Discussion Paper No. 17-051

Guilt by Association: How Scientific Misconduct Harms Prior Collaborators

Katrin Hussinger and Maikel Pellens

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 17-051

Guilt by Association: How Scientific Misconduct Harms Prior Collaborators

Katrin Hussinger and Maikel Pellens

Guilt by Association:
How Scientific Misconduct Harms Prior Collaborators

Katrin Hussinger[a,b,c] and Maikel Pellens[b,c,*]

[a] *University of Luxembourg (Luxembourg)*
[b] *Centre for European Economic Research (ZEW), Mannheim (Germany)*
[c] *K.U. Leuven, Dept. of Managerial Economics, Strategy and Innovation (Belgium)*

Abstract: Recent highly publicized cases of scientific misconduct have raised concerns about its consequences for academic careers. Previous and anecdotal evidence suggests that these reach far beyond the fraudulent scientist and her career, affecting coauthors and institutions. Here we show that the negative effects of scientific misconduct spill over to uninvolved prior collaborators: compared to a control group, prior collaborators of misconducting scientists, who have no link to the misconduct case, are cited 8 to 9% less afterwards. We suggest that the mechanism underlying this phenomenon is stigmatization by mere association. The result suggests that scientific misconduct generates large indirect costs in the form of mistrust against a wider range of research findings than was previously assumed. The broad fallout of misconduct implies that potential whistleblowers might be disinclined to make their concerns public in order to protect their own reputation and career.

Keywords: scientific misconduct; prior collaborators; stigma

JEL: O31, O33

---

*: corresponding author. Address: Centre for European Economic Research (ZEW), department of Economics of Innovation and Industrial Dynamics, L7, 1, D-68161 Mannheim, Germany. Fax: +49 (0)621 1235-170. Email: maikel.pellens@zew.de
Acknowledgements: The authors thank Paul David, Rainer Frietsch, Ben Jones, Ben Martin, Andy Toole, and John Walsh for valuable comments and inputs. The authors also thank Dimitris Kafteranis for excellent research assistance. Previous versions of the paper were presented in the LEI & BRICK workshop on the Organization, Economics, and Policy of Scientific Research, the DRUID Society Meeting, the Technology Transfer Society Conference, the Asia-Pacific Innovation Conferences, the EARIE, and research seminars at TU München, the University of Luxembourg, ZEW, EPFL Lausanne, Georgia Tech, and the University of Kassel. Any errors remain the author's own.

1. Introduction

Scientific misconduct affects many more than those who commit it. Consider for instance the case of the Center for Developmental Biology in Kobe, Japan, where in 2014 two retractions due to scientific misconduct led to a change in directors, half of its laboratories closing, merging, or moving elsewhere, and a budget slash of 40% (Cyranoski, 2015). In other cases, such as the one involving Dutch social psychologist Diederik Stapel, the work of graduate students was called into question due to the unethical actions of their supervisor.[1]

The question arises how far the ripple effect of scientific misconduct reaches. Prior research has shown that co-authors of misconducting scientists experience significant drops in their publication flows (Mongeon and Larivière, 2014). While research on actual misconduct cases is scarce and due to the limited amount of cases rather descriptive (Lubalin and Matheson, 1999; Pozzi and David, 2007; Redman and Merz, 2008; Resnik and Dinse, 2012; Reynolds, 2004; Rhoades, 2004), we can refer to a related strand of the previous literature that has focused on retractions of journal articles. Retractions can occur in response to both scientific misconduct, but can also be due to honest mistakes (Azoulay et al., 2014a; Fang et al., 2012; Van Noorden, 2011). Despite the fact that a common cause of retraction is the honest reporting of a mistake by an author, retractions have been shown to have negative implications for the citations to prior and future work of the authors themselves (Azoulay et al., 2015, 2014a; Jin et al., 2013; Lu et al., 2013), articles conceptually related to the questionable science (Lu et al., 2013), and the narrowly defined research field as a whole (Azoulay et al., 2014a).

[1] Cf. the report on Mr. Stapel's fraudulent research (Levelt Committee et al., 2012): the work of many of the doctoral dissertations overseen by him needed to be reassessed since they were based on fraudulent data.

Here we show that the damage of scientific misconduct reaches further, affecting prior collaborators who collaborated in the past with a scientist found guilty of scientific misconduct, but who are not involved in the misconduct case. Our empirical analysis shows that prior collaborators face a citation penalty of 8 to 9% in the aftermath of a scientific misconduct case. We base this result on misconduct cases investigated by the U.S. Office of Research Integrity (ORI) between 1993 and 2008, which oversees misconduct investigations for research funded by the National Institute of Health (NIH) and the Public Health Service (PHS). Hence, our database is based on the complete list of investigated and well documented scientific misconduct cases at the world's largest funder of medical research.

This observation fits a theory of stigma spreading through mere association (Pontikes et al., 2010), where scientists are punished by the scientific community for being associated with a misconducting – and hence deviant from the norms – scientist. Crucially, they are punished for associating with scientists who afterwards turns out to have engaged in misconduct, but they did not choose to associate with a known fraudulent scientist. The effect we observe can be interpreted as an extended loss of trust in fellow scientists in the wake of misconduct. Trust plays an important role in science, considering its cumulative nature (Merton, 1973) and the lacking incentives and resources for replication of results (Hamermesh, 2007). While a loss of trust in the work of misconducting scientists can be seen as rational behaviour and as a means of self-regulation of the scientific community (Azoulay et al., 2015), it turns into wasteful ignorance of public knowledge if it spills over to others.

To the best of our knowledge, there is no solution to this problem. Transparency about guilt, a presumably obvious measure, apparently is a necessary but not sufficient precondition as the U.S. ORI already publishes detailed information about the investigated cases, including the names of the scientists found guilty. This should help the community to distinguish fraudulent scientists from innocent bystanders, yet, we do not find innocent bystanders to emerge unscathed from a misconduct scandal.

One important implication of our findings is that incentives for scientists to whistle blow on unethical behaviour of their collaborators are very low: scientists are unlikely to draw attention to foul play if they expect to be considered to be complicit by association. Such behaviour hinders the self-correction mechanism that science relies on. This implies that it is even more important to prevent misconduct at the bench, where it can be corrected without affecting the careers of others.

We contribute to the previous literature in several ways. First, by drawing on the theory of social stigmatization (Goffman, 1963; Pontikes et al., 2010) we provide a general explanation for previously documented negative implications of scientific misconduct and retractions on closely related parties (Mongeon and Larivière, 2015) and the research field (Azoulay et al., 2014a). These articles employ three different mechanisms to describe their findings. The first one is Bayesian updating of beliefs (Azoulay et al., 2015, Jin et al., 2013), predicting that after the surfacing of retractions, involved scientists are considered to be of lower 'quality' than before. Second, a loss of trust may lead scientists to avoid citations to work associated with misconduct – even indirectly – in order to protect their integrity (Fuchs and Westervelt, 1996). Lastly, the actual or feared invalidation of part of the scientific field by a retraction may incentivize scientists to move to different topics (Azoulay et al., 2014a). The concept of stigmatization allows us to connect these different mechanisms. Caused by a *loss of trust*, social stigmatization can be the reason for a *Bayesian updating of beliefs* about the "quality" of prior collaborators and the means by which the members of the scientific community *protect their own scientific integrity*.

Second, our study focuses only on investigated cases of scientific misconduct in the field of medical research. Other than most of the previous studies that are based on retractions (Azoulay et al., 2015, 2014a; Jin et al., 2013; Lu et al., 2013; Mongeon and Larivière, 2015), we base our findings on an analysis of the Findings of Research Misconduct published by the ORI, the central U.S. authority that is responsible for investigating scientific misconduct cases in biomedical research funded by the NIH throughout the country, between 1993 and 2008. This data constitutes a complete list of cases of scientific misconduct committed by NIH grant recipients and grant recipients at the PHS investigated by the ORI. We thus avoid the considerable uncertainty inherent to retraction-based analyses and instead base ourselves on confirmed cases of scientific misconduct (Azoulay et al., 2014a; Fang et al., 2012; Van Noorden, 2011). In doing so, we expand the literature analysing misconduct at the ORI (Lubalin and Matheson, 1999; Pozzi and David, 2007; Redman and Merz, 2008; Resnik and Dinse, 2012; Reynolds, 2004; Rhoades, 2004; Wright et al., 2008).

Third, we contribute to understanding the consequences of misconduct on science by empirically showing that the implications of scientific misconduct spread beyond the scientists involved in the project and the fraudulent scientist's institution. Specifically, our analysis shows that, misconduct scandals affect prior collaborators. Hence, the potential fallout of a misconduct scandal is much wider than one would assume.

The remainder of the article is organized as follows. In the next section, we develop our theoretical framework. Section three provides an overview on the institutional framework in the U.S. Section four details our methodology, before section 5 presents the data along with descriptive statistics. The results are presented in section six. Section seven concludes.

2. Conceptual framework

This section describes the key features of the scientific system that help our understanding of the incentives for scientific misconduct. A second part draws from sociological research in order to introduce the concept of social stigmatization and apply it to the context of science and misconduct.

2.1. Science and misconduct

Trust is an important pillar of the scientific system (Dasgupta and David, 1994). This is due to its distinctive features, including the way in which scientists compete, the cumulative nature of science (Merton, 1973), and the freedom that scientists enjoy when choosing their projects (Stern, 2004, Aghion et al., 2008).

Scientists compete in a race for priority, gaining reputation when they are first to publish new discoveries in scientific journals. The rewards for publishing include access to further resources, prestigious jobs offers, contacts to peers in science, as well as other lucrative opportunities inside and outside the academic sector (Merton, 1973). The race for priority is a winner-takes-all game: the second to finish wins (almost) nothing, even though her investment might have been large. This structure is beneficial for society as it sets incentives for scientists to disclose their discoveries as soon as possible, thus allowing follow-up research in a speedy manner and avoiding the duplication of research streams (Merton, 1957, Dasgupta and David, 1994, Stephan, 2012). The nature of scientific competition hence pushes scientific progress. The absence of formal intellectual property rights in the normal scientific process further accelerates the dissemination and adaption of scientific results. The fact that third parties can freely access, exploit and modify a published idea is well in lines with the scientists' pursuit of maximizing reputation (Fleming and Sorenson, 2004; Murray and O'Mahony, 2007)

The downside of the winner-takes-all competition in science is that it creates incentives to cheat. As in sports, a laggard might find it beneficial to use forbidden means in order to increase the chances of winning a race. Classic game theory predicts that it can be a rational strategy to cheat in these settings, at least when the chances of being caught are relatively low (Nalebuff and Stiglitz, 1983). This seems to be the case in science because incentives for replication and other forms of double checking published results are low (Dewald et al., 1986; Hamermesh, 2007). Kiri et al. (2015) as well as Lacetera and Zirulia (2011) investigate the peer review system in a game theoretical setting. The predictions of their models include a certain positive equilibrium level of misconduct, and the worrisome observation that more stringent verification procedures do not necessarily lead to less misconduct.[2]

In order to establish the importance of trust in science, two more features of the system of science production need to be understood. The first one is the well-known cumulative nature of science (Merton, 1973). Science evolves along specific lines of research where scientists use the prior insights of their colleagues as foundation and stepping stones for their own research (Azoulay et al., 2014a; Mokyr, 2002). The second distinct feature is that science also differs from alternative knowledge production systems insofar as scientists are enjoying substantial discretion in choosing

[2] Lacetera and Zirulia (2011) consider four policy levers through which misconduct might be discouraged: facilitating replication studies, easing the "publish or perish" paradigm, increasing punishment for being caught, and involving journal editorials boards in checking for fraud. Although common knowledge dictates that increasing either should result in less fraud, they find that none of these policies is able to eliminate fraudulent behaviour and that some may in fact lead to more fraud. In their model, a reduction of costs for checking for fraud leads to a change in the degree of novelty of research that is conducted. Alleviating pressure to publish leads to more rather than to less fraud. Higher punishment for fraud can deter an author from cheating, but might also reduce incentives for checking. A more active role of the editorial board could crowd out incentives for the reader to check the published results. In a follow up paper, Kiri et al. (2015) focus explicitly on the interplay between investigating efforts in high quality research and efforts spend on validating research results produced by others. More validation leads to a higher research quality level, and low quality research can be eliminated if the incentives for verification and confirmatory research are high enough. However, a high number of scientists performing validation activities may reduce the overall level of validation activities as well as the overall quality level of research as the rewards need to be shared among many. Moreover, in research areas with a high level of collaboration, scientists may collude and not validate each others' research findings if they have doubts.

their research projects (Aghion et al., 2008; Stern, 2004). As it is not possible to personally verify all prior research results related to a planned project, science depends in great degree on trust. It is thus only rational that scientists shy away from the results of authors that have been involved in retractions in the past (Azoulay et al., 2015, 2014a; Lu et al., 2013; Mongeon and Larivière, 2015).

In line with this, it has been observed that scientists, eager to protect their reputation and wanting to avoid wasteful investments, avoid starting research projects in research areas characterized by a larger amount of retractions (Azoulay et al., 2014a; Lu et al., 2013). Closely related to our study, Azoulay et al. (2014a) investigate the impact of retractions on cumulative science. They show that intelectually related articles receive fewer citations after a retraction occurs, with a stronger decline when the retraction is due to misconduct. Furthermore, scientists tend to avoid research fields in which retractions occur. While Azoulay et al. (2014a) are interested in the cumulativeness of research, employing retractions as a disruption to a research line, we are interested in the implications of scientific misconduct on prior collaborators of the fraudulent scientist investigating the role of trust in scientific communities.

2.2.Stigma and mere association

The fact that retractions spread beyond the scientists involved in the actual publication has been explained in the prior literature by rational Bayesian updating of beliefs (Azoulay et al., 2015; Jin et al., 2013), and the fear to start projects standing on “shaky shoulders” (Azoulay et al., 2014a). Since we are interested in the effect of misconduct on prior collaborators – who are more distant in time and in the sense that they are not directly involved in the fraud and are not necessarily thematically related to the fraudulent work - we draw in this section from sociology where the notion that bystanders suffer from being unjustifiably associated with others’ actions is well-established, and finds expression in the concept of stigmatization (Goffman, 1963).

A stigma is defined as a perceived undesirable deeply discrediting attribute that disqualifies an individual from full social acceptance (Goffman, 1963). Attributes evoking a stigma can evolve along the attributes of an individual's social identity which include physical appearance, professional activities and self-concept (Goffman, 1963). Stigmatization can hence be described in the general case as a flexible process involving labelling, stereotyping, separation, status loss and discrimination (Link and Phelan, 2001). Put into the context of the scientific community, individuals who commit scientific misconduct are likely to be stigmatized because they disrespect the norms of the scientific community. They are labelled to be cheaters, stereotyped into being generally untrustworthy, seen as different from the social norm of the scientist, and rejected by the scientific community through the ignoring of their findings and exclusion from common necessities such as scientific funding.

Stigma can spread by virtue of association: an individual can be stigmatized purely because of an association, rather than because of discrediting attributes (Goldstein and Johnson, 1997; Mehta and Farina, 1988). 'Stigma by association' (Goldstein and Johnson, 1997) or 'courtesy stigma' (Goffman, 1963) or 'associative stigma' (Mehta and Farina, 1988) is well studied in laboratory settings (e.g. Goldstein & Johnson, 1997; Mehta & Farina, 1988; Neuberg, Smith, Hoffman, & Russell, 1994) and has been applied to a variety of different sources for stigmatization such as mental illness (e.g. Lefley, 1987), homosexuality (e.g. Swim et al., 1999), and criminal behaviour (e.g. Levenson and Tewksbury, 2009). It has also been observed in corporate settings through industry-wide implications of firm-specific scandals (Barnett and Hoffman, 2008; Barnett and King, 2008; Kostova and Zaheer, 1999; Smith et al., 2010) and wage losses of executives associated with firm wrongdoing (Groysberg et al., 2016).

Stigma transfers quickly: A single juxtaposition with a stigmatized individual can be enough to create further stereotyping (Risen et al., 2007). Stigma by association occurs for myriad reasons[3] and is persistent. At its core, stigmatization by association is found to exist because of the assumption that individuals choose to associate with the stigmatized individual, thus justifying stereotyping of the associated individual (Pontikes et al., 2010).

Essential to our setting is that stigmatization by association can occur ex post – an individual is stereotyped because he chooses to associate with the stigmatized group – or ex ante – an individual experiences stigmatization because of association with a group before it became stigmatized. The latter case is called stigma by mere association (Pontikes et al., 2010). Pontikes et al. (2010) show the existence of stigmatization by mere association among Hollywood actors in the times of the Red Scare: being merely associated with a blacklisted individual significantly reduced actors' chances of being employed in feature films. The authors conclude that "stigma by association can lead to false positives and harm many innocents" (p.456).

Our setting is one of mere association. Scientists do not expect their collaborators to engage in illicit behaviour. Yet, when a scandal surfaces they see part of their oeuvre tainted by the misconduct stigma. In the worst case, they are not complicit in producing a piece of false science and are stereotyped into fraudsters themselves. Empirical evidence supports this idea: co-authors of retracted studies see their citation rates drop (Jin et al., 2013; Mongeon and Larivière, 2015). In our case, the scientists of interest were not directly involved in false science. However, they can still be stigmatized because of stereotyping: the (original) misconducting author is classified as being a fraudster (even though evidence might exist for only one case of misconduct), which makes all of his collaborators, past, present and future, associated to fraud by mere association. This forms

[3] An extreme example of arbitrary stigmatization is found in a study presenting lower expectations of job candidates that sit next to an obese female (Hebl and Mannix, 2003).

the main hypothesis of the paper: mere association with scientific misconduct, through prior collaboration, leads scientists to be caught in misconduct scandals.

The question arises whether such behavior can be seen as rational. In our view, the answer is yes. As scientists need to resort to heuristics, including reputation, to assess the quality of the work of their colleagues (Merton, 1968, p.59), association to misconducting authors can be interpreted as a negative quality signal. Rational herding behavior (Banerjee, 1992) may add to this. If scientists see that peers shy away from a researcher's prior work they may believe that these colleagues have more information about the honesty of the researcher in question and her involvement in the misconduct case.

The next question is whether such behavior is welfare enhancing or destroying. Ignoring research results is welfare enhancing when they are invalid or "shaky" (Azoulay et al, 2014a). Ignoring valid research results is welfare destroying. We believe that the ignorance of research of prior collaborators of a fraudulent scientist is to a large extend welfare destroying, as they are not directly connected to fraudulent research, apart from an indirect association through prior collaboration, and incidental overlaps in topic. Neither of these, however, are informative of the quality of the ignored research.

3. Institutional Context

The empirical analysis is based on Findings of Research Misconduct published by the Office of Research Integrity (ORI) in the U.S. Since 1993, the ORI has the responsibility to oversee and direct activities concerning matters of research integrity within the Public Health Services. This includes intramural and extramural scientific misconduct in research funded by the National Institutes of Health (NIH) and other agencies of the PHS, but does not include activities concerning regulatory research integrity by the Food and Drug Administration.

When the ORI receives allegations of scientific misconduct, either directly or through the funding institutes, the ORI decides whether or not to pursue the allegation through an inquiry. Note that institutions can also pursue an inquiry independently. If grounds are found to continue, the inquiry becomes an ORI investigation. If evidence of scientific misconduct is found, the ORI publishes the results in its Findings of Research Misconduct, as well as in its annual report.[4] These reports transparently and in great detail describe the misconduct case: the persons involved, where and when misconduct took place, the nature of the misconduct, any publications that are affected by it, and any administrative sanctions taken such as institutional oversight, exclusion from referee or advisory boards, exclusion from government contracting, and/or exclusion from further NIH or PHS funding. This level of transparency allows the identification of scientists found to be guilty of scientific misconduct.[5] The ORI maintains the following definition of research misconduct: "Research misconduct means fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results. Fabrication is making up data or results and recording or reporting them. Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record. Plagiarism is the appropriation of another person's ideas, processes, results or words without giving appropriate credit. Research misconduct does not include honest error or differences of opinion" (Office of Research Integrity, 2011).

There are previous studies focusing on the ORI cases of scientific misconduct (Lubalin and Matheson, 1999; Pozzi and David, 2007; Redman and Merz, 2008; Resnik and Dinse, 2012; Reynolds, 2004; Rhoades, 2004; Wright et al., 2008). Earlier studies (Rhoades, 2004, Reynolds,

[4] See http://ori.hhs.gov/historical-background. The first ORI annual report was published in 1994.
[5] A more detailed description of the ORI's course of action in misconduct investigation and statistics on previous investigations can be found in Roades (2004).

2004, Pozzi and David, 2007) provide detailed descriptive evidence on the ORI misconduct cases, looking into the different types of accusations and outcomes, sources of funding, trends over time, etc. Resnik and Dinse (2012) investigate the correlation between information provided by the journal retraction note and the outcome of the ORI investigation. Redman and Merz (2008) consider the career consequences of being found guilty by the ORI, finding that being found guilty of misconduct by the ORI associates with severe drops in publication rates (with approximately a third dropping out of publishing completely), and a high incidence of leaving university employment. Lubalin and Matheson (1999) survey whistle blowers and individuals accused of having been involved in scientific misconduct but exonerated, investigating long-run and short-run professional and personal consequences for both groups. They find that in the short-run, during the period of investigation, whistleblowers fare worse than those accused of misconduct. While both groups report little long-term impacts, those accused of misconduct report worse consequences in terms of mental and physical health as well as self-esteem and self-identity. Lastly, Wright et al. (2008) point out that the majority of misconduct cases is associated with a lack of oversight by mentors. In more than half of the investigated cases, stress was reported as a main or contributing factor to misconduct.

4. Method

We aim at investigating how scientific misconduct affects the reputation of prior collaborators, as proxied through the accumulation of citations. Citations to publications are a widely used indicator for the importance of scientists and their scientific findings, reflecting to which extent results and insights are used as building stones for future research. They have been shown to correlate with other measures of a scientist's influence such as awards, honours and Nobel laureateships (Cole and Cole, 1967; Inhaber and Przednowek, 2007; Myers, 1970) as well as with peer judgements (Aksnes and Taxt, 2004). Keeping the quality of the scientist and other factors constant, a drop in citations reflects a loss of trust in the scientists' work by the scientific community. We start the analysis with a descriptive exercise to detect a change in the time trend of the citations received by an innocent prior collaborator. This shows whether a drop in citations occurs around the point in time when the misconduct case was published. Since a so-detected dip in the trend could be driven by time-varying macro effects we employ a control group of comparable scientists that were not associated to a misconduct case. These scientists serve as a benchmark, a control group, and allow us to abstract from influences other than the publication of the misconduct case that could impact the citations that the publications of a researcher that was collaborating with a misconducting scientist, the treatment group, receive.

To evaluate whether there is a causal effect of the publication of a misconduct case on the citations to prior collaborators, we employ a difference-in-difference method that allows us to compare the evolution of citations to prior collaborators of misconducting scientists relative to citations to the control group of scientists. We control for scientist career age and unobservable ability in a multivariate setting. More specifically, we estimate an equation of the form:

$$\mathrm{Citations}_{\mathrm{it}} = \psi_1 \mathrm{T}_{\mathrm{t}} + \psi_2 \mathrm{Pt}_{\mathrm{i}} + \psi_3 \mathrm{T}_{\mathrm{i}} * \mathrm{Pt}_{\mathrm{t}} + \beta \Gamma_{\mathrm{i}} + \varphi_{\mathrm{t}} + \xi_i + \ln(Publications_{it}) + \varepsilon_{\mathrm{it}}$$

Where $\mathrm{Citations}_{\mathrm{it}}$ represents the total number of citations accumulated by articles published by author i in year t by the time the data were collected. In the remainder of the analysis, we refer to this measure as 'aggregate citations'.[6] As citation counts follow a count

[6] The data were collected in late 2014. As we only consider misconduct cases up to 2008, this leaves reasonable time for citations to accumulate, even three years after the *Findings* have been published. We control for any remaining time-driven variation in citations through a set of publication year dummies. We additionally show the robustness of our results to employing a citation flow indicator, representing the number of citations an author received in a year regardless of the year in which the cited article was published.

distribution, we estimate a Poisson model. We offset aggregate citations by publication output by including the natural logarithm of the number of publications issued by the author in the year, with coefficient constrained to one, to account for differences in publication output.

T_t and Pt_i represent the core of the model. The first variable is an indicator variable that takes the value one if the author is in the treatment group, and zero otherwise. The second variable takes value zero in pre-treatment years, and one after treatment. The estimated coefficients of these variables, ψ_1 and ψ_2, thus capture any systematic differences in aggregate citations between the treatment and control group, and between any shared differences pre- and post-treatment. The main result of the model is provided by ψ_3, which captures the average difference in change of aggregate citations between control and treatment observations after the misconduct was discovered. If scientists in the treatment group experience a drop in aggregate citation after being associated with scientific misconduct, while scientists in the control group do not, ψ_3 takes a negative and significant coefficient.

Γ_i represents a vector of individual-specific factors that affect citations. One such factor is career age, as measured by the time since the authors' first publication in a scientific journal. This accounts for the fact that authors have changing levels of commitment to publishing as their career progresses (e.g. Stephan and Levin, 1992). We include career age in linear and squared terms to account for possible life time effect of scientific productivity of researchers.

Another factor is talent, or ability. As this is usually impossible to observe directly by the econometrician, we control for inherent differences in citation rates either through unobserved (fixed) effects models or by applying a pre-sample average estimator (Blundell et al., 1999).[7] Finally, φ_t captures common time trends through a set of year dummies, and ξ_i captures that misconduct can have a quite heterogeneous impact through a set of case dummies.

5. Data

5.1. Data sources and construction of the database

The analysis is based on the *Findings of Research Misconduct* as published by the Office of Research Integrity (ORI). The *Findings* are concerned with misconduct cases of National Institute of Health (NIH) grant recipients and grant recipients at the Public Health Service (PHS). The analysis makes use of the 36 cases published between 1993 and 2008 in which at least one scientific publication with a Pubmed[8] identifier was affected. We focus on misconduct cases involving retractions or corrections only[9] in order to restrict the analysis on misconduct with real scientific impact.

We should note that research funded by the NIH or PHS is not representative of all research performed in biomedicine: given the eminence associated with being NIH-funded, we expect misconduct cases investigated by the ORI to be more high-profile than the average misconduct case. This is corroborated by the fact that these misconduct cases involve significant sanctions, barring researchers temporarily or permanently from NIH funding, contracting with the PHS, or subjecting them to institutional oversight (Redman and Merz, 2008). Given that the

[7] The pre-sample mean estimator suggested by Blundell et al. (1999) accounts for unaccounted heterogeneity due to unobservables by including an additional parameter in the model which contains the pre-sample average outcome. This restrains our models to prior collaborators whose publication outputs has been observed for at least four year at the time the *Findings* were published.
[8] Pubmed is a search engine for the bibliometric database MEDLINE (Medical Literature Ananlysis and Retrieval System Online). This database covers more than 5,500 biomedical journals.
[9] Others include faking of credentials, of affiliations, and fraud in applications, often occuring in the application phase for an NIH grant.

NIH accounts for a substantial share of biomedical research funding in the U.S., 28% in 2008 (Dorsey et al., 2010), and that grant recipients constitute top researchers, it is a relevant setting to study nevertheless.

We retrieved all of the misconducting author's publications and citations from the Scopus publication database in late 2014. Through these publications we identified prior collaborators of the misconducting authors in the five years prior to the publication of the case in the *Findings*. Throughout the analysis, we take articles, conference papers, notes, reviews, and short surveys into account. Letters, books, and other document types are not taken into consideration.

Our dependent variable is the aggregate number of citations that the articles published in the focal year receive over the years until late 2014 when the data was collected. As we only consider misconduct cases up to 2008, this leaves reasonable time for citations to accumulate, even three years after the *Findings* have been published. In the robustness check section, we also use the annual citation flow as an alternative dependent variable. This data is available from Scopus from 1996 onwards only.

Authors who also collaborated on work mentioned in the *Findings* were excluded from the set of collaborators to ensure that the effects found only stem from association, and not direct implication. We also removed all articles co-authored with the misconducting scientist, as citations to these articles could be affected by negative attention directed to the misconducting author. Furthermore, we removed authors who were co-authors of more than one misconducting scientist, as these are subject to multiple treatments. After data cleaning we arrived at 929 unique co-authors that published with the misconducting scientists in the five years prior to the publication of the *Findings*. Our final sample is somewhat smaller, consisting of 856 treated co-authors as we need to observe publication outputs at least four years before the misconduct was published in order to be able to apply a pre-sample correction for unobserved ability.

5.2. Construction of control group

To construct a control group we need to identify prior collaborators comparable to the treated prior collaborators, i.e. those that collaborated with a fraudulent NIH or PSH grant recipient in the past. We achieve this by selecting prior collaborators of a scientist funded by a similar grant of the same agency as the misconducting scientist. Selection based on grant receipt is essential since we observe a positive selection when focusing on scientists that receive grants. To do this, we used the NIH grant database made available by Pierre Azoulay, and is referred to as Grant2pmid database.[10] This database covers NIH grants since 1971, and lists grant numbers, general information about the grant, and any publications (indexed by Pubmed identification numbers) which list the grant as source of funding.

The control group was built as follows. We selected a NIH grant which was listed as funding on the corrected or retracted publication mentioned in the *Findings* ('treatment grant'). We then match a control grant (using the NIH grant database) with similar characteristics in terms of medical research area, type of grant, grant year, and grant duration as the treatment grant ('control grant'). We match on research area by selecting grants issued by the same NIH institute. The National Institutes of Health consist of 27 institutes and centers, which are divided thematically. Thus, grants issued by a center should be in comparable fields. Examples of institutes are the National Cancer Institute (NCI), the National Human Genome Research Institute (NHGRI), and the National Institute of Diabetes and Digestive and Kidney Diseases

[10] http://pazoulay.scripts.mit.edu/Data.html

(NIDDK). Selecting the grant of a similar type ensures that the grantees are in a comparable career position. While the database does not include information on the size of the grant, we proxy for this through the duration of the grant.

We then selected a random author of a random publication supported by the control grant to serve as a match for the misconducting grant recipient and used Scopus to identify this author's prior collaborators in the five years before the *Findings* associated with the corrected or retracted publication were published. This results in a final control group of 1,149 co-authors. The process is summarized in Table 1.

Table 1: Summary of Data Gathering Protocol

1. For each misconduct case, select a NIH grant acknowledging funding of one of the papers listed as affected in the *Findings*
    - Pick the grant on the earliest paper listed in the *Findings* which acknowledges funding
    - Pick the earliest grant in case the paper acknowledges more than one NIH grant
2. Match the so retrieved 'treatment' NIH grants to NIH 'control' grants with similar characteristics as the treatment grants, considering granting institute, grant type, grant duration, and grant year as matching parameters.
3. For each control grant, select one publication funded by it
    - In case more than one paper was supported by the control grant, the publication published as closely as possible to the treatment publication was chosen.
4. Select a random author of these publications to form the control group of the misconducting authors
5. Collect the collaborators of the control authors in the five years before the relevant *Findings* were published.

6. Results

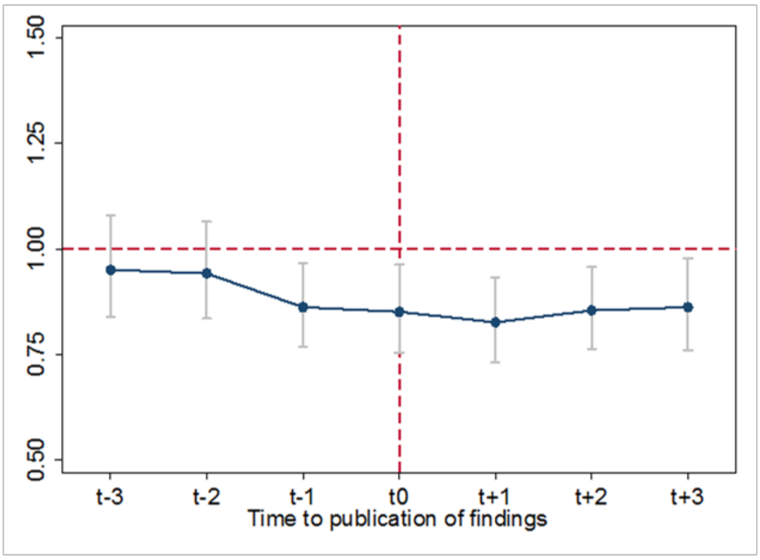
6.1. Visual Inspection

To start with a visual description of the effect of the publication of the *Findings* on prior collaborators compared to control authors, we estimate the evolution of the aggregate citations over time. To do so, we estimate a Poisson model on the author-year level interacting the treatment indicator with each year relative to the publication of the *Findings*, controlling for calendar year and case effects. We further control for the number of articles published in the year. The resulting coefficients of the interaction terms thus represent the difference between treatment and control collaborators over time.

Figure 1 plots the effects and 95% confidence intervals. Three and two years before the misconduct is revealed, there is no significant difference in aggregate citations between the articles published by treatment and control authors, confirming that the control authors are similar to the treatment authors in that time period. In the year before the misconduct is revealed, aggregate citations drop to 86% compared to those of control authors, and hover around this proportion until $t+3$.

The fact that the aggregate citations start decreasing in t_{-1}, i.e. before the scientific misconduct case is published, is likely explained by the presence of rumours of misconduct during the investigation: we take the end of the investigation, the publication of the *Findings*, as treatment, but reputational damage could occur much earlier. The ORI's investigation takes on average approximately half a year to complete, and rumours of potential misconduct could spread throughout the scientific community even before allegations reach the ORI (Rhoades, 2004). The dip in $t-1$ is likely to be further deepened by the fact that we plot aggregate citations and not the flow of citations in *t*: articles published just before the misconduct case was revealed could still be affected, as a large part of their citations would only emerge after the *Findings* have been published.

Figure 1: Average aggregate citations to prior collaborators' publications relative to the control group



Note: This figure represents a plot of coefficients stemming from a Poisson model where prior collaborator's aggregate citations are regressed onto year and case effects, an exposure term capturing publication output in the year, and 7 interaction terms which represent the treatment effect for each year before or after the *Findings* were published. Coefficients are exponentiated and represent Incidence Rate Ratios. Confidence bands represent 95% confidence intervals, calculated based on robust standard errors clustered by misconduct case.

6.2.Descriptive statistics

Table 2 provides summary statistics and correlations. The average prior collaborator of a grant recipient in the sample publishes 4.39 papers per year, which received 282.23 aggregate citations. The average prior collaborator enters the sample relatively late in her career, 16.94 years after their first publication. It should be underscored here that the sample under study does not form a representative sample of biomedical researchers, but rather of researchers that collaborate with recipients of prestigious NIH grants.

Table 3 compares the means between the treatment and control groups for the key variables. Treated prior collaborators publish slightly more than those in the control group (treated: 4.75, control 4.11 difference significant at $p<0.01$), but the control scientists are cited equally often (treatment: 274 aggregate citations per year, control: 288, difference not significant at $p<0.1$). In terms of control variables, treated authors are approximately half a year younger than control authors (treatment: 16.62 years since first publication, control: 17.18, significant at $p<0.01$) and tend to have lower measures of pre-sample citations average per publication (treatment: 41.92, control: 50.78, significant at $p<0.01$), which are defined as the average aggregate citations per publication the author received in each observed year before t-3.

Table 2: Summary statistics (2005 prior collaborators, 13206 observations)

	Summary Statistics					Correlation Matrix					
	Mean	St. Dev	Min	Median	Max	1	2	3	4	5	6
1. Publications	4.39	6.12	0	2	121	1					
2. Aggregate Citations	282.23	492.21	0	91	4257	0.69	1				
3. Years since first publication	16.94	9.43	1	16	56	0.22	0.17	1			
4. Pre-sample aggregate citations per publication average	46.91	62.17	0	30.71	1046.5	0.03	0.15	-0.13	1		
5. Treatment dummy	0.44	0.5	0	0	1	0.05	-0.01	-0.03	-0.07	1	
6. After treatment indicator	0.55	0.5	0	1	1	0.02	0.01	0.18	-0.01	0.02	1

Note: 42 prior collaborators that did not publish between t-3 and t+3 have been dropped from the sample used for the descriptive statistics. Aggregate citations: total number of citations received by papers published in year *t* at time of data extraction.

Table 3: Mean comparison treated and control samples

	Treated		**Control**		**Diff**
	Mean	St. Dev	Mean	St. Dev	
Publications	4.75	7.08	4.11	5.23	***
Aggregate citations	274.31	473.09	288.37	506.48	.
Years since first publication	16.62	9.48	17.18	9.48	***
Pre-sample aggregate citations per publication average	41.92	65.17	50.78	59.46	***
After treatment indicator	0.56	0.50	.054	0.50	**
Authors	856		1149		
Observations	5769		7437		

Notes: Sample restricted to author-years with at least one publication to match baseline estimation sample. Diff: two-sample t-test. Stars indicate significance level of difference: *: $p<0.10$, **: $p < 0.05$, ***: $p< 0.01$

6.3. Estimation results

Table 4 presents the results of Poisson regression of prior collaborator's aggregate citations on the treatment effects and controls. In all specifications, we find that the aggregate citations per publication by treated prior collaborators drops compared to prior collaborators in the control group in the years after the *Findings* were published. The point estimates of all specifications are significant at least $p < 0.05$ and show an average drop between 8% and 9% in the main specifications and up to 12% in the alternate specifications.

The first column shows the model without taking into account any observed or unobserved scientist-specific controls, only controlling for case and common year effects. The former is justified by the idea that cases of misconduct bring about highly heterogeneous amounts of (media) attention in the scientific community and severity. Therefore, the effect on the careers of prior collaborators is also likely to be partly case-specific. The results show that treated prior collaborators' aggregate citations do not differ significantly from control collaborators. We also find positive trend for both groups, which is captured by the dummy indicating the after-treatment. The treatment effect of being a treated prior collaborator (as inferred from the interaction between being in the treated group and the after-treatment indicator) is negative and highly significant, with a marginal effect of 11.3% at the mean.[11] In other words, prior collaborators suffer a loss of 11.3% in aggregate citations after a scandal is published.

Model 1, however, does not control for publication volume. Therefore, we offset citation counts by publication counts in model 2.[12] This controls for the volume of scientific output and

[11] Calculated as the difference in expected incidence rate ratio at the mean, i.e. 1-exp(-0.12).
[12] In practice, we include the natural logarithm of the number of publications in year *t* as an explanatory variable with its coefficient fixed at one. This thus means that we have to condition on the subsample with at least one publication per year, which is the case for 80% of the sample.

allows us to interpret the coefficients as a change in citation rate. Offsetting changes the interpretation of the results: whereas treated prior collaborators gathered weakly significantly more citations than control prior collaborators, they gather fewer per publication. Offsetting also lowers the treatment effect: the drop of 11.3% of citations at the mean translates to a drop of 7.7% in the citation to publication rate.

In model 3 we enrich the offset baseline model by controlling for the career age of the scientist through the time since first publication in linear and quadratic form. We additionally include a proxy for unobserved ability in the form of the pre-sample average aggregate citations per publications. We find an inverted-U shaped relation between career age and citation counts, and a positive significant relationship between pre-sample citations per publications and current citation outputs. Including these factors slightly increases the treatment effect from 7.7% to 8.6%.

The results in columns 2 and 3 are based on the assumption that there is a one-to-one relationship between citations and publications. That is, the model assumes that the coefficient of the natural logarithm of publications is one. In model four, we relax this assumption. We find that the estimated treatment effect is the same as before, and the coefficient of the publication offset is highly significant ($p<0.01$) and estimated at 1.02.

Figure 1 showed an initial drop in aggregate citations among treatment prior collaborators, as compared to prior collaborators in the control group, at t_{-1}. We speculate that this can be a result of inaccuracy of the treatment effect: while we are certain that the misconduct has been published with the publication of the *Findings*, the misconduct could also have been made public through other channels before then. We test whether this concern explains our findings through model 5, where we move the treatment one year forward, to t_{-1}. While the results are qualitatively the same, the treatment effect is estimated slightly higher at 11.3% in this specification. As an alternative robustness test, we provide results when disregarding years t_{-1} and t_0 in column 6. We thus compare the aggregate citations three and two years before the findings of misconduct to one to three years after. This shows an even higher estimate of the treatment effect at 12.2%.

Lastly, we present in column 7 the results of fixed effects Poisson estimation with robust standard errors as described in Wooldridge (Wooldridge, 1999). Using fixed effects instead of pre-sample estimators does not yield different conclusions. As in model 2 and 3, the treatment effect is estimated at 8.6%.

Table 4: Poisson regression estimates of aggregate citations of prior collaborators of misconducting authors and control group

Dependent: Prior collaborator's Aggregate citations	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model	Baseline	Offset by Publications	Incl. controls	Relax offset	Lead treatment	Drop t-1 and t0	QML F.E.
Treatment group	0.13	-0.09*	-0.03	-0.04	0.005	0.01	
	(0.08)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
After-treatment period	0.08*	0.08**	0.10***	0.10***	0.04	0.02	0.08**
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.08)	(0.04)
Treatment group* After-treatment period	-0.12**	-0.08**	-0.09**	-0.09**	-0.12***	-0.13***	-0.09**
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Years since first publication			0.02***	0.02***	0.02***	0.02***	
			(0.01)	(0.01)	(0.01)	(0.01)	
Years since first publication^2/100			-0.03***	-0.03**	-0.03***	-0.04***	
			(0.01)	(0.01)	(0.01)	(0.01)	
Pre-sample aggregate citations per publication average/10			0.03***	0.03***	0.03***	0.03***	
			(0.003)	(0.003)	(0.003)	(0.003)	
Ln(number of publications)				1.02***			0.91***
				(0.02)			(0.02)
Intercept	4.761***	3.95***	3.71***	3.70***	3.70***	3.75***	
	(9.96)	(0.37)	(0.37)	(0.37)	(0.37)	(0.39)	
Year effects	YES	YES	YES	YES	YES	YES	YES
Case effects	YES	YES	YES	YES	YES	YES	NO
Individual fixed effect	NO	NO	NO	NO	NO	NO	YES
Number of observations	13206	10756	10756	10756	10756	7526	10679

Notes: Poisson estimation of aggregate citations received by articles published in year. Models 2,3,5,6: Estimates offset by publications in year by including ln(number of publications) with coefficient fixed at 1. Model 4 and 7: ln(number of publications) included as control without restriction. Prior author-years without publications not included in columns 2-7. Years since first publication^2 and pre-sample aggregate citations per publication scaled down for readability. Cluster-robust standard errors in parentheses. Stars indicate significance level of coefficient: *: $p<0.10$, **: $p < 0.05$, ***: $p< 0.01$

6.4. Further analysis and robustness checks

6.4.1. Effect on publication output

Table 5 presents estimation results for Poisson regression models using the volume of scientific output as the dependent variable. The models presented in columns 1-3 do not show a statistically or economically significant drop in the number of papers published by the prior collaborators after the *Findings* were revealed. This strengthens the interpretation of the observed drop in citations being the result of stigmatization – lower perceived quality or reliability of published science, and not of effort exercised by the authors, in which case a drop in publications should also be observed.

Table 5: Poisson regression estimates of publications by prior collaborators of misconducting authors

Dependent: Prior collaborator's publication count	(1)	(2)	(3)
Model	Baseline	Incl. Controls	QML FE
Treatment group	0.27***	0.18***	
	(0.07)	(0.05)	
After-treatment period	-0.02	0.003	-0.0005
	(0.03)	(0.03)	(0.02)
Treatment group* After-treatment period	-0.03	-0.02	-0.03
	(0.03)	(0.03)	(0.03)
Years since first publication		0.07***	
		(0.01)	
Years since first publication^2/100		-0.15***	
		(0.02)	
Pre-sample publication average		0.15***	
		(0.01)	
Intercept	0.80***	0.07	
	(0.14)	(0.18)	
Year effects	YES	YES	YES
Case effects	YES	YES	NO
Individual fixed effect	NO	NO	YES
Number of observations	13206	13206	13204

Notes: Poisson estimation of yearly publication counts of prior collaborators. Years since first publication^2: coefficient scaled down for readability. Cluster-robust standard errors in parentheses. Stars indicate significance level of coefficient: *: p<0.10, **: p < 0.05, ***: p< 0.01

6.4.2. Effect on misconducting authors

Table 6 shows the effects on misconducting authors themselves. The results (columns 1-3) show a negative and weakly significant treatment effect in the range of 26 to 33%. Previous studies focusing on retractions which include scientific misconduct cases along with retractions due to honest mistakes identified a 'career effect' on the order of 10% (Azoulay et al., 2015). In comparison, we are analysing a dataset of investigated and declared misconduct cases of highly eminent researchers. Note also that the statistical significance of these coefficients is hindered by the much smaller estimation sample.

Columns 4 to 6 show estimates of publication output by misconducting authors. In line with previous findings (Redman and Merz, 2008), authors found guilty of scientific misconduct by the ORI have a high chance of dropping out of publishing: 21% of authors do not publish in the three years after treatment, compared to 5% in the three years before. This is a stark increase compared to the control group, which only shows 3% non-publishing authors in the after-treatment period. This pattern is mimicked in the regression estimates, which report drops in publication output by misconducting authors between 36 and 44% (columns 4-6). However, this effect is only statistically significant in the fixed effects specification where we control for author-specific unobserved effects, indicating a strong degree of remaining unobserved heterogeneity.

Table 6: Poisson regression estimates of citations rates and publication output of misconducting authors

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent	Misconducting author's aggregate citations			Misconducting author's publications		
Model	Baseline	Incl. controls	QML FE	Baseline	Incl. controls	QML FE
Treatment group	-0.02 (0.19)	-0.01 (0.26)		-0.61** (0.25)	-0.47* (0.28)	
After-treatment period	-0.22** (0.11)	-0.04 (0.12)	-0.08 (0.13)	0.24 (0.18)	0.11 (0.11)	-0.11 (0.11)
Treatment group* After-treatment period	-0.65*** (0.23)	-0.49** (0.25)	-0.30 (0.19)	-0.47 (0.33)	-0.44 (0.32)	-0.59** (0.27)
Years since first publication		-0.02 (0.04)			-0.02 (0.04)	
Years since first publication^2/100		0.03 (0.06)			0.02 (0.08)	
Pre-sample aggregate citations per publication average/10		0.12*** (0.04)				
Pre-sample publications average					0.27*** (0.04)	
Ln(number of publications)			0.96*** (0.10)			
Intercept	3.59*** (0.07)	3.63*** (0.54)		2.29*** (0.64)	0.73 (0.53)	
Year effects	YES	YES	YES	YES	YES	YES
Case effects	YES	YES	NO	YES	YES	NO
Individual fixed effect	NO	NO	YES	NO	NO	YES
Number of observations	255	255	253	331	331	331

Notes: Poisson estimation of yearly aggregate citation counts of misconducting authors (columns 1-3) and annual citation output (columns 4-6). Aggregate citation count estimates offset by publications in year by including ln(number of publications) with coefficient fixed at 1 (QML F.E: not fixed). Prior author-years without publications not included in citation estimates sample. Years since first publication² and pre-sample aggregate citations per publication average scaled down for readability. Cluster-robust standard errors in parentheses. Stars indicate significance level of coefficient: *: p<0.10, **: p < 0.05, ***: p< 0.01

6.4.3. Robustness Checks

6.4.3.1. Robustness Check I: Exiting misconducting authors as a driver of the citation drop

As a first robustness check, we explore the possibility that part of the drop in aggregate citations might be caused by the fact that misconducting authors leave the scientific community (Redman and Merz, 2008). The results presented in the previous section indicated that misconducting authors publish significantly less after the misconduct case surfaces. Since the likelihood to cite a prior collaborator might be disproportionately high, the drop of misconducting authors can be an alternative explanation for the patterns observed.

Whereas our dataset does not allow us to assess the origin of every individual citation, we perform a back-of-the-envelope calculation as a robustness check. We first estimate a publication trend for each misconducting author, based on pre-treatment observations, as a proxy of the counterfactual publication pattern had the author not been caught committing scientific misconduct. We then make the most conservative assumption that all of these publications (observed in the pre-treatment and estimated counterfactual situation in the post-treatment period) cite all prior publications of the collaborators. We then subtract this maximum amount of potential citations that the collaborator could have received from the misconducting scientist from the dependent variable and re-run our regressions.[13] If the citation drop would be caused by the misconducting author leaving academia our previously found treatment effect should disappear.

The results, shown in column 1 of table 5, however, display that the treatment effect is still significant and negative. This exercise should be seen as an upper limit estimate of the potential citation relation between misconducting author and prior collaborator, it indicates that the effect is not driven by the misconducting author dropping out of scientific publishing.

[13] Due to the fact that we use estimated citations for the counterfactual situation it can happen that the aggregate citations minus the estimated value turn negative. In this case we set them to 0.

Table 7: Robustness checks: excluding citations by misconducting authors

Dependent: Prior collaborator's aggregate citations	(1)	(2)	(3)
Model	Baseline	Incl. Controls	QML FE
Treatment group	0.33***	0.39***	
	(0.07)	(0.07)	
After-treatment period	0.24***	0.25***	0.23***
	(0.05)	(0.05)	(0.05)
Treatment group*	-0.24***	-0.25***	-0.27***
After-treatment period	(0.05)	(0.05)	(0.05)
Years since first publication		0.02**	
		(0.01)	
Years since first publication^2/100		-0.03*	
		(0.02)	
Pre-sample citations per publication average/10		0.03***	
		(0.004)	
Intercept	3.14***	2.88***	0.88***
	(0.60)	(0.61)	(0.03)
Year effects	YES	YES	YES
Case effects	YES	YES	NO
Individual fixed effect	NO	NO	YES
Number of observations	10756	10756	10144

Notes: Poisson estimation of yearly aggregate citation counts, excluding citations by misconducting authors. Citations by misconducting authors have been estimated post-treatment by extrapolating pre-treatment trend and assuming that every publication by misconducting author cites entire body of prior collaborator's work at that time. Aggregate citation count estimates offset by publications in year by including ln(number of publications) with coefficient fixed at 1. Prior collaborator-years without publications not included in sample. Years since first publication^2 and pre-sample citations per publication scaled down for readability. Cluster-robust standard errors in parentheses. Stars indicate significance level of coefficient: *: p<0.10, **: p < 0.05, ***: p< 0.01

6.4.3.2. Robustness Check II: Citation flows

We acknowledge the fact that a drop of aggregate citations does not necessarily coincide with a drop in the citation flow. Hence, we replicate the regressions for aggregate citations reported above for the citation flow in Table 8. We take a three-year fixed overlapping citation window (Rehn et al., 2014, p. 19) and offset for the volume of publications that might be cited through the number of articles published in the same time window. As Scopus only offers detailed information on citation flows from 1996 onwards (Rehn et al., 2014 p. 16) we restrict the estimation sample to this period. We do not include a pre-sample mean estimator in these models as it imposes a particularly strong constraint on the sample, and instead focus on the QML F.E. Poisson model to control for unobserved heterogeneity. While the models without controls or offset (Table 8, column 1), with offset but without controls (column 2), or with offset and controls (column 3) show little effect, the results from a QML F.E. Poisson regression shows an estimated drop of 9% in citation flow after publication of the Findings (column 4), which is statistically significant at $p<0.05$. The difference between the QML F.E. specification and the others highlights the importance of controlling for unobserved individual-specific effects in this specification.

Table 8: Robustness checks: Citation flow

Dependent: Prior collaborator's citation flow	(1)	(2)	(3)	(4)
Model	Baseline	Offset	Controls	QML F.E.
Treatment group	0.09 (0.10)	-0.06 (0.07)	-0.06 (0.07)	
After-treatment period	-0.04 (0.05)	0.03 (0.03)	0.03 (0.03)	0.04 (0.03)
Treatment group* After-treatment period	0.02 (0.07)	-0.02 (0.05)	-0.01 (0.05)	-0.09** (0.04)
Years since first publication			0.01 (0.01)	
Years since first publication^2/100			-0.02 (0.02)	
Ln(Publication stock)				1.00*** (0.04)
Intercept	3.74*** (0.59)	1.24*** (0.31)	1.12*** (0.32)	
Year effects	YES	YES	YES	YES
Case effects	YES	YES	YES	NO
Individual fixed effect	NO	NO	NO	YES
Number of observations	10235	9731	9731	9479

Notes: Poisson estimate of prior collaborator's yearly citation flow. Citation flow calculated in a three year citation window. Publication Stock: number of articles published in same window. Model 2-3: estimates offset by ln(Publication Stock) with coefficient fixed at 1. Model 4: ln(Publication Stock) included as control without restriction. Model 2-4: author-years with 0 publication stock not included. Years since first publication^2 scaled down for readability. Cluster-robust standard errors in parentheses. Stars indicate significance level of coefficient: *: $p<0.10$, **: $p < 0.05$, ***: $p< 0.01$

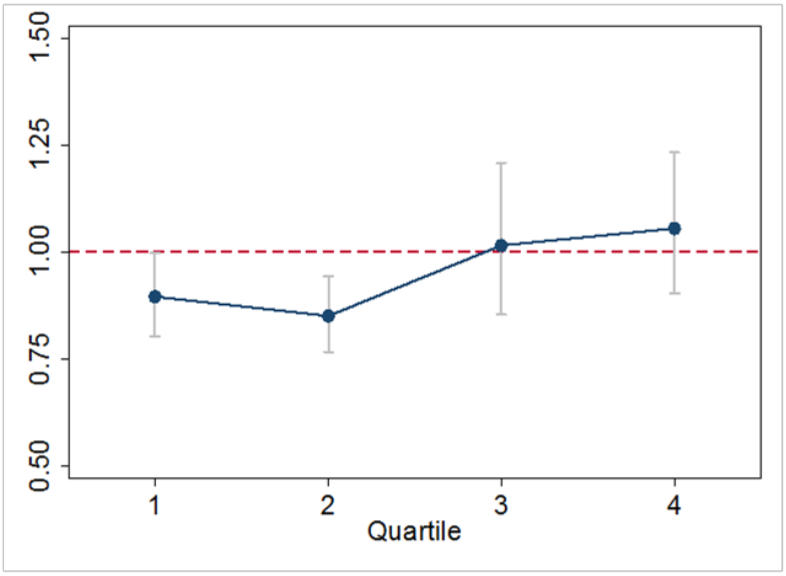
6.4.3.3. Effect heterogeneity: author reputation and scandal size

Finally, we explore two sources of potential heterogeneity in the effects documented above. We consider the standing of the prior collaborators before the misconduct case surfaces and the scope of the scandal. One could expect that scientists with a better standing are more affected by stigmatization and that also a scandal with a broader scope would have more significant negative effects. On the other hand, previous work has shown that the most eminent authors seem to be able to mostly avoid the fallout associated with retractions, whereas less eminent author suffer disproportionally (Jin et al., 2013)

Starting with the standing of scientists, we estimated differential effects by the citation stock of either the misconducting author or the prior collaborator, along their location in the citation stock distribution of the sample in the year before treatment. Citations proxy the standing or the visibility of the scientist within the academic community. The results are shown in Figure 2, for heterogeneity along the misconducting author's pre-treatment citation stock, and in Figure 3 for that of prior collaborators. The regression output can be found in table A.1. As Figure 2 shows, the negative effects shown above are driven by the bottom half of the misconducting authors' citation distribution. For the third and fourth quartile, the point estimates turn positive but statistically nonsignificant at $p < 0.05$. In other words, stigmatization of prior collaborators seems to occur mainly when the misconducting author has (comparatively) low reputation himself.

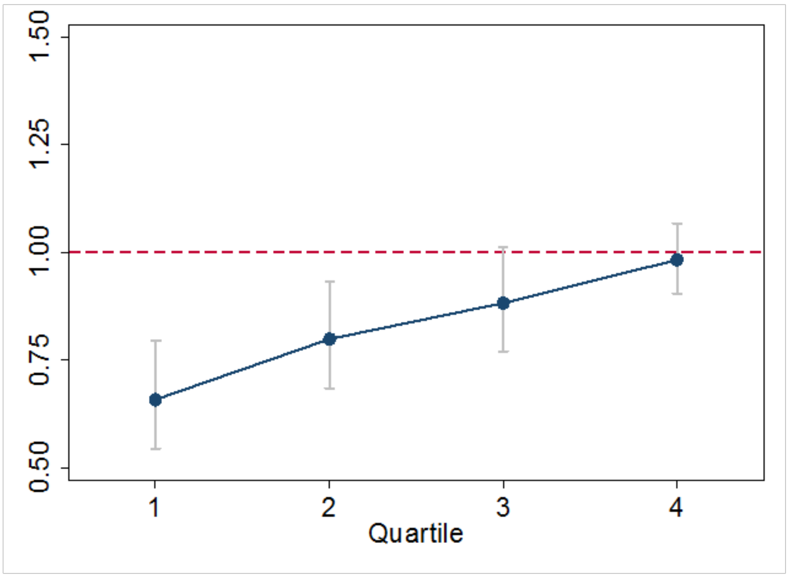
Figure 3 plots the results when differentiating by the prior citation stock of the prior collaborator herself. Here too, the results show that the effects are strongest for authors in the lower quartile, who experience a 34% drop in citations ($p<0.01$). The second quartile is estimated at a drop of 20% ($p<0.01$). For the third quartile the effect is estimated at 11% ($p<0.10$), and the fourth quartile shows no significant drop at $p<0.10$. These results confirm the "reverse Matthew effect" documented previously by Jin et al. (2013) also for prior collaborators. The main effects thus seem to be strongly moderated by, on the one hand, the prior reputation of the misconducting author, and, on the other, the prior reputation of the potential stigmatization victim. This is in line with the interpretation of the results presented here as social stigmatization: especially those prior collaborators who lack a well-established reputation suffer from the stereotyping inherent to stigmatization by mere association. The effects are at their worst when both are on the left tail of the distribution.

Figure 2: Treatment effect on prior collaborator by pre-treatment citation quartile of misconducting author



Note: This figure represents a plot of coefficients stemming from a Poisson model where prior collaborator's aggregate citations are regressed onto year and case effects, years since first publication in linear and squared terms, pre-sample aggregate citations per publication average, and an exposure term capturing publication output in the year. The plot shows the treatment effect interacted with 4 quartile indicators of misconducting author's pre-treatment citation stock. Coefficients are exponentiated and represent Incidence Rate Ratios. Confidence bands represent 95% confidence intervals, calculated based on robust standard errors clustered by misconduct case. Full regression output is presented in table A.1

Figure 3: Treatment effect on prior collaborator by pre-treatment citation quartile of prior collaborator



Notes: This figure represents a plot of coefficient estimates stemming from Poisson model where prior collaborator's aggregate citations is regressed onto year and case effects, years since first publication in linear and squared terms, pre-sample aggregate citations per publication average , and an exposure term capturing publication output in the year. The plot shows the treatment effect interacted with 4 quartile indicators of prior collaborator's pre-treatment citation stock. Coefficients are exponentiated and represent Incidence Rate Ratios. Confidence bands represent 95% confidence intervals, calculated based on robust standard errors clustered by misconduct case. Full regression output is presented in table A.1

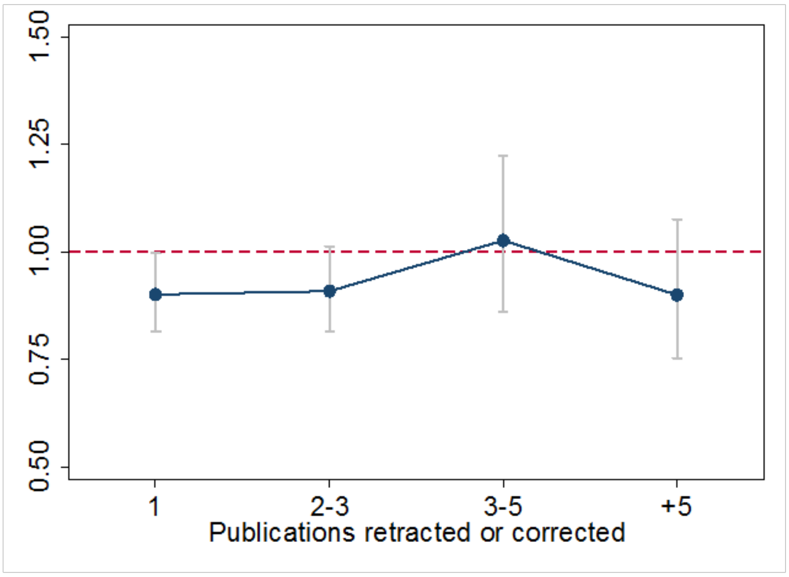
In a second set of estimations we compare the effects across the scope of the scandal. Ceteris paribus, it is reasonable to expect a stronger impact on reputation of a larger misconduct scandal. To conceptualize the scope of a scandal, we make use of the number of papers corrected or retracted, the number of citations received by papers which were corrected in the wake of the scandal, and the highest impact factor associated with a paper which was corrected or retracted in the wake of the scandal. One to ten papers were affected by the misconduct cases included in the study.[14]

Figures 4 through 6 show the results of estimations differentiating between different scopes of the scandal. The full results are presented in table A.2. Along all three measures, we cannot detect a substantial trend in effect size along the scope of the scandal. This result might

[14] The median case involved the retraction or correction of 2 articles. Approximately 75% of the articles affected were retracted, the others corrected. The results presented in this section are similar when differentiating between retracted and corrected articles.

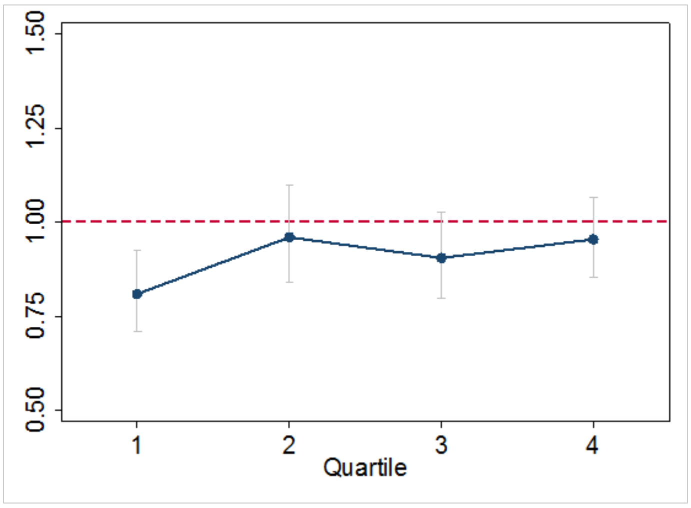
be explained by the fact that our analysis is based on a positive selection in terms of scandal scope since all our cases have been formally investigated and found guilty by the ORI, which in all but a few cases in the sample led to exclusion from highly prestigious NIH grant funding. There might be a stronger link between scandal scope and consequences (for prior collaborators) among the broader range of misconduct cases outside of the ORI.

Figure 4: Treatment effect on prior collaborator by size of scandal number of articles affected



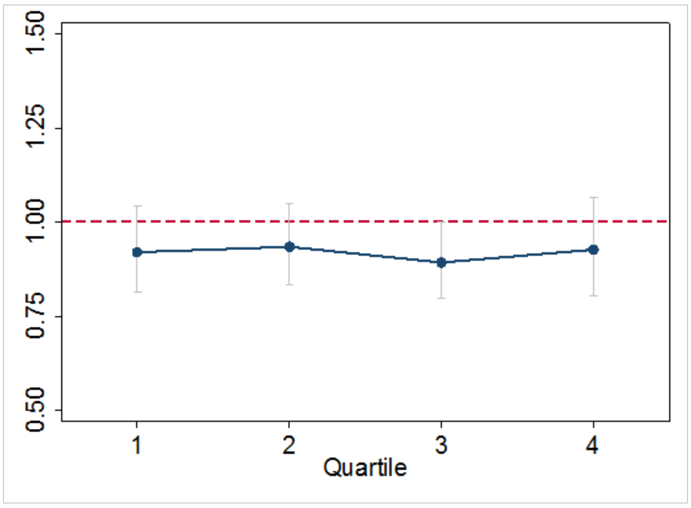
Note: This figure represents a plot of coefficients stemming from a Poisson model where prior collaborator's aggregate citations are regressed onto year and case effects, years since first publication in linear and squared terms, pre-sample aggregate citations per publication average, and an exposure term capturing publication output in the year. The plot shows the treatment effect interacted with 4 groups of the number of articles retracted or corrected in the wake of the scandal. Coefficients are exponentiated and represent Incidence Rate Ratios. Confidence bands represent 95% confidence intervals, calculated based on robust standard errors clustered by misconduct case. Full regression output is presented in table A.2

Figure 5: Treatment effect on prior collaborator by size of scandal: total citation impact



Note: This figure represents a plot of coefficients stemming from a Poisson model where prior collaborator's aggregate citations are regressed onto year and case effects, years since first publication in linear and squared terms, pre-sample aggregate citations per publication average, and an exposure term capturing publication output in the year. The plot shows the treatment effect interacted with 4 quartile indicators of the importance of the affected work, as measured through the total number of citations received by the work. Coefficients are exponentiated and represent Incidence Rate Ratios. Confidence bands represent 95% confidence intervals, calculated based on robust standard errors clustered by misconduct case. Full regression output is presented in table A.2

Figure 6: Treatment effect on prior collaborator by size of scandal: highest impact factor



Note: This figure represents a plot of coefficient estimates stemming from a Poisson model where prior collaborator's aggregate citations are regressed onto year and case effects, years since first publication in linear and squared terms, pre-sample aggregate citations per publication average, and an exposure term capturing publication output in the year. The plot shows the treatment effect interacted with 4 quartile indicators of importance of the affected work, as measured through the highest impact factor among the articles retracted or corrected in the wake of the misconduct scandal. Coefficients are exponentiated and represent Incidence Rate Ratios. Confidence bands represent 95% confidence intervals, calculated based on robust standard errors clustered by misconduct case. Full regression output is presented in table A.2

7. Conclusions and discussion

The fact that we can detect the impact of misconduct in prior collaborators suggests that scientific misconduct affects many more than previously imagined. By documenting negative consequences of scientific misconduct for innocent prior collaborators, our results show that the implications of scientific misconduct spread beyond the fraudulent researcher, suggesting that articles by prior collaborators are less often used as building blocks for future research. This is an unwarranted and wasteful disregard of valid research findings, which slows down scientific progress and represents a cost to society. These findings are consistent with the theory of stigmatization through mere association (Goffman, 1963; Goldstein and Johnson, 1997; Mehta and Farina, 1988; Pontikes et al., 2010).

Why would stigmatization through mere association happen in the scientific setting? The answer, in our opinion, lies in the reliance on trust in the scientific enterprise. While misconduct should in principle be detected during the peer review process or through scrutiny by readers, it is prohibitively costly – in terms of money and time - to verify some results to a sufficient degree, and scientists need to a certain extent assume honesty of their peers (Dewald et al., 1986). At the same time, scientific research is becoming ever more complex and multidisciplinary so that scientists need to rely more on the expertise of their collaborators (Jones, 2009; Wuchty et al., 2007). Not in the least, researchers do not have strong incentives to perform replication studies, as they receive relatively little attention and are unlikely to change thinking (Hamermesh, 2007). Yet these studies are important. One recent replication study reported that of 100 replication attempts in psychology, only 39 were successful (Baker, 2015). Another study found that 47 out of 53 'landmark' studies in cancer research could not be replicated (Begley and Ellis, 2012). Given that scientists need to assume the honesty of their colleagues and are hard pressed for testing it, it should come as no surprise that they fall back on heuristics and stereotyping for appraising the reliability of their colleagues.

Whereas some evidence exists that those accused but not found guilty of misconduct can at least mend their reputation through exoneration (Greitemeyer and Sagioglou, 2015), there seem to be no clear course of action for dealing with the indirect reputational effects documented here. The problem becomes even clearer when one notes that in this setup, misconducting scientists were clearly named, and the nature of their misdoings was transparently reported. Other, less transparent contexts might lead to even worse effects for prior collaborators.

Since means to counteract stigma at the institutional level are limited, we turn to options that the stigmatized collaborator themselves might have. Psychiatric research discusses three ways of responding to stigma:[15] protest, education, and contact (Corrigan and Watson, 2002). Evidence suggests those who know more about the "disease" – in our context the association of misconduct – tend to stigmatize less than others (Corrigan and Penn, 1999). Contact with stigmatized persons also reduces stigmatization (Kolodziej and Johnson, 1996). In the present context, this means that prior collaborators of colleagues found guilty of scientific misconduct should actively take part in the scientific community. Attempting to suppress the stigma can however also worsen stigmatization (Monteith et al., 1998).

Given these findings, we believe that it is important to create awareness within the scientific community about guilt by association. Stigmatized scientists may for instance proactively use scientific conferences as a forum to discuss their case with the scientific community. This is in line with the psychiatric recommendation to "transform the person from a patient to an advocate" (e.g. Byrne, 2000). Such efforts could also be supported by scientific journals by providing a forum for innocent scientists with indirect ties to misconducting colleagues. At least in case of stigma related to mental disease, education campaigns have been proven to be successful (Wolff et al., 1996).

[15] This research is mostly concerned with stigmatization associated with mental disease.

One alarming implication of our findings is that scientists might refrain from whistleblowing on their colleagues if they know to expect a backlash through association. Such behaviour would stand in sharp contrast to the self-correction function that science relies on. A survey or 68 whistle-blowers at the ORI (Lubalin and Matheson, 1999) supports that a majority of whistle-blowers experiences negative outcomes of their whistleblowing, affecting their career advancement, professional activities, mental health and personal life. The question arises whether self-correction suffices as a mechanism to ensure validity of research results; or whether in times of increased pressure on scientists to publish paired with unpreceded levels of complexity and multidisciplinarity of research call for different quality control mechanisms. Some have proposed that stigmatization might serve a function as a deterring factor for future misconduct. However, the effectiveness of stigma for deterrence is debated in the broader criminal setting (Funk, 2004; Harel and Klement, 2007; Rasmusen, 1996), with some evidence indicating that, while inducing deterrence effects for non-offenders, stigmatization increases the crime rate for offenders (Funk, 2004). In our setting, stigmatization might make future misconduct more attractive by reducing its costs for the already stigmatized innocent collaborators. In this sense, modelling the effect of stigmatization and stigma by association and scientific misconduct, might make for an interesting avenue for future theoretical research.

Our study makes several contributions to the small, but developing literature on the consequences of scientific misconduct. First of all, we contribute to prior studies providing possible explanations for citation penalties for scientists around the misconducting scientist himself. Mechanisms proposed previously include the Bayesian updating of beliefs (Azoulay et al., 2014a; Jin et al. 2013), a loss of trust (Lu et al, 2012) and the ambition to protect the own scientific integrity (Azoulay et al., 2014a). We introduce the concept of social stigmatization which we borrowed from sociology as an overarching concept to explain the three mechanisms. Caused by a *loss of trust*, social stigmatization can be the reason for an *Bayesian updating of* *beliefs* about the "quality" of prior collaborators and the way for the members of the scientific community too *protect the own scientific integrity*.

Second, we show that the ripple effect of scientific misconduct reaches further than the fraudulent author's close collaborators or colleagues. Given the significance of the citation penalty that these distantly related scientists experience and the increasing trend of collaboration in science, our results show that the costs of scientific misconduct have been underestimated so far. This seems to be especially the case for scholars who have not developed a strong reputation for themselves yet, as these seem to be hit particularly hard by stigmatization through mere association.

Third, we present the first study to analyse a clean sample of scientific misconduct cases in this context. Prior studies often focus on retraction which are readily available from publication databases such as the Web of Science or Scopus and include scientific misconduct cases along with retractions due to honest mistakes (Azoulay et al., 2012, 2104a). The difference in magnitude of the effects on misconduct authors (robustness check 2) that we present as compared to prior studies should remind us of the important difference between both.

From a more general perspective, our result is related to research about 'superstar extinction' in that it confirms that shocks to coauthorship networks seem to ripple out (Azoulay et al., 2014b, 2010). Our results can also be interpreted as an extreme case of negative citations, which have been shown to prelude a future decline in citations (Catalini et al., 2015).

Our study is not free of limitations. The downside of using a clean sample of misconduct cases for quantitative analysis is a smaller number of cases. In our context, where we are interested in the effects of misconduct on prior collaborators this however is of minor importance because of the large number of collaborators NIH grant recipients have. A more important point to stress is that the NIH grant recipients are a positive selection and might not have a lot in common with the average scientists. This implies that our results represent upper limit estimates and should be treated as such. An interesting avenue for follow-up studies could be to employ network methods for a network of co-authors in order to investigate how far the effect of scientific misconduct reaches.

References

Aghion, P., Dewatripont, M., Stein, J., 2008. Academic freedom, private sector focus and the process of innovation. RAND J. Econ. 39, 617–635. doi:10.1111/j.1756-2171.2008.00031.x

Aksnes, D.W., Taxt, R.E., 2004. Peer reviews and bibliometric indicators: a comparative study at a Norwegian university. Res. Eval. 13, 33–41. doi:10.3152/147154404781776563

Azoulay, P., Bonatti, A., Krieger, J.L., 2015. The career effects of scandal: evidence from scientific retractions. NBER Work. Pap. No. 21146.

Azoulay, P., Fons-Rosen, C., Zivin, J.S.G., 2014b. Does Science Advance One Funeral at a Time ? Barcelona GSE Work. Pap. No. 857.

Azoulay, P., Furman, J.L., Krieger, J.L., Murray, F.E., 2014a. Retractions. Rev. Econ. Stat. doi:10.1162/REST_a_00469

Azoulay, P., Graff Zivin, J.S., Wang, J., 2010. Superstar Extinction. Q. J. Econ. 125, 549–589. doi:10.1162/qjec.2010.125.2.549

Baker, M., 2015. Over half of psychology studies fail reproducibility test. Nature. doi:10.1038/nature.2015.18248

Banerjee, A. V, 1992. a Simple Model of Herd Behavior. Q. J. Econ. 107, 797–817. doi:10.2307/2118364

Barnett, M.L., Hoffman, A.J., 2008. Beyond Corporate Reputation: Managing Reputational Interdependence. Corp. Reput. Rev. 11. doi:10.1007/s10994-011-5260-9

Barnett, M.L., King, A.A., 2008. Good Fences Make Good Neighbors: a Longitudinal Analysis of an Industry Self-Regulatory Institution. Acad. Manag. J. 51, 1150–1170.

Begley, C.G., Ellis, L.M., 2012. Drug development: Raise standards for preclinical cancer research. Nature 483, 531–3. doi:10.1038/483531a

Blundell, R., Griffith, R., Van Reenen, J., 1999. Market share, market value and innovation in a panel of British manufacturing firms. Rev. Econ. Stud. 66, 529–554. doi:10.1111/1467-937X.00097

Byrne, P., 2000. Stigma of Mental Illness and ways of diminishing it. Adv. Psychiatr. Treat. 6, 65–72.

Catalini, C., Lacetera, N., Oettl, A., 2015. The incidence and role of negative citations in science. Proc. Natl. Acad. Sci. 112, 13823–13826. doi:10.1073/pnas.1502280112

Cole, S., Cole, J.R., 1967. Scientific Output and Recognition : A Study in the Operation of the Reward System in Science. Am. Sociol. Rev. 32, 377–390.

Corrigan, P.W., Penn, D.L., 1999. Lessons from social psychology on discrediting psychiatric stigma. Am. Psychol. 54, 765–776. doi:10.1037/0003-066X.54.9.765

Corrigan, P.W., Watson, A.C., 2002. The paradox of self-stigma and mental illness. Clin. Psychol. Sci. Pract. doi:10.1093/clipsy/9.1.35

Cyranoski, D., 2015. Collateral Damage: how a case of misconduct brought a leading Japanese biology institute to its knees. Nature 520, 600–603.

Dasgupta, D., David, P., 1994. Towards a new economics of science. Res. Policy 23, 487–521.

Dewald, W.G., Thursby, J.G., Anderson, R.G., 1986. Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. Am. Econ. Rev. 76, 587–603. doi:10.2307/1806061

Dorsey, E.R., de Roulet, J., Thompson, J.P., Reminick, J.I., Thai, A., White-Stellato, Z., Beck, C.A., George, B.P., Moses, H., 2010. Funding of US biomedical research, 2003-2008. Jama 303, 137–43. doi:10.1001/jama.2009.1987

Fang, F.C., Steen, R.G., Casadevall, A., 2012. Misconduct accounts for the majority of retracted scientific publications. Proc. Natl. Acad. Sci. 109, 17028–17033. doi:10.1073/pnas.1220833110

Fleming, L., Sorenson, O., 2004. Science as a map in technological search. Strateg. Manag. J. 25, 909–928. doi:10.1002/smj.384

Fuchs, S., Westervelt, S.D., 1996. Fraud and trust in science. Perspect. Biol. Med. 39, 248–269.

Funk, P., 2004. On the Effective Use of Stigma as a Crime Deterrence. Eur. Econ. Rev. 48, 715–728.

Goffman, E., 1963. Stigma: Notes on a spoiled identity. Jenkins, JH & Carpenter.

Goldstein, S.B., Johnson, V. a., 1997. Stigma by Association: Perceptions of the Dating Partners of College Students With Physical Disabilities. Basic Appl. Soc. Psych. 19, 495–504. doi:10.1207/s15324834basp1904_6

Greitemeyer, T., Sagioglou, C. (2015). Does exonerating an accused researcher restore the researcher's credibility? PLoS ONE 10(5): e0126316. Doi.10.1371/journal.pone.0126316.

Groysberg, B., Lin, E., Serafeim, G., Abrahams, R., 2016. The Scandal Effect. Harv. Bus. Rev.

Hamermesh, D.S., 2007. Replication in Economics. Can. J. Econ. 40, 715–733.

Harel, A., Klement, A., 2007. The Economics of Stigma: Why More Detection of Crime May Result in Less Stigmatization. J. Legal Stud. 36, 355–377.

Hebl, M.R., Mannix, L.M., 2003. The weight of obesity in evaluating others: A mere proximity effect. Personal. Soc. Psychol. Bull. 29, 28–38. doi:10.1177/0146167202238369

Inhaber, H., Przednowek, K., 2007. Quality of Research and the Nobel Prizes. Soc. Stud. Sci. 6, 33–50. doi:10.1177/030631277600600102

Jin, G. Z., Jones, B., Lu, S. F., Uzzi, B. (2013). The Reverse Matthew Effect: Catastrophe and Consequence in Scientific Teams. NBER working paper no. 19489.

Jones, B.F., 2009. The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? Rev. Econ. Stud. 76, 283–317. doi:10.1111/j.1467-937X.2008.00531.x

Kiri, B., Lacetera, N., Zirulia, L., 2015. Above a Swamp: A Theory of High-Quality Scientific Production. NBER Work. Pap. 21143.

Kolodziej, M.E., Johnson, B.T., 1996. Interpersonal contact and acceptance of persons with psychiatric disorders: a research synthesis. J. Consult. Clin. Psychol. doi:10.1037/0022-006X.64.6.1387

Kostova, T., Zaheer, S., 1999. Organizational legitimacy under conditions of complexity: The case of the multinational enterprise. Acad. Manag. Rev. 24, 64–81. doi:10.5465/AMR.1999.1580441

Lacetera, N., Zirulia, L., 2011. The economics of scientific misconduct. J. Law, Econ. Organ. 27, 568–603. doi:10.1093/jleo/ewp031

Lefley, H., 1987. Impact of mental illness in families of mental health professionals. J. Nerv. Ment. Dis. 175, 613–619.

Levelt Committee, Noort Committee, Drenth Committee, 2012. Flawed science: Fraud. Res. Pract. Soc. Psychol. Diederik Stapel 1–104.

Levenson, J., Tewksbury, R., 2009. Collateral damage: family members of registered sex offenders. Am. J. Crim. Justice 34, 54–68.

Link, B.G., Phelan, J.C., 2001. Conceptualizing Stigma. Annu. Rev. Sociol. 27, 363–385.

Lu, S.F., Jin, G.Z., Uzzi, B., Jones, B., 2013. The retraction penalty: evidence from the Web of Science. Sci. Rep. 3, 3146. doi:10.1038/srep03146

Lubalin, J.S., Matheson, J., 1999. The Fallout: What Happens to Whistleblowers and Those Accused But Exonerated of Scientific Misconduct? Sci. Eng. Ethics 5, 229–250.

Mehta, S.I., Farina, A., 1988. Associative Stigma: Perceptions of the Difficulties of College-Aged Children of Stigmatized Fathers. J. Soc. Clin. Psychol. 7, 192–202. doi:10.1521/jscp.1988.7.2-3.192

Merton, R.K., 1973. The sociology of science. The University of Chicago Free Press, Chicago, IL.

Merton, R.K., 1968. The Matthew Effect in Science. Science (80-.). 159, 56–63.

Mokyr, J., 2002. The Gifts of Athena: Historical Origins of the Knowledge Economy. Princeton University Press, Princeton, NJ.

Mongeon, P., Larivière, V., 2015. Costly Collaborations : The Impact of Scientific Fraud on Co-authors' Careers. J. Assoc. Inf. Sci. Technol. 67, 535–542.

Monteith, M.J., Sherman, J.W., Devine, P.G., 1998. Suppression as a stereotype control strategy. Personal. Soc. Psychol. Rev. 2, 63–82. doi:10.1207/s15327957pspr0201_4

Murray, F., O'Mahony, S., 2007. Exploring the Foundations of Cumulative Innovation: Implications for Organization Science. Organ. Sci. 18, 1006–1021. doi:10.1287/orsc.1070.0325

Myers, C.R., 1970. Journal citations and scientific eminence in contemporary psychology. Am. Psychol. 25, 1041–1048. doi:10.1037/h0030149

Nalebuff, B.J., Stiglitz, J.E., 1983. Prizes and Incentives: Towards a General Theory of Compensation and Competition. Bell J. Econ. 14, 21–43. doi:10.2307/3003535

Pontikes, E.G., Negro, G., Rao, H., 2010. Stained red: A study of stigma by association to blacklisted artists during the "Red Scare" in Hollywood, 1945 to 1960. Am. Sociol. Rev. 75, 456–478. doi:10.1177/0003122410368929

Pozzi, A., David, P.A., 2007. Empirical realities of scientific misconduct in publicly funded tesearch: What can we learn from ORI investigations of U.S. cases in the biomedical and behavioral sciences? ESF-ORI first world Conf. Res. Integr. Foster. responsible Res. 1–88.

Rasmusen, E., 1996. Stigma and Self-Fulfilling Expectations of Criminality. J. Law Econ. 39, 519–543. doi:10.1086/467358

Redman, B.K., Merz, J.F., 2008. Scientific misconduct: do the punishments fit the crime? Science 321, 775. doi:10.1126/science.1158052

Resnik, D.B., Dinse, G.E., 2012. Scientific retractions and corrections related to misconduct findings. J. Med. Ethics 46–50. doi:10.1136/medethics-2012-100766

Reynolds, S.M., 2004. ORI findings of scientific misconduct in clinical trials and publicly funded research, 1992-2002. Clin. Trials 1, 509–516. doi:10.1191/1740774504cn048oa

Rhoades, L., 2004. ORI closed investigations into misconduct allegations involving research supported by the Public Health Service: 1994-2003. Investig. pdf 1–50.

Risen, J.L., Gilovich, T., Dunning, D., 2007. One-shot illusory correlations and stereotype formation. Personal. Soc. Psychol. Bull. 33, 1492–1502. doi:10.1177/0146167207305862

Smith, R.H., Corredoira, R., Goldfarb, B., Greenwood, B., Hallen, B., 2010. Managing the Message: The Effects of Firm Actions and Industry Spillovers on Media Coverage Following Wrongdoing. Acad. Manag. J. 55, 1079–1101.

Stephan, P., 2012. How Economics Shapes Science. Harvard University Press, Cambridge, MA.

Stern, S., 2004. Do Scientists pay to be scientists? Manage. Sci. 50, 835–853. doi:10.1287/mnsc.1040.0241

Swim, J.K., Ferguson, M.J., Hyers, L.L., 1999. Avoiding Stigma by Association: Subtle Prejudice Against Lesbians in the Form of Social Distancing. Basic Appl. Soc. Psych. 21, 61–68.

Van Noorden, R., 2011. The Trouble with Retractions. Nature 478, 26–28.

Wolff, G., Pathare, S., Craig, T., Leff, J., 1996. Public education for community care. A new approach. Br. J. Psychiatry 168, 441–447.

Wooldridge, J.M., 1999. Distribution-free estimation of some nonlinear panel data models. J. Econom. 90, 77–97. doi:10.1016/S0304-4076(98)00033-5

Wright, D.E., Titus, S.L., Cornelison, J.B., 2008. Mentoring and research misconduct: An analysis of research mentoring in closed ORI cases. Sci. Eng. Ethics 14, 323–336. doi:10.1007/s11948-008-9074-5

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science (80-.). 316, 1036–1040.

Appendix

Table A.1: Poisson regression estimates of citations rates of misconducting authors interacted with prior citation stock of misconducting author or prior collaborator

Dependent: Prior collaborators' aggregate citations	(1)	(2)
Model	By prior citation stock of misconducting authors	By prior citation stock of collaborator
Treatment group	-0.02 (0.05)	-0.03 (0.05)
After-treatment period	0.09** (0.04)	0.10*** (0.04)
Treatment group* After-treatment period* quartile 1	-0.11** (0.06)	-0.42*** (0.10)
Treatment group * After-treatment period* quartile 2	-0.16*** (0.05)	-0.22*** (0.08)
Treatment group * After-treatment period* quartile 3	0.02 (0.09)	-0.12* (0.07)
Treatment group * After-treatment period* quartile 4	0.05 (0.08)	-0.02 (0.04)
Years since first publication	0.02*** (0.01)	0.02*** (0.01)
Years since first publication^2/100	-0.03** (0.01)	-0.03** (0.01)
Pre-sample citations per publication/10	0.03*** (0.00)	0.03*** (0.00)
Intercept	3.71*** (0.37)	3.76*** (0.36)
Year effects	YES	YES
Case effects	YES	YES
Individual fixed effect	NO	NO
Number of observations	10756	10756

Notes: Poisson estimation of prior collaborator's aggregate citations. Aggregate citation count in year offset by publications in year by including ln(number of publications) with coefficient fixed at 1. Quartile indicates quartile of citation stock at t-1 of respectively misconducting author and prior collaborator. Author-years without publications not included in citation estimates sample. Years since first publication^2 and pre-sample citations per publication scaled down for readability. Cluster-robust standard errors in parentheses. Stars indicate significance level of coefficient: *: $p<0.10$, **: $p < 0.05$, ***: $p< 0.01$

Table A.2: Poisson regression estimates of citations rates of misconducting authors interacted with prior citation stock of misconducting author or prior collaborator

Dependent: Prior Collaborator's aggregate citations	(1)	(2)	(3)
Model	Number of publications affected	Citation-weighed affected publications	Highest IF of affected papers
Treatment group	-0.03	-0.03	-0.03
	(0.05)	(0.05)	(0.05)
After-treatment period	0.10***	0.10***	0.10***
	(0.04)	(0.04)	(0.04)
Treatment group*	-0.10**	-0.21***	-0.08
After-treatment period* quartile 1	(0.05)	(0.07)	(0.06)
Treatment group *	-0.10*	-0.04	-0.07
After-treatment period* quartile 2	(0.06)	(0.07)	(0.06)
Treatment group *	0.03	-0.10	-0.11**
After-treatment period* quartile 3	(0.09)	(0.06)	(0.06)
Treatment group *	-0.11	-0.05	-0.08
After-treatment period* quartile 4	(0.09)	(0.06)	(0.07)
Years since first publication	0.02***	0.02***	0.02***
	(0.01)	(0.01)	(0.01)
Years since first publication^2/100	-0.03***	-0.03**	-0.03**
	(0.01)	(0.01)	(0.01)
Pre-sample citations per publication	0.03***	0.03***	0.03***
average/10	(0.00)	(0.00)	(0.00)
Intercept	3.67***	3.69***	3.72***
	(0.35)	(0.36)	(0.37)
Year effects	YES	YES	YES
Case effects	YES	YES	YES
Individual fixed effect	NO	NO	NO
Number of observations	10756	10756	10756

Notes: Poisson estimation of prior collaborator's aggregate citations. Aggregate citation count in year offset by publication counts by including ln(number of publications) with coefficient fixed at one in the model. Column 1: quartile represents number of publications corrected or retracted in the misconduct case (1: 1 publication affected or corrected, 2: 2-3, 3: 3-5, 4: 5+). Column 2: quartile represents quartile of citation-weighted corrected or retracted publications in the misconduct case. Column 3: quartile represents quartile of highest impact factor among publications corrected or retracted in the context of the misconduct case. Author-years without publications not included. Years since first publication^2 and pre-sample citations per publication scaled down for readability. Cluster-robust standard errors in parentheses. Stars indicate significance level of coefficient: *: p<0.10, **: p < 0.05, ***: p< 0.01