholars at Risk. "Free to Think." Report of the Scholars at Risk Academic Freedom Monitoring Project." (2021).

Series RWP23-020.

Shattuck, J. and Risse, M., 2021. Reimagining Rights and Responsibilities in the United States: Equal Access to Public Goods and Services.

Soucek, B., 2021. Diversity Statements. *UC Davis L. Rev., 55*, p.1989.

Teplitskiy, M., Duede, E., Menietti, M. and Lakhani, K.R., 2022. How status of research papers affects the way they are read and cited. *Research Policy, 51*(4), p.104484.

the dog-which-didn't-bark." Harvard Kennedy School Faculty Research Working Paper

Voerman-Tam, D., Grimes, A. and Watson, N., 2023. The economics of free speech: Subjective wellbeing and empowerment of marginalized citizens. *Journal of Economic Behavior & Organization, 212*, pp.260-274.

Warburton, N., 2009. Free speech: A very short introduction. OUP Oxford.

Whittington, K.E., 2018. Free speech and the diverse university. *Fordham L. Rev., 87*, p.2453.

Widmann, R., Rose, M.E. and Chugunova, M., 2022. Allegations of Sexual Misconduct, Accused Scientists, and Their Research. *Max Planck Institute for Innovation Competition Research Paper*, (22-18)

Wight, C., 2021. Critical dogmatism: Academic freedom confronts moral and epistemological certainty. *Political Studies Review, 19*(3), pp.435-449.

Wood, P.W., 2022. Free Speech in Academia. *Tex. Rev. L. & Pol., 27*, p.761.

Wu, A.H., 2018, May. Gendered language on the economics job market rumors forum. In AEA Papers and Proceedings (Vol. 108, pp. 175-179). 2014 Broadway, Suite 305, Nashville, TN 37203: *American Economic Association.*

Ziman, J. (1999). Rules of the game of doing science. *Nature*, *400*(6746), 721-721.

Zhang, F.J., 2023. Political endorsement by Nature and trust in scientific expertise during COVID-19. *Nature Human Behaviour*, *7*(5), pp.696-706.

8. Tables

Table 1: Descriptive statistics for the dependent variable, number of citations

treat	post	n	sd	min citations	mean citations	median citations	max citations
treated	pre	34951	3.26	0.00	2.04	1.00	19.00
treated	post	35847	3.15	0.00	1.86	1.00	19.00
control	pre	317703	3.11	0.00	1.91	1.00	19.00
control	post	326829	3.09	0.00	1.81	1.00	19.00

Table 2: Descriptive statistics of authors in the treated group in paper level analysis (n = 246)

variable	mean	sd	median	min	max
treatment year	2017.21	4.15	2019.00	2001.00	2021.00
citations per year	43.01	111.99	5.00	0.00	1341.00
year of first publication	1983.68	11.27	1981.00	1970.00	2015.00
academic age at treatment	33.53	11.63	35.00	1.00	51.00
number of publications per year	3.03	5.57	1.00	0.00	173.00
cumulative publnications	43.54	80.74	17.00	1.00	868.00
publication stock	15.36	25.59	7.72	0.00	301.48
citation stock	196.81	563.78	19.24	0.00	7015.62
citation over publication stock	8.27	15.26	2.92	0.00	213.39

Notes: academic age at treatment is simply the difference between the year in which the treated scholar is targeted and the year of her first publication.

Table 3: Parallel trends

Dependent Variable:	Number of Citations			
Model:	(1)	(2)	(3)	(4)
Variables				
treat × period-4	0.0220	0.0236	0.0236	0.0140
	(0.0237)	(0.0247)	(0.0249)	(0.0284)
treat × period-3	0.0347	0.0328	0.0328	0.0307
	(0.0216)	(0.0219)	(0.0221)	(0.0270)
treat × period-2	-0.0030	-0.0048	-0.0048	-0.0008
	(0.0162)	(0.0165)	(0.0166)	(0.0201)
treat × period0	-0.0212	-0.0190	-0.0190	-0.0218
	(0.0138)	(0.0140)	(0.0141)	(0.0169)
treat × period1	-0.0083	-0.0068	-0.0068	-0.0036
	(0.0184)	(0.0187)	(0.0188)	(0.0202)
treat × period2	-0.0412**	-0.0426**	-0.0426**	-0.0345*
	(0.0190)	(0.0192)	(0.0193)	(0.0199)
treat × period3	-0.0551**	-0.0597**	-0.0597**	-0.0595**
	(0.0239)	(0.0233)	(0.0235)	(0.0240)
treat × period4	-0.0380	-0.0403	-0.0403	-0.0403
	(0.0343)	(0.0330)	(0.0332)	(0.0338)
Fixed-effects				
Article	Yes	Yes	Yes	Yes
Year		Yes	Yes	Yes
Journal-Year			Yes	Yes
Fit statistics				
Observations	759,244	759,244	759,244	637,227
Squared Correlation	0.74364	0.74672	0.74672	0.73959
Pseudo R^2	0.51316	0.51449	0.51449	0.51158
BIC	3,222,777.4	3,217,803.3	3,355,790.2	2,730,696.7

Clustered standard-errors in parentheses at article, treated scholar, and matched group level.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 4: Main result

Dependent Variable:	Number of citations			
Model:	(1)	(2)	(3)	(4)
Variables				
post	-0.0647***	0.0709***	0.0709***	0.0714***
	(0.0179)	(0.0091)	(0.0092)	(0.0114)
treat × post	-0.0426***	-0.0421***	-0.0421***	-0.0406**
	(0.0159)	(0.0157)	(0.0158)	(0.0188)
Fixed-effects				
Article	Yes	Yes	Yes	Yes
Year		Yes	Yes	Yes
Journal-Year			Yes	Yes
Fit statistics				
Observations	759,244	759,244	759,244	637,227
Squared Correlation	0.74224	0.74649	0.74649	0.73934
Pseudo R^2	0.51222	0.51446	0.51446	0.51156
BIC	3,226,340.0	3,217,734.9	3,355,721.8	2,730,597.7

Clustered standard-errors in parentheses at article, treated scholar, and matched group level.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 5: Percentage of scholars in sub-groups

Variable	percentage of scholars
controversial outside their field	0.51
often controversial	0.32
right wing	0.73
extreme	0.10
fired, terminated or resigned	0.14
turning point	0.00
male	0.87

Table 6: Heterogeneity I

Dependent Variable:	Number of citations								
Model:	no extr, no ftr	+ no contr	+ extramural	+ acc. free	+ gpt no hate	+ gpt extramural	+ gpt acc.free	+ yes chicago	+ no chicago
Variables									
post	0.0714***	0.0743***	0.0529***	0.1099***	0.0729***	0.1132***	0.0599***	0.0795***	0.0694***
	(0.0099)	(0.0121)	(0.0191)	(0.0140)	(0.0091)	(0.0185)	(0.0154)	(0.0231)	(0.0145)
treat × post	-0.0486***	-0.0543***	-0.0770**	-0.0368**	-0.0259**	-0.0382*	-0.0659**	-0.0286	-0.0595**
	(0.0174)	(0.0204)	(0.0375)	(0.0180)	(0.0129)	(0.0228)	(0.0297)	(0.0328)	(0.0244)
Fixed-effects									
Article	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics									
Observations	596,106	448,637	237,547	211,067	645,415	138,269	307,846	87,277	361,273
Squared Correlation	0.74768	0.74821	0.75652	0.74100	0.74851	0.74211	0.75283	0.75524	0.74627
Pseudo R^2	0.51435	0.50900	0.50993	0.50740	0.51887	0.51361	0.50614	0.52252	0.50498
BIC	2,635,004.6	2,011,449.9	1,032,129.3	938,059.0	2,859,818.2	606,572.0	1,359,507.5	390,918.0	1,590,854.2

Clustered standard-errors in parentheses at article, treated scholar, and matched group level.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 7: Heterogeneity II

Dependent Variable:	Number of citations				
Model:	> 5 pubs	Tenure	No tenure	Support	No support
Variables					
post	0.0735***	0.0764***	0.0597***	0.0770***	0.0686***
	(0.0108)	(0.0103)	(0.0178)	(0.0228)	(0.0096)
treat × post	-0.0371**	-0.0357**	-0.0811**	-0.0087	-0.0487***
	(0.0175)	(0.0175)	(0.0397)	(0.0327)	(0.0179)
Fixed-effects					
Article	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
so_id-publication_year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	618,535	664,218	94,903	148,820	609,966
Squared Correlation	0.74889	0.74455	0.75834	0.73618	0.74878
Pseudo R^2	0.51553	0.51248	0.52622	0.51538	0.51377
BIC	2,720,519.2	2,920,437.5	399,643.5	602,166.6	2,707,774.9

Clustered standard-errors in parentheses at article, treated scholar, and matched group level.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 8: Future productivity

Dependent Variables:	Number of citations	Number of publication	Number of citations	Number of publication
Model:	(1)	(2)	(3)	(4)
Variables				
treat × factor(period)-4	0.0520	0.0464		
	(0.0537)	(0.1028)		
treat × factor(period)-3	0.0267	-0.0227		
	(0.0420)	(0.0852)		
treat × factor(period)-2	0.0164	-0.0933		
	(0.0344)	(0.0772)		
treat × factor(period)0	-0.0279	-0.0352		
	(0.0237)	(0.0810)		
treat × factor(period)1	-0.0648*	-0.2727***		
	(0.0337)	(0.0744)		
treat × factor(period)2	-0.1833***	-0.2977***		
	(0.0451)	(0.0819)		
treat × factor(period)3	-0.2112***	-0.3729***		
	(0.0544)	(0.0910)		
treat × factor(period)4	-0.2582***	-0.4643***		
	(0.0712)	(0.1134)		
post			0.0089	0.0183
			(0.0146)	(0.0306)
treat × post			-0.1552***	-0.2153***
			(0.0470)	(0.0644)
Fixed-effects				
Scholar	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	17,591	15,581	17,558	15,143
Squared Correlation	0.92851	0.66876	0.92672	0.66689
Pseudo R^2	0.92709	0.48193	0.92486	0.53802
BIC	290,977.5	83,800.0	288,979.1	86,461.6

Standard-errors clustered at the scholar level in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 9: Test mechanism for publications

Dependent Variable:	Number of publication	
Model:	Support	No support
Variables		
post	0.0950*	-0.0064
	(0.0555)	(0.0355)
treat × post	-0.1782*	-0.2523***
	(0.0982)	(0.0613)
Fixed-effects		
Scholar	Yes	Yes
Year	Yes	Yes
Fit statistics		
Observations	4,006	10,637
Squared Correlation	0.61983	0.67776
Pseudo R^2	0.45725	0.48097
BIC	20,970.7	56,829.3

Clustered (Scholar) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

9. Figures

Figure 1 Number of affected scholars per year

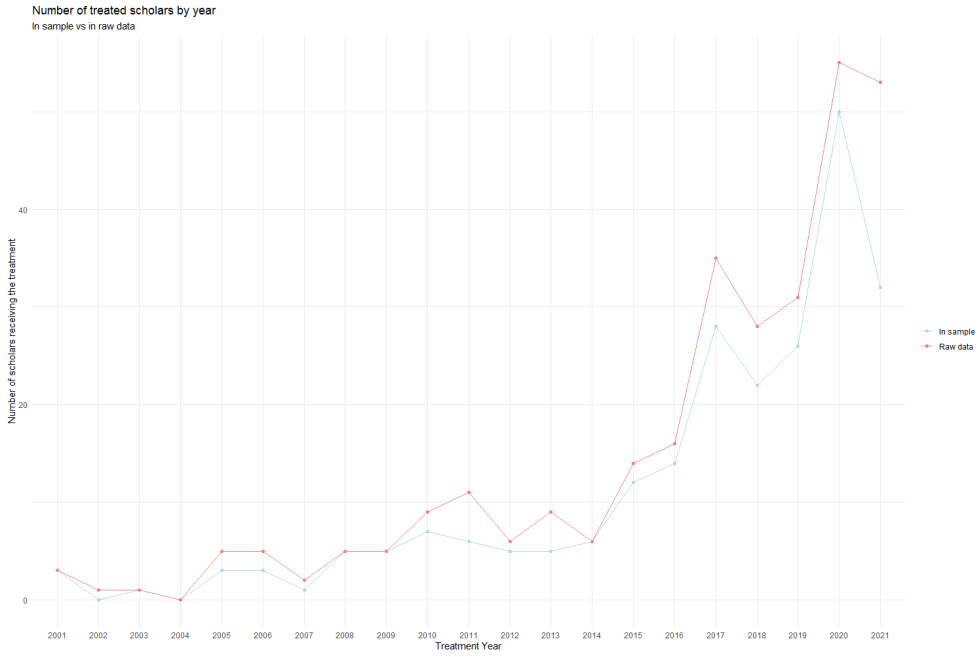

Figure 2 Parallel trends for different datasets

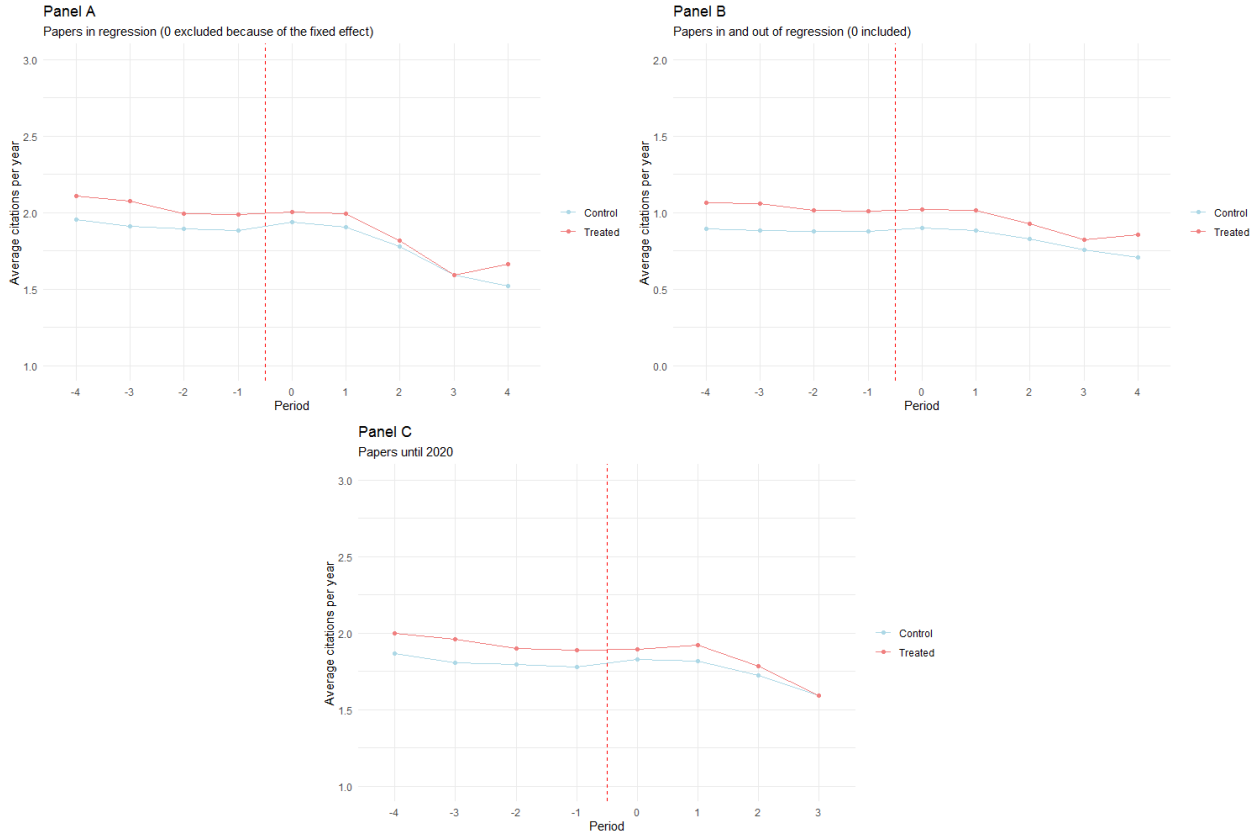

Figure 3 Distribution of the number of citations

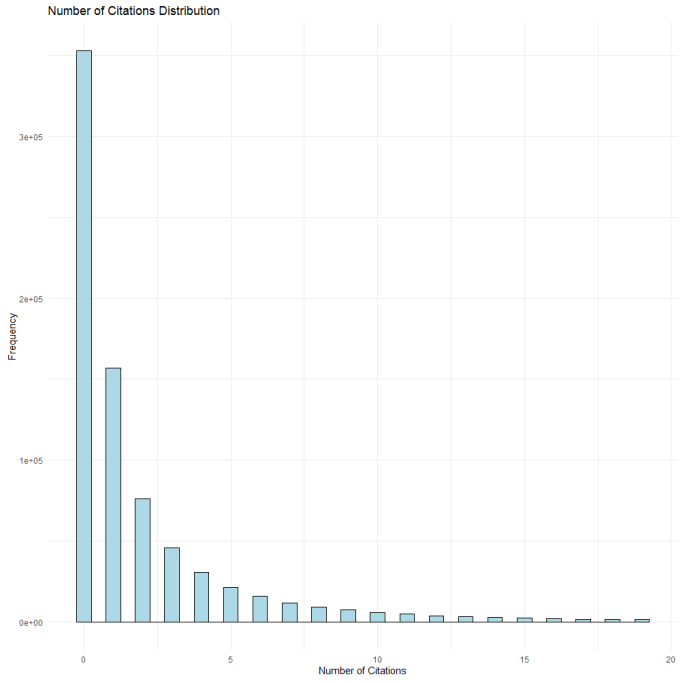

Figure 4 Test mechanism 1: citations from scholars affiliated to the same institutions of the affected scholars vis-à-vis citations from scholars affiliated to other institutions

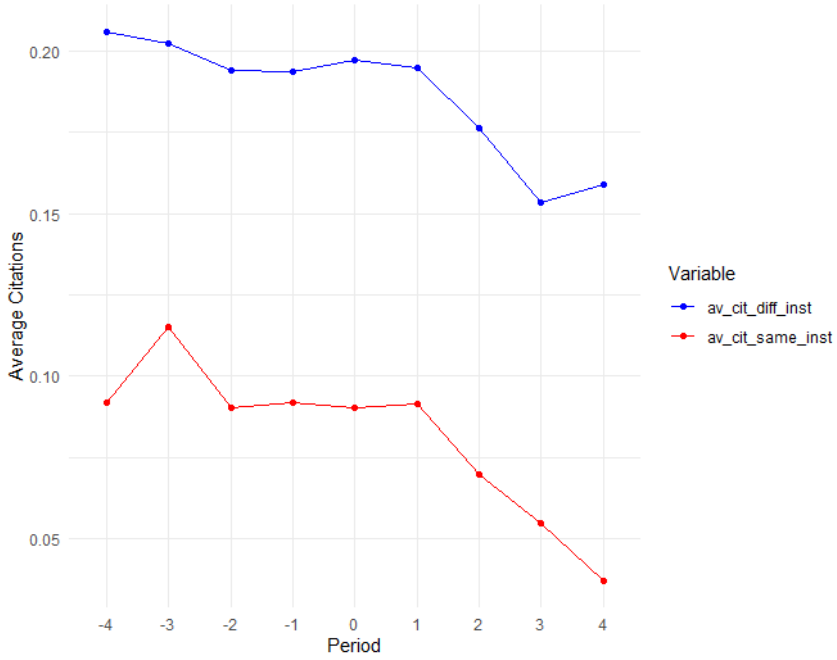

Figure 5 Test mechanism 2: citations from scholars in the co-authorship network of the affected scholars vis-à-vis citations from scholars not in the co-authorship network