

The Reverse Matthew Effect: Consequences of Retraction in Scientific Teams*

Ginger Zhe Jin, University of Maryland and NBER

Benjamin Jones, Northwestern University and NBER

Susan Feng Lu, Purdue University and CCER

Brian Uzzi, Northwestern University

Abstract

Teamwork pervades modern production, yet teamwork can make individual roles difficult to ascertain. The “Matthew Effect” suggests that communities reward eminent team members for great outcomes at the expense of less eminent team members. We study this phenomenon in

* We thank Alex Entz, Yiyang Liu, Huan Meng, and Ari Bellin for excellent research assistance and seminar participants at Case Western, Harvard, Northeastern, Northwestern, Purdue, University of Chicago, the 2013 International Industrial Organization Conference, and the Collegio Carlo Alberto for helpful comments and suggestions. We also thank the University of Maryland, the Northwestern University Institute on Complex Systems; the Army Research Laboratory under cooperative Agreement Number W911NF-09-2-0053; and Defense Advanced Research Projects Agency grant BAA-11-64, Social Media in Strategic Communication, for financial support. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. government. Contact emails: jin@econ.umd.edu; bjones@kellogg.northwestern.edu; lu428@purdue.edu; uzzi@kellogg.northwestern.edu.

reverse, investigating credit-sharing after damaging events. Our context is article retractions in the sciences. We find that retractions impose little citation penalty on the prior work of eminent coauthors, but less eminent coauthors experience substantial citation declines, especially when teamed with eminent authors. These findings suggest a “Reverse Matthew Effect” for team-produced negative events. A Bayesian model provides a candidate interpretation.

Keywords: teamwork, reputation, information, science, retraction, Matthew Effect.

JEL: J24, L15, L23, O3

1. Introduction

Teamwork is pervasive in modern production contexts, with benefits often related to the division of labor in executing tasks and/or creative advantages in driving innovation.¹ Yet team production raises challenges, including challenges in finding appropriate reward structures for team participants. Indeed, in many team production contexts, the joint output is observable but the separate inputs of individual team members are difficult to discern, which makes the assignment of credit difficult.² In situations where the output of the individual is not directly observed, reputation may become a cornerstone not only in providing effort incentives but also in shaping how outsiders assign credit within a team.

In a classic study, Robert K. Merton suggested the “Matthew Effect” as a fundamental issue in an important team production context, science (Merton 1968). Like many team production contexts, science is a setting where the joint output of the team is observable but the individual

¹ See, e.g., Bacon (1620) and Smith (1776) or modern analyses such as Becker and Murphy (1992), Woodman et al. (1993), and Jones (2009).

² See the large literature discussed in Section 2.

contributions of the team members are less clear. Merton argued that, in this setting, more eminent team members tend to limit the credit received by less eminent team members.³ In Merton's analysis, the community, upon witnessing a great contribution, assumes that the already eminent team member was the key producer while less well-known team member(s) were less important contributors who deserve less credit. However, empirical evidence on the foundational question of how credit is shared across team members remains limited.

Using scientific publications as an example, this paper considers the individual consequences of working in teams. Our question, however, concerns not the rewards of "good" events, but rather the consequences of "bad" events. Namely, we look at the effect of article retractions in team production settings and examine whether eminent team members attract or repel blame compared to less eminent team members. On the one hand, one might imagine that eminent individuals receive disproportionate credit for the joint output, whether good or bad, as the presumed leader of the enterprise. On the other hand, one may imagine that eminent individuals have such established reputations that they escape blame for bad events, leaving any blame to accrue to junior team members. Thus we may imagine a "Reverse Matthew Effect" through which less eminent team members experience greater, negative consequences.

³ Merton coined the Matthew Effect after the biblical passage "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath" (Matthew 25: 29, King James Version).

In our empirical analysis, we collect retracted articles in the Web of Science where the retracted paper was authored in a team and where the authors have a single retraction event.⁴ We then investigate citations to the prior publications of each author involved in the retracted work. To examine the effect of retraction, we match each of these prior publications (the “treated” papers) with a set of other publications (the “control” papers) that were published in the same field-year and received similar citations every year before the retraction event. This approach allows us to identify the effect of retraction via difference-in-differences estimation. This identification strategy builds from the observation that the content of prior work is unchanged, so that changes in citations to this work, compared to counterfactual control papers, reveal the effect of the retraction shock.⁵

Using standard measures of eminence, we find four central results following retraction events. First, less established team members experience substantial citation declines to their prior work. Second, by contrast, eminent team members experience little or no citation consequences for their prior work. Third, less established team members are especially negatively affected in the presence of an eminent team member. This interaction effect suggests that eminence may act not

⁴ That is, in our main analysis, we do not look at extreme cases where an author is revealed to be a systematic fraud as such cases make the credit assignment problem straightforward. We will, however, also consider the multiple retraction cases as a falsification exercise and show that, as expected when the guilty party is obvious, prior reputation no longer matters.

⁵ Using citations to prior scientific work to assess the effects of shocks was pioneered as an identification strategy in Furman and Stern (2011). While the retraction event is not a natural experiment, for ease of exposition, we will refer to the two groups as “treated” papers and “control” papers.

only to protect oneself, but also to hurt others on one's production team. Fourth, and related, we find that while the citation losses experienced by ordinary team members are exacerbated by the presence of eminent team members, these citation losses are attenuated in the presence of "rookies" – coauthors who had no prior work and are yet more junior to the ordinary coauthor. These results persist across a variety of robustness checks. These findings, where the already "rich" have an advantage over the relatively "poor" in the context of negative events, and where the effects on ordinary individuals depend on the standing of other team members, provide the paper's central results.

Given these findings, and building from reasoning in Merton's original Matthew Effect paper (Merton 1968), we further present a simple Bayesian model as a candidate explanation for the empirical results. In the model, the community attempts to infer each individual's tendency to produce bad output given different priors about each individual and the possibility that anyone might make a mistake. Eminence is defined as a prior reputational state featuring precise beliefs that an individual is a high quality type. When bad output is revealed, the model shows that (1) being eminent helps you; (2) the presence of a more eminent team member hurts you but the presence of a less eminent team member helps you; and (3) eminent teammates hurt you less when you yourself are eminent. The empirical results thus appear broadly consistent with a Bayesian inference problem, where the community assigns blame given priors over the individuals involved and their interactions. While simple, the model captures the suite of empirical findings in an intuitive manner and identifies key primitives that may extend to a broad set of teamwork settings.

2. Literature and Context

Teamwork is a ubiquitous feature of modern production and organizations, where collaborative work is seen from assembly lines to entrepreneurial teams to surgical suites and

appears across industrial, agricultural, and service sectors (e.g., Cohen and Bailey 1997, Wuchty et al. 2007). Yet teamwork raises challenges, including challenges in finding appropriate reward structures (e.g., Holmstrom 1982, Welbourne et al. 1995; Wageman and Baker 1997; Bikard et al. 2015). When individuals join together in production, it can be difficult for outsiders to discern the separate inputs of individual team members. This information gap can undermine an organization or community's capacity to reward team members appropriately (e.g., Holmstrom 1982) and can lead outsiders to rely on additional sources of information in making inferences, including the existing reputations of the parties involved (e.g., Merton 1968).

Indeed, information challenges may be overcome through reputation and learning in many contexts, as suggested by large theoretical and empirical literatures.⁶ Merton's "Matthew Effect" provides a canonical analysis (Merton 1968). On the one hand, an eminent team member can enhance demand for the product (a research article in Merton's setting, where an eminent author attracts greater attention to the output) thus creating a positive spillover on other team members by elevating attention to their work. On the other hand, and according to Merton's primary analysis, the presence of an eminent team member may limit credit for others as the community infers that the eminent team member is more responsible for the output. Thus, while partnering with a high-reputation teammate may enhance demand for the given output, it may also make it difficult for the less-established teammate to become substantially rewarded herself. Such a credit allocation effect, should it be operating, may in turn create additional challenges in team production settings. Indeed, credit allocation is the fundamental consideration in classic theories

⁶ Reputation's role amidst information problems has been emphasized in economics, sociology, and management literatures, with classic analyses including Shapiro (1983) and Rao (1994).

of teamwork and organizations (e.g., Holmstrom 1982, Aghion and Tirole 1994) and may also impact career progress; for example, if young team members struggle to garner credit for their efforts, their interest in the career itself may dim (Stephan 2012, Jones 2010). Understanding reward systems in team production thus appears as a key for understanding team function, team assembly, and career choice, and hence appears as a potentially critical issue for modern management and the economy at large given the prevalence of teamwork today.

Recent literature has examined the reputation effect specifically in the setting of science. Simcoe and Waguespack (2011) show that attention to proposed Internet standards increases substantially when the presence of an eminent author's name is revealed as opposed to hidden. Azoulay et al. (2013) show that citations increase to a researcher's prior body of work after the researcher becomes a Howard Hughes Medical Investigator, a high-status award in the biomedical sciences. Both studies indicate that positive reputational shocks can improve community awareness or perceptions of existing output. By contrast, Lu et al., (2013) and Azoulay et al. (2015) study negative reputational shocks in science, demonstrating penalties from retraction. Azoulay et al. (2017) show that retraction penalties differ by author standing across different retractions.

This paper departs from prior literature by focusing on credit allocation *within teams*. Namely, we examine the allocation of retraction penalties among team members when individual inputs to a team-produced retraction appear unobservable to outsiders. The setting of team science allows us to examine not just how established reputations influence community use, but how differential reputations *within* a team influence individual-specific consequences. We thus embrace the centerpiece of Merton's seminal analysis, examining the role of an individual's prior reputation and the potential entanglement of reputations in assigning rewards within teams. The communication hypothesis, normally an advantage, suggests that eminence may attract extra

attention to the negative event and thus amplify consequences for the individuals involved. The credit hypothesis suggests two distinct alternatives. On the one hand, a strong reputation may protect an individual in case of falsehood, where the community infers that a less-established team member was responsible for the problem. Thus the Matthew Effect may also work in reverse, with eminence not only attracting good credit but also deflecting bad credit. On the other hand, the credit hypothesis may suggest that the community sees the eminent individual as being “in charge” and directing events, in which case the eminent individual may take the blame for mistakes, just as they get credit for successes.

Given a rich set of plausible mechanisms, we treat our analysis primarily as an empirical question and seek to establish first-order facts. Having presented these facts, we then return to theory in Section 5 and provide a simple Bayesian interpretation that emphasizes the credit-inference aspects of the problem. This theoretical approach shows how strong prior beliefs can both insulate one’s own reputation and deflect consequence onto others.

Azoulay et al. (2017), in a related contribution, find that eminent scientists can be especially harshly penalized in the wake of a retraction in cases involving fraud or misconduct. The sentiment of their empirical results differ from ours, a difference that can be attributed to distinctions in the research question, sample composition, and empirical approach. In term of the research question, Azoulay et al (2017) compare retraction penalties across different retraction events. They use all retraction cases including solo-authored retractions and multiple retractions from the same author. In doing so, they largely focus on one author per retraction (the principal investigator) and therefore examine variation by author standing *between* teams and *between* retractions. Their context is one where the blameworthy party is typically obvious and where eminent authors have more reputation to lose in the severe case of misconduct. In contrast, we

address a team production issue *within the same retraction*, i.e. whether eminent team members receive more or less blame than their less eminent teammates, and further focus on cases where individual responsibility is unclear. Hence, we focus on *within-team* variation and study single retraction cases, for which the uncertainty about who to blame is substantial. We discuss these distinctions further below when we present our data, sample and empirical approach.

3. Data and Empirical Framework

Our data comes from the largest known repository of scientific knowledge, the Web of Science (WOS) from Thomson Reuters, which includes more than 32 million research articles published in over 15,000 journals worldwide since 1945. This database includes bibliographic information for each paper (authors, journal, publication year, etc.) together with citation linkages between each paper. The WOS further includes retraction notices, which describe the time and reasons for each retraction and whether the errors are reported by the authors.

3.1. Treated Papers

In our study, we focus on changes in citations to an author's *prior published work*. We focus on prior work, i.e., papers published before the retraction event, because this work is in a fixed published form, allowing us to isolate changes in usage of this work from changes in the work itself. Moreover, focusing on prior published work allows us to construct counterfactual cases by matching the prior work to other papers in the WOS that followed very similar citation profiles prior to the retraction event. We refer to each prior publication by authors involved in the retraction as a treated paper.

To build the sample of prior work, we confront a typical challenge in the WOS, where neither author names nor affiliations are uniquely identifiable. For example, different authors may share the same name. Relying on the name alone would then lead to the inclusion of work not

written by that author. To address this, we track the publication history of an author via her self-citation network, assuming that researchers tend to cite their own works in the same field.

In our primary analysis, we focus on “single” retraction events, where the coauthors are involved in only one retraction between 1993 and 2009. These cases present the community with an inference challenge in determining who is to blame within the team, raising the possibility of Matthew Effect like outcomes. By contrast, authors with multiple retractions, represent (extreme) cases where an author is revealed to have produced many false works, which makes the inference challenge for the community straightforward. We consider multiple retractions, where the blameworthy party becomes obvious, as a falsification test in Section 5.

The retraction notices in the WOS indicate whether or not the errors were reported by the authors themselves. Lu et al. (2013) show that retractions trigger citation losses to an author's prior work but these penalties disappear if the author(s) self-report the error.⁷ Therefore, to examine how retraction affects authors by differential eminence, our retraction sample focuses on cases where retractions were not self-reported.

In the sample period we located 513 singular retraction events and 95% of these retracted papers (489) were written by more than one author. Among these team-authored retractions, 57.3% (280) were not self-reported, 32.3% (158) were self-reported, and 10.4% (51) had unclear or unknown retraction reasons. For our main retraction sample, we identified each authors' prior work published before the retraction. Changes in citations to these papers are the objects of our

⁷ The absence of citation losses with self-reported retractions may indicate that the community interprets these events as innocent mistakes, and/or there may be some offsetting advantage through self-reporting in signaling the authors' trustworthiness. See Lu et al. (2013).

empirical analysis. The procedure for identifying prior work of an author, which is based on their citation network, is described in the Online Appendix.

3.2 Control Papers

Because citation patterns differ across disciplines and by time since publication, we construct a control group to match each “treated” paper in the pre-retraction period. The underlying assumption is that both treated and control articles will continue the same course of citations if there were no retraction influencing the treated paper. This methodology draws on an identification approach first used in the context of scientific outputs by Furman and Stern (2011).

For a treated paper i published in field f and year p , we search for control papers within the same field and the same publication year. Using the WOS, we are able to search across millions of papers to find controls that are minimally distant within the same field, where field is defined by the 252 field categories that WOS uses to classify thousands of journals. In particular, for each non-treated paper j in this pool, we define the arithmetic distance between i and j as

$$AD_{ij} = \sum_{t=p}^{r-1} (c_{it} - c_{jt}) \quad (1)$$

and the Euclidean distance between i and j as:

$$ED_{ij} = \left[\sum_{t=p}^{r-1} (c_{it} - c_{jt})^2 \right]^{1/2} \quad (2)$$

where c_{it} denotes the citations paper i receives in year t and r is the year of retraction. Both distances measure the citation discrepancy between papers i and j . Arithmetic distance allows positive and negative differences to offset each other while Euclidean distance is direction free.

The quality of control group matching is assessed in Figure A1 of the Online Appendix. Because we access the entire WOS, we can find substantially closer controls than is normally the

case in other applications of this methodology (Furman and Stern 2011; Furman et al. 2012; Azoulay et al. 2017). For example, focusing on the ten papers with the lowest Euclidean distance to a treated paper, the upper-left panel of Figure A1 shows that the average Euclidean distance between the ten controls and the treated paper has high density around zero. The density drops smoothly at higher distances except for the bin of 50 or more (which is driven by some treated papers that were exceptionally highly cited before retraction). As shown in the bottom-left panel of Figure A1, the average arithmetic distance between these ten controls and the treated paper has substantially more density on the negative side, so that these controls on average underestimate the citation flow of the treated papers. Focusing instead on the single control paper with the lowest Euclidean distance, we are able to find a perfect match for 36.1% of the treated papers. When we cannot find a perfect match, the arithmetic distance of the single best control is negative on average, though it is more evenly distributed on both sides of zero than the ten-control sample.

To achieve a sample that balances close matches with sample size, we consider the two nearest neighbors, one from above (with positive AD) and one from below (with negative AD). As shown in the bottom-right panel of Figure A1, the density of the average arithmetic distance of these two controls is either exactly zero or concentrated in the neighborhood of zero. In particular, the two nearest neighbors yield an average of zero arithmetic distance for a large share (68.5%) of our treated papers. This sample, with zero distance, is the main sample used in our analysis. In practice, we have 276 retraction events where authors have closely-matched prior work.⁸

⁸ We lose four retraction cases by focusing on prior publications that have close control matches prior the retraction event.

Our control approach is novel to the economics of science literature. Compared to the traditional control approach that attempts to match papers within the same journal and year (Azoulay et al, 2017), our method uses a larger pool of candidate control papers and enables us to find matches with an average of zero arithmetic distance on pre-retraction citation counts.⁹

Overall, by focusing on these 276 team-authored, single retraction events that were not self-reported, our sample includes 732 authors. The mean number of prior publications for these authors is 24.5. The mean number of prior publications for these authors where the two nearest-neighbor controls have zero average arithmetic distance is 16.8 giving a main treatment sample of 12,290 prior publications. This sample, with each treatment paper and its two controls, includes 419,239 paper-year observations. Note that some prior publications will be counted more than once if multiple authors in the sample collaborated on them.¹⁰

3.3 Definitions of Author Eminence

We construct three standard measures for an author's eminence: publication counts, total citations received, and the h-index. The h-index (Hirsch 2005) attempts to account for publication quantity and quality in a single measure: the number h is the largest scalar for a given scholar such that the scholar has published h papers each of which has been cited at least h times. These

⁹ While our matching approach allows for extremely close matches in citation dynamics, journal-level matching may more closely match the narrow subfield of the retracted article.

¹⁰ The estimation sample of 12,290 prior publications from retraction authors is constituted by 10,209 unique prior publications, some of which are shared by multiple authors. We cluster standard errors by the retraction event (i.e., the 276 cases) to allow for correlated shocks across the prior work within a given author and across authors involved in the same retraction event.

measures, which are commonly used as indications of eminence in the scientific community, are calculated using the papers and citations within the WOS. They are calculated for each author in the year just prior to the retraction event.

Taking each treated author as an observation, Figure A2 plots the distribution of the h-index at the time of retraction. In the main part of our statistical analysis, we define the “absolute eminence” of an author using the continuous measures of paper counts, total citations, or h-index. As alternative measures, we also define simple dummy variables to indicate whether an author is in the top 10th percentile of the eminence measure.

Because we focus on retractions of team-authored papers, we also define relative measures of social standing based on whether an author has the highest or second highest standing in the team at the time of retraction. These authors are referred to as “relatively eminent.” Compared to the absolute measure of author eminence, relative eminence helps us examine differential standings within a team, even if all team members have high or low eminence metrics in absolute terms. The relative eminence measure can also help filter out heterogeneity in the absolute measures across different academic fields.

3.4 Summary Statistics

Table 1 provides two panels of summary statistics: Panel A, at the author level, considers the standing of each treated author at the time of retraction; Panel B, at the paper level, considers summary statistics for the retracted papers and prior work. Panel A shows that authors of a retracted paper had, at the time of retraction, a mean of 24 prior publications, 1,071 citations, and an h-index of 10. Whether measured by total counts of prior work, total counts of citation, or h-index, these author measures appear dispersed and right-skewed. Defining eminence by whether an author’s prior-retraction h-index is among the top 10 percentile, Panel A shows that eminent

authors have much more publications, receive much more citations, and have been publishing over a greater number of years than ordinary authors.

The retracted papers have 5.9 authors per paper on average (Panel B). Among the prior publications of these authors, 45.5% were published in the 2000s, 40.0% were published in the 1990s, and 14.5% were published in the 1980s. The mean yearly citation count for the prior publications is 3.0. With our sample ending in 2009, the mean age of a prior publication in 2009 is 11.6 years. The mean age of an author's prior publications in the year that author experiences a retraction is 8.5 years.

3.5 Estimation Equation

Our identification strategy employs difference-in-differences. We examine the citation effects of retraction shocks comparing the pre-post differences for treatment papers with the pre-post differences for control papers, while further comparing these differences across authors with different standings. The regression model is

$$\Pr(y_{iat}) = f(\alpha_{ia} + \mu_t + \beta_1 \cdot \text{Treat}_i \cdot \text{Post}_{kt} + \beta_2 \cdot \text{Standing}_a \cdot \text{Treat}_i \cdot \text{Post}_{kt} + \beta_3 \cdot \text{Standing}_a \cdot \text{Post}_{kt} + \beta_4 \cdot \text{Post}_{kt}) \quad (3)$$

where i indexes article, a indexes author, t indexes year since publication, and k indicates a treatment-control paper group. The dependent variable, y , denotes counts of citations to article i at time t for author a . Fixed effects for each paper and author with a retraction (α_{ia}) and each year since publication (μ_t) capture the mean citation pattern of articles. Treat_i is a dummy variable that equals 1 if article i is a treatment paper, and Post_{kt} is a dummy variable that equals 1 if year t is after the retraction event for a given treatment and control group k . Standing_a measures the

eminence of the treated author in the year prior to the retraction.¹¹ For clarity in interpretation, we normalize $Standing_a$ as a z-score, so that $Standing_a = 0$ corresponds to the average treated author and $Standing_a = 1$ indicates an author one standard deviation above the mean. For the three standing measures, the means and standard deviations are given in Table 1.

The coefficient β_1 captures the effect of the retraction shock on citations to prior work of ordinary authors, compared to closely-matched control papers. The coefficient β_2 captures any difference in the effect on authors with an eminence measure one standard deviation above that of the average treated author. We estimate (3) using the standard Poisson model for count data. While there are 10,209 unique prior publications in the treated sample, to be conservative we cluster the standard errors by the retraction event, giving 276 paper groups.¹²

The key identification assumption is that the prior work would continue the same course of citations as its control papers had the retraction not occurred. Later, we will present a placebo test to further support this assumption. To the extent that this assumption may be less valid if the prior work is published close to the retraction time and therefore provides a shorter time window for matching control papers, we will also later exclude such cases as a robustness check.

4. Results

As a first look at the raw data, Figure A3 shows the citation flows to prior publications before and after retraction, separating the data by author standing. On the horizontal axis, zero demarcates

¹¹ Note that the interaction $Standing_a * Treat_i$ is absorbed by the paper-author fixed effect (α_{ia}).

¹² This approach allows arbitrary correlations in the errors across time for a given treated paper, across treated papers by the same author, and across all treated papers by distinct authors who were later involved in the same retraction event.

the year of retraction. The solid blue line shows treated papers, and the dashed red line shows control papers. In the upper row we separate out the author with the greatest h-index on the team (left panel) from the other team members (right panel). The bottom row distinguishes the top two highest h-index authors from the other authors of the retracted paper.

These graphs suggest that the post-retraction citation decline is noticeably negative for more ordinary authors, while relatively eminent authors experience no citation loss. These pictures of the raw data group papers from fields with different citation dynamics and also group papers with different lengths of observed citation histories. The rest of this section analyzes the data using regression models, presents our central findings, and considers robustness checks.

4.1 Main Results

Pooling the data across authors in our sample, we first confirm that retraction has a significant negative spillover effect on citations to the authors' prior work. The regression results are presented in Figure 1, drawing on the approach of Lu et al. (2013).¹³ Compared to the control papers, the annual flow of citations to prior publications falls 4.8% ($p < .0001$) in the first two years post retraction and 13.0% ($p < 0.0001$) five or more years post retraction. This suggests that retractions lead to substantial citation declines to prior work in team-authored papers, which is consistent with the results shown in Lu et al. (2013) for retracted papers more generally.

4.1.1 Absolute Standing

Table 2 reports results from our main specification. We highlight the difference-in-differences coefficient on *Treated * Post ($t \geq 1$)* and the relative effect on individuals with greater

¹³ This graph differs slightly from the analysis in Lu et al. (2013) because, here, we are interested in and present team-authored cases, where Matthew Effect like outcomes may emerge.

standing from the coefficient on *Standing * Treated * Post (t>=1)*.¹⁴ The latter indicates whether a treated author with greater absolute standing at the time of retraction experiences different citation consequences for their prior work. There are three columns in the table, differing by measures of eminence, using total prior publications, total prior citations, and the h-index respectively.

All measures show that the main effect (for those with the mean absolute standing measure) is negative and statistically significant. Meanwhile, the three continuous measures show that higher absolute standing offsets the negative main effect, with statistically significant interactions when using total prior citations or the h-index. Broadly, the coefficients are of similar magnitude across the three measures. Focusing on column (3), a retraction leads to a 10.8 percentage point decline in yearly citations to prior work for an average author. This main effect is offset by a 2.9 percentage point smaller decline in citations per one standard deviation increase in absolute eminence.¹⁵ This finding suggests that having higher standing at the time of retraction may help alleviate the reputational harm due to retraction. Being more eminent suggests a protective effect. Figure 2A repeats the analysis of Figure 1 but now observing how the citation losses to prior work differ between eminent and non-eminent authors.¹⁶ Eminent authors are defined as those with an h-index in the upper 10th percentile, while other authors are classified as non-eminent.

¹⁴ We separate out the retraction year itself ($t=0$) because the exact time of retraction could occur early or late within the year.

¹⁵ The marginal effect (in percent) of a one-unit change in a variable is $\exp(\text{coefficient})-1$.

¹⁶ The econometric specification, as indicated in the figure, includes separate post period dummies for 1-2, 3-4, and 5+ years after the retraction event.

Commensurate with Table 2 and Figure 1, we see large citation declines to the prior work of non-eminent authors and this decline increases with time after the retraction. By contrast, eminent authors see modest if any decline in citations to their prior work.

4.1.2 Standing Relative to Coauthors

Beyond one's own absolute standing, we further consider the implications of coauthors' relative standing, as emphasized by Merton (1968). To capture relative standing within the team, we separate out those authors who have the highest standing on the team, even if they don't have high standing in an absolute sense. In particular, we define a dummy equal to one if a treated author has the highest measured standing or, separately, if the author is among the two individuals on the team with the greatest standing. As before, author standing is measured in the year prior to the retraction and is alternatively defined using the total number of prior publications, the total citations received, and the h-index.

Table 3 reports the results. As before, the main effect for those with low relative standing is negative and statistically significant across all specifications. When looking at the highest standing author (Columns 1-3), we consistently see large, offsetting positive point estimates, which are significant at the 10% level when using the total number of prior citations or the h-index. Looking at the two authors with highest relative standing (Columns 4-6), we see larger point estimates and greater statistical significance across the measures. Moreover, the estimates for relatively low-standing authors become increasingly negative, suggesting that the top two individuals may more neatly divide high and low standing individuals within the typical team.

Figure 2B repeats the analysis of Figure 2A but now using relative standing, where the relatively eminent authors are defined as the top team member by h-index, while the relatively non-eminent authors are the other team members. We again see large citation declines to the prior

work of non- eminent authors, and larger declines with time after the retraction. By contrast, the most eminent team member sees modest if any decline in citations to his or her prior work.

4.1.3 Team Configuration

Further tests generalize the empirical model (3) to consider more textured team configurations. In particular, using binary absolute eminence measures (the top 10th percentile as the cutoff), we consider four different configurations among the authors of the retracted paper. These regressions include dummy variables to indicate whether (i) one's own standing is ordinary and the highest-standing coauthor is ordinary, (ii) one's own standing is ordinary but a coauthor is eminent, (iii) one's own standing is eminent and the highest-standing coauthor is ordinary, and (iv) one's own standing and a coauthor are both eminent (the omitted category in the regression). Here, the coauthor refers to the best coauthor in a team. The results are presented in Table 4, columns (1)-(3), with each column using a different measure of standing.

We see that the spillover effect on prior work is most negative when one has ordinary standing and is in the presence of an eminent coauthor. This finding generalizes across the standing measures with varying statistical significance. Taking column (3), for the h-index, the loss on prior work is 15.2% larger when you are ordinary and your coauthor is eminent, compared to the baseline where you were also eminent yourself. Indeed, being eminent yourself suggests little citation losses to your prior work and regardless of the standing of your coauthors, which is seen both in the main effect (you and a coauthor are eminent) and in the interaction effect where you are eminent and your highest standing coauthor is not.

The above approach considers an author's own standing and its interaction with the highest standing coauthor. While simple and transparent, other approaches may be additionally informative as team configurations can be more complex. In particular, teams typically contain

“rookie” coauthors, i.e. those with no prior publication history in our data. As the least established members of the team, these individuals may play important roles in modulating the effect of retractions on the other coauthors.

Table 5 presents additional analyses in light of rookie coauthors. Focusing on the h-index, the first column repeats our basic analysis in Table 2 column 3 but now adds team size fixed effects and the percentage of rookie coauthors on the retracted paper.¹⁷ The earlier findings regarding author standing are robust. The new finding is that the presence of rookie coauthors tends to limit substantially the citation losses for the other authors. The second and third columns of Table 5 further examine the role of rookie coauthors for eminent and ordinary authors separately. Here we see that the presence of rookie coauthors has a weak effect for the eminent (who already experience little citation loss) but can substantially offset the losses for ordinary authors. For ordinary authors, moving from no rookie coauthors to all rookie coauthors offsets 88% of the citation losses.

Taken together, Tables 2 through 5 show a consistent pattern. After retraction, the average author experiences large citation losses to their prior work. The citation loss for ordinary authors is amplified when working with an eminent coauthor and attenuated when working with rookie coauthors. Eminent authors, meanwhile, show little citation losses to their prior work, regardless of the standing of their coauthors. A variety of additional tests discussed below further support these results and tend to strengthen their magnitudes or statistical precision.

4.2 Additional Tests and Robustness Checks

¹⁷ The team size fixed effects are interacted with the treatment and post dummies; the inclusion or exclusion of these team size fixed effects has little effect on the results.

We consider here additional tests to explore the robustness of the results and further sharpen the empirical findings. These analyses are presented in Tables A1, A2 and A3, which further investigate the main results but with changes to the sample or econometric specification. Table A1 repeats the analysis of Table 3, focusing on relative standing in the team to see if relatively ordinary authors continue to experience large citation losses to their prior work while the relatively eminent authors experience smaller losses. Table A2 repeats the analyses of Table 4, examining whether ordinary authors experience especially large citation losses in the presence of an eminent coauthor. Table A3 repeats the analysis of Table 5, examining whether the citations losses are milder in the presence of rookie coauthors.¹⁸

4.2.1 Recent Papers

Older papers may receive fewer ongoing citations. Because eminent authors may have an older distribution of papers than ordinary authors do, this tendency could contribute to smaller citation losses among the relatively eminent. Figure A4 shows that the mean annual citations to treated papers falls to two in the tenth year after publication. We therefore reconsider our analysis excluding prior articles published more than ten years earlier than the retraction year. As a result, 69.8% of treated papers and 50.5% of paper-year observations are kept in the subsample.

Tables A1-A3 reconsider our core findings for this restricted sample, with the results presented in column (2) in each table. We see that the results are robust. For example, in Table A1, citations fall by 14.4% for lower-standing authors after retraction and the difference with eminent researchers is 11.0%, which is very similar to the results for the main sample. The results

¹⁸ For brevity, these analyses use the h-index as the measure of author standing. Appendix

Tables A4-A13 provide additional results using the other standing measures.

for team configuration in Tables A2 and A3 are again robust, with similar magnitudes and statistical significance as with the main specifications.

4.2.2 Actively Cited Papers

A related approach restricts the sample to publications that are being positively cited at the time of retraction. This issue is somewhat different from old papers per se because zero citations could occur soon after publication, especially for ordinary authors who do not have many high quality publications. To deal with this issue, we exclude all prior work that has zero citations in the year before retraction. Compared to the main sample, this subsample includes 68.9% of treated papers and 59.1% of paper-year observations. The results are presented in column (3) of Tables A1-A3. We see again that the results all remain robust.

4.2.3 Citation Distance

Another related issue is that the (relatively abundant) prior work of eminent authors may on average be farther in idea space or social space from the retracted paper. To the extent that scientific communities and reputations tend to be field-specific, eminent authors may experience relatively mild citation declines on average if their prior work tends to sit outside the focal field and community of the retraction.¹⁹ To assess this possibility, we reexamine our results in a sample restricted to low citation distance from the retracted paper. Namely, we consider the differential effects of author standing within the subsample of treated papers that are one degree of separation in the backwards citation network from the retracted article (i.e., prior papers that were directly

¹⁹ That said, it is less clear how such differences in prior work would explain our main results around team configuration – i.e., that ordinary authors experience worse losses in the presence of eminent coauthors and milder losses in the presence of rookies.

cited by the retraction article). This restriction is substantial: it reduces the treatment sample to only 10.8% of the treated papers and 8.0% of the paper-years observations.

Looking at Table A1 column (4), we see that once again ordinary authors experience large citation losses to their prior work and that this effect is substantially offset for eminent authors. The magnitudes are somewhat greater on both dimensions than with the full sample. Thus, the attenuation of citation losses that is seen with eminence appears robustly within the narrow sample of the most closely related prior work. This finding indicates that the relatively mild citation losses experienced by eminent authors comes not because they have more prior work that is more distant, but rather appears among uniformly “near” prior work. Tables A2 and A3 tend to show broadly similar results to the main sample although with somewhat greater noise, which is perhaps not surprising given the large drop in sample size. The exacerbating role of eminent coauthors on ordinary coauthors is noisier than in the main sample (Table A2), while the attenuating role of rookie coauthors is similar and slightly larger than in the main sample (Table A3). Table A6 considers these results with a broader range of standing measures and shows similar and more statistically significant results using other standing measures.

Note also that, since we use self-citations to compile prior work for a given author, our sample is relatively likely to capture an author’s prior work in closer fields (Wuchty et al, 2007) but may more weakly capture prior work written by that author in distant fields. If retraction effects weaken with distance from the focal field, and if eminent coauthors are more likely than less established teammates to have diverse research areas, then sampling closely-related work would tend to understate the magnitude of the Reverse Matthew Effect. That is, the differential advantage of eminence would be greater than the advantage already seen in the empirical results.

Overall, after restrictions on the treated sample, including by age of prior work, ongoing citations to prior work, or citation distance to prior work, we see that within “near” prior work, the findings continue to be characterized by relatively large citation losses for ordinary authors, relatively muted losses for eminent authors, and broadly similar amplification/attenuation of losses depending on the presence of eminent or rookie team members,

4.2.4 Citation Losses Excluding Self Citations

Retractions may also affect future publishing prospects, and differentially for eminent and non-eminent authors. Citation declines to prior work might then potentially reflect less a direct community response and more a decline in the capacity of authors to cite their own prior work, once any differential retraction effects on an author’s career take hold. To further focus on the community response, we reconsider the analysis excluding self-citations from the citation counts. These results are presented in column (5) of Tables A1-A3. The findings are similar to the earlier results. Interestingly, the magnitudes are, if anything, slightly larger. This finding, which nets out self-citations, further points toward the effect on prior work coming from the broader community, as opposed to the citation behavior of the retraction authors themselves.

4.2.5 Further Robustness Checks

We conduct additional robustness checks estimating different samples and models. First, we replace our Poisson estimation with OLS estimation. The OLS results are reported in column (6) of Tables A1-A3 and appear broadly similar to the Poisson results. Second, we cluster the standard errors by treatment-control paper group instead of retraction event. These results, presented in column (7) of Tables A1-A3, show increased statistical precision and confirm that the results we have presented are conservative. Third, we consider an alternative and noisier set of control papers, taking the 9th and 10th nearest controls for each treated paper, rather than the

two nearest controls. As shown in column (8) of Tables A1-A3, the magnitudes of the results appear broadly similar although, not surprisingly, the noisier controls lead to somewhat less precise estimates. Fourth, we exclude prior work that has a short citation history before retraction, which could hurt our ability to find effective counterfactual controls. Specifically, we exclude prior work published within three years of the retraction. Results are shown in column (9) of Tables A1-A3 and appear slightly stronger than our baseline specification. Fifth, we consider a specification that also includes author position (first, middle and last) to control for the author's role in the retracted teamwork and, as shown in column (10) of Tables A1-A3, the results are again robust.²⁰ This last specification will be further discussed in Section 5.

4.2.6 Placebo Test

As a final check on our approach, we consider a placebo exercise to see whether the evolution of control paper citations is sensitive to author standing in the absence of retraction. In particular, using our control papers, we examine whether papers matched according to very similar initial citation patterns also have similar later citation patterns regardless of standing.²¹ We find that standing does not predict future citation paths, conditional on initially similar

²⁰ An alternative test includes the career age of an author in the regressions to control for the author's role in the retracted paper. See Table A13 in the Online Appendix.

²¹ Specifically, we randomly sample 500 pairs of control papers. For each author on these 1,000 papers, we build their body of prior work and determine eminence measures for each author. By construction, each control paper in a given pair has similar citation behavior up to the retraction event year. We then analyze whether control papers with higher standing authors diverge in their citations, after the retraction event year, from control papers with lower standing authors.

citation paths, as detailed in Table A14. This analysis further suggests that our control strategy is effective for estimating counterfactual citation paths in the absence of retraction.

5. Interpretations and Discussion

The above empirical analyses establish several striking facts regarding retraction shocks and their differential effects across team members. We call these results a “Reverse Matthew Effect”, as they echo the ideas that animate Merton’s Matthew Effect, only now in the reverse case where we consider bad events. We find that retraction shocks lead to substantial declines in citations to the prior work of ordinary coauthors. By contrast, for eminent coauthors, retraction shocks provoke much less if any citation loss to their prior work. Furthermore, citation losses for ordinary coauthors are especially severe in the presence of an eminent coauthor on the retracted publication but less severe in the presence of rookie coauthors.

This section further discusses the empirical results in light of the ideas that Merton proposed. Returning to Merton’s credit mechanism, we first formalize the idea that the community makes ex-post inferences about individual contributions in team settings given prior reputations and the uncertainty over who was responsible for the output. A simple Bayesian model of this mechanism is shown to provide a parsimonious, candidate explanation for the empirical results. We then discuss potential alternative interpretations and examine a falsification test where the community can easily infer the bad actor.

5.1 A Model

Let there be two types of agents, who differ in their tendency to produce “good” output. The community does not observe an individual's type directly but rather makes inferences about it by observing the individual’s output. The community's belief about the individual's type characterizes

that individual's reputation.²² In particular, let an output have a quality characteristic that takes one of two states, $Y \in \{Y_{good}, Y_{bad}\}$. An individual can have a high or low tendency to produce good output. Let an individual's type be $\theta \in \{H, L\}$, representing a "high" and "low" type individual, respectively, where the low type produces "bad" output with a greater frequency than the high type

$$\Pr[Y_{bad}|L] > \Pr[Y_{bad}|H] \quad (4)$$

and we use the shorthand $p_\theta = \Pr[Y_{bad}|\theta]$. An individual's "reputation", R , is defined as the probability that the individual is the high type, $R = \Pr[H]$. In summary, the background probability of producing bad output depends on the author's type. How to distinguish the type given the observed output is the heart of the inference problem.

5.1.1 Solo Production

To develop basic intuition, first consider the reputational updating for an individual who, working alone, produced output with characteristic Y . Let the individual i have a given prior reputation, R_i . Bayes rule says that the posterior belief about i 's type, which we denote R_i' is

$$R_i' = \Pr[H_i|Y] = \frac{\Pr[Y|H_i] \Pr[H_i]}{\Pr[Y]}.$$

Using the law of total probability in the denominator and definitions above, we can thus express the reputational change upon retraction as

$$\frac{R_i'}{R_i} = \frac{1}{R_i + \frac{\Pr[Y|L]}{\Pr[Y|H]}(1 - R_i)}. \quad (5)$$

²² In our empirical context, a "bad" output concerns the possibility that a given paper, regardless of how important it may otherwise seem, contains a severe enough mistake so that the paper will be retracted (i.e., the paper is not actually true).

Given that low types are more likely to produce bad output, as defined in (4), it follows by inspection of (5) that the individual's reputation will fall after a bad event and rise after a good event.²³ Note also that in the extreme case, where $R_i = 1$, the individual is fully protected from the reputational consequences of retraction; as is standard with a Bayesian model, having a tight prior about the individual means that new events will have little further effect on beliefs.

5.1.2 Team Production

We now consider the richer case of team production, which allows us to characterize how the reputation of one team member can influence the credit another receives. In particular, let the output be produced by a team of two people, indexed $i \in \{1,2\}$, who have independent priors. Again following Bayes' Rule, the two-person analogue to the reputational updating problem after an event with characteristic Y is now²⁴

$$\frac{R_1'}{R_1} = \frac{1}{R_1 + \frac{\Pr[Y|L_1, L_2](1 - R_2) + \Pr[Y|L_1, H_2]R_2}{\Pr[Y|H_1, L_2](1 - R_2) + \Pr[Y|H_1, H_2]R_2}(1 - R_1)}. \quad (6)$$

²³ We have defined $\Pr[Y_{bad}|L] > \Pr[Y_{bad}|H]$. Therefore, for a bad event the denominator is greater than 1 and the reputation deteriorates. For a good event, it also follows from (4) that $\Pr[Y_{good}|L] < \Pr[Y_{good}|H]$ and so the denominator is less than 1 and reputation improves.

²⁴ In particular, by Bayes' Rule, the posterior belief about individual 1's type can be written

$$R_1' = \Pr[H_1|Y] = \frac{\Pr[Y|H_1, L_2] \Pr[H_1, L_2] + \Pr[Y|H_1, H_2] \Pr[H_1, H_2]}{\Pr[Y]}$$

Using the law of total probability to rewrite $\Pr[Y]$, the definition of R_1 , and rearranging, one obtains the expression in the text.

Reputational updating for the given team member thus depends on three elements: (a) the team member's own prior reputation, R_1 ; (b) the prior reputation of the other team member, R_2 , raising the possibility of Matthew Effect type outcomes; and (c) the production technology mapping individual types to joint output. This last feature is encapsulated by the $\Pr[Y|\theta_1, \theta_2]$ terms.

5.1.3 The Reverse Matthew Effect

As seen in (6), the reputational update will depend on the production technology for the (observed) joint output characteristic, Y . That is, how do the individual contributions of the team participants determine the probability of a given output state? In the context of our empirical analysis, we focus on bad events, where the paper is false. For clarity, and to emphasize the “Reverse Matthew Effect” case, we can use $Y_{good} = 1$ representing that the output is “true” and $Y_{bad} = 0$ representing that the output is “false”.

The production technology for false output may naturally have a "weak link" technology. That is, if an input to the paper is false (the data is faked, the empirical or computational analyses are wrong, etc.), the paper itself turns out to be false, so that the quality of the joint output is

$$Y = \min\{y_1, y_2\}$$

where the individual contribution is $y_i \in \{1, 0\}$, representing a true or false input, respectively.

With this production technology, the probability that the joint output is false is then $\Pr[Y = 0|\theta_1, \theta_2] = 1 - (1 - p_{\theta_1})(1 - p_{\theta_2})$, where $p_{\theta} = \Pr[y = 0|\theta]$. Reputational updating will occur according to the following Lemma.

Lemma (Reverse Matthew Effect)

$$(i) R_1' \leq R_1; (ii) \lim_{R_1 \rightarrow 1} R_1' / R_1 = 1; (iii) \frac{\partial(R_1' / R_1)}{\partial R_2} \leq 0; \text{ and } (iv) \lim_{R_1 \rightarrow 1} \left(\frac{\partial(R_1' / R_1)}{\partial R_2} \right) = 0.$$

The proof is given in the Online Appendix.

These results can capture the empirical findings and provide some precise intuition for them. The first result states that reputation declines upon retraction. This result corresponds to the broad finding where the team members experience citation losses on average to their existing work. It is also consistent with the retraction penalties reported in Lu et al. (2013) and Azoulay et al. (2017). The second result states that a high reputation acts to limit the reputational decline from the retraction. This result corresponds to the findings in Table 2, where an already eminent team member experiences more limited negative consequences on average.

The last two results focus on the reputational entanglement across individuals that may emerge in a teamwork setting and thus speak most precisely to a “Reverse Matthew Effect”. The third result states that the greater the reputation of your teammate, the worse the effect on you. Thus, the Bayesian model predicts that the presence of an eminent team member exacerbates the reputational losses for the other team member. At the same time, the fourth result shows that eminence is protective against this spillover effect. Thus, while an eminent teammate can hurt you, it does not hurt you if you yourself are eminent. These theoretical results are closely consistent with the findings in Table 4, where ordinary authors experience worse effects the more eminent the coauthor (result iii), yet eminent authors see little effect from eminent coauthors (result iv). The empirical results in Table 5 also broadly correspond to these findings, where now we consider what happens when someone is paired with especially junior coauthors (i.e., rookies). Ordinary authors experience much smaller citation losses when paired with rookies (result iii), while eminent authors see relatively little influence from rookies (result iv).

These results are all intuitive in a Bayesian context, where the community is trying to infer the source of a mistake and must adjudicate between the team members and the background chance of a mistake. A well-established reputation deflects blame away from you and toward both your

teammate and background bad luck. If the teammate also has a well-established reputation, then the community will tend to blame background bad luck, and both individuals face relatively mild consequences. An unformed reputation, however, attracts blame, and the more so the better your teammate's reputation. Overall, this theoretical approach can provide a natural and parsimonious interpretation of the key empirical results of the paper.

It is useful to compare Azoulay et al. (2017)'s model with ours. Azoulay et al. (2017) assume that research communities can classify whether a retraction is due to misconduct or an honest mistake. If the research community already agrees that a retraction event is due to misconduct of an identifiable bad actor, the retraction will tarnish the bad actor's reputation. If the community characterizes a retraction as honest mistake, it attributes the retraction to background noise and hence does not update much on the author's reputation. This explains why Azoulay et al. only find significant retraction penalties in the cases of fraud or misconduct but not in the cases of honest mistakes. In comparison, we focus on the events where there is significant uncertainty as to who contributes to the bad output in a team-produced single retraction. In light of this uncertainty, our theory describes how the community makes inferences from the bad outcome and each author's prior reputation.

5.2. An Alternative Credit Inference Hypothesis

Within the class of credit inference explanations, an alternative inference problem involves task allocation within the team. In particular, science teams may feature a hierarchal nature, where eminent authors lead in the research design rather than in the technical analysis, where problems are more likely to emerge. In this view, eminent authors may receive less blame when retraction occurs because they are seen as unlikely to be responsible for the relevant tasks.

One way to test this idea is to control for position in the author list for the retracted paper. Noting that positioning in the author list typically informs the hierarchy of the team in science and engineering, we reconsider our main results adding dummies variables for the last author (usually the principle investigator) and middle authors (who play lesser roles). As shown in Column 10 of Tables A1-A3, adding such author-position variables to the regression model has little effect on the main results.²⁵

Another way to test this idea is to examine citation effects based not on author eminence at the time of the retraction but at the time the research was conducted, when task allocation would be determined. To do so, we constructed past-standing measures using the eminence measures for an author in the year the problem paper was published. Then we examined both types of author standing (at the time of retraction and at the time of publication) in the regression. For ease of interpretation, both types of standing are measured by a dummy for whether the absolute standing is in the top 10 percentile of all treated authors at that time. As shown in the first three columns of Table A15, being eminent at the time of retraction substantially reduces the citation losses using two of the three standing measures, while being eminent at the time of publication does not. This result appears inconsistent with a task allocation hypothesis. The last three columns of Table A15 restrict the sample to authors who had ordinary standing when the problem paper was published. Some of these authors became eminent and others remained ordinary by the time of retraction. The results suggest that ordinary authors who became eminent later, measured by total publications or

²⁵ Table A12 provides the regression results with these additional coefficients reported. The author position fixed effects in these regressions are found to be highly insignificant.

h-index, see little if any citation loss. These results further suggest that task allocation does not appear to be a key explanation for our main findings.

5.3 “Bad Actors” as a Falsification Exercise

We can further conduct a falsification test by studying a context where the guilty actor is obvious and hence prior reputation should no longer matter in allocating blame across team members. Namely, we can study “multiple retraction” episodes where a single common author appears across multiple team-authored papers that were retracted. These cases point strongly at the common author as the blameworthy party. To undertake this analysis, we repeat our sampling and econometric strategy for all multiple retraction cases in the WOS where there is a single common author. We define a “bad actor” as the common author across these multiple retraction cases and define “innocent actors” as the coauthors on these retracted papers. Appendix Table A16 provides basic summary statistics for the multiple retractions cases.

Two additional features distinguish this exercise from the study of single retraction episodes. First, multiple retraction cases are more noteworthy events, often involving systematic fraud, which can attract substantial attention. Hence the scale and scope of effects may naturally be different from single retraction events. Second, multiple retraction cases often occur over a string of years, which makes the timing in the econometric strategy less clean. To operationalize the analysis, we will use the retraction of the first paper to define the event year.

Table A17 presents the regression results. In column 1, we limit the sample to the “bad actors” and find they experience large losses in citations to their prior work. This is consistent with Azoulay et al. (2017). In column 2, we limit the sample to “innocent actors” and find the interesting result that they experience citation increases to their prior work, which may reflect increased attention that comes after retraction, as we discuss below. In column 3, we consider the full sample

of these authors and see that the citation decline for the “bad actors” appears especially large. Notably, and in line with the purpose of this falsification exercise, interactions with author standing are never statistically significant and are of inconsistent sign across specifications. Thus, prior reputation does not appear germane when the identity of the bad actor is known – either for the bad actors themselves or their innocent coauthors. This finding, as a falsification exercise, can further support an inference-based interpretation of our main results: prior reputation matters in episodes when the identity of the responsible actor is unclear.

5.4 The Communication Hypothesis

Merton’s Matthew Effect also emphasizes a “communication” hypothesis, where eminence attracts attention to the output and for which there is evidence in the literature (Simcoe and Waguespack 2011, Azoulay et al. 2013). In the standard Matthew Effect, which considers “good” events, this communication effect may help the less established team member, offsetting the credit sharing issue. Namely, even if the less established team member receives little credit *share*, a widely noticed output can make the impact large in absolute terms. With a “bad” event, the communication hypothesis could exacerbate effects on less established team members, as the presence of an eminent team member may make bad events more widely noticed.²⁶

Our empirical analysis, which examines differential effects within a team, studies the credit allocation aspect of the Matthew Effect rather than the communications hypothesis, where

²⁶ That said, it is also possible that less eminent scholars have more to gain (or less to lose) from fraud and thus, in equilibrium, may experience greater scrutiny of their papers and hence be more susceptible to retraction ex-ante (Lacetera and Zirulia 2011). Interestingly, this theoretical insight provides another advantage eminent scholars may have with regard to retraction.

attention can influence everyone in the team. The one place where we may see a suggestive role of attention per se is the case of innocent actors in the multiple retraction analysis of the prior section. Here we see that the innocent team members actually experience a gain in citations to their prior work, which is consistent with increased attention to these individuals (coupled with the community's inference that they are unlikely to be at fault). This finding is consistent with Simcoe and Waugespack (2011), although in this case the increased attention is not driven by eminence but rather newsworthy events.

More generally, while a communication mechanism may be operating in our primary context of single retractions, it does not appear capable of providing an alternative explanation for the results. Namely, were this mechanism all that was happening, then eminence should worsen the citation losses in general. Given that we find the opposite result – that ordinary authors experience substantially worse effects than eminent authors – the communication hypothesis does not appear to dominate. Nonetheless, the basic communication mechanism may still be operating in tandem with other forces. For example, if high standing is protective, then the communication channel may worsen things more for the less eminent in the presence of eminent team members, exacerbating the credit inference effects.

6. Conclusion

We have considered the consequences of bad events in team production. Our empirical context investigates journal article retractions in the sciences and demonstrates a striking asymmetry: Eminent authors experience little or no change in citations to their prior work after a coauthored retraction, while less eminent coauthors experience large citation losses, and especially in the presence of an eminent coauthor. We thus find a “Reverse Matthew Effect,” developing

Merton's canonical ideas about team production, showing that the less established team members appear especially vulnerable in the aftermath of negative events.

While our setting is science, the primitives of our setting – teamwork, difficulty in directly observing individual inputs, and differential reputations – generalize across many production contexts. For example, entrepreneurial teams mix publicly unobserved inputs into a collective output, and judgments about which individuals shaped the outcome may create important reputational consequences for serial entrepreneurs in attracting future financing and new teams (Hsu 2008). Medical errors, legal malpractice, and accounting fraud may all suggest inference challenges in assigning individual blame for collective failures in surgery, litigation and accounting practices. Similarly, the financial performance of venture capital, private equity, and hedge funds may all bear on the reputations of the individuals in the investment team.

Conceptually, Section 5.1 provides a team-production framework that can explain our findings assuming (1) individual inputs are not observed, (2) a bad input can ruin the collective output, and (3) Bayesian updating. In such contexts, team-based failures may create especially large reputational damage for less established team members, and especially when the team includes well-established individuals. The conceptual framework thus may act as a guide to potential generalizability of the findings. Empirical investigations of additional contexts provide exciting avenues for future work.

The findings around credit sharing also raise a rich set of additional theoretical issues. The link between reward allocation and effort incentives is the subject of an enormous literature on relational contracts whose predictions depend on information structures and the contracting environment (e.g., Holmstrom 1982, Aghion and Tirole 1994, Rayo 2007). Other authors have considered credit-sharing implications for team assembly (e.g., Bar-Isaac 2007, Costa and

Vasconcelos 2010, Bikard et al. 2015), leading to multifaceted but somewhat ambiguous results. More generally, literatures on the sources of team effectiveness (e.g., Cohen and Bailey 1997) and the emergence of teams within social networks (e.g., Reagans et al. 2004) also bear on the link between credit considerations and team formation. Given the empirical findings in this paper, in which reward allocation is found to be asymmetric across team members, further empirical and theoretical research on how reputational considerations influence team function and team assembly choices appears to be an important avenue for future work.

References

Aghion, Philippe and Jean Tirole, “The Management of Innovation,” *Quarterly Journal of*

Economics, 109(4): 1185-1209

Azoulay, Pierre, Toby Stuart, and Yanbo Wang (2013). “Matthew: Effect or Fable”,

Management Science, 60(1): 92-109.

Azoulay, Pierre, Jeffrey L. Furman, Joshua L. Krieger, and Fiona E. Murray (2015).

“Retractions”, *Review of Economics and Statistics*, 97(3).

Azoulay, Pierre, Alessandro Bonati and Joshua L. Krieger (2017). “The Career Effects of

Scandal: Evidence from Scientific Retractions”, *Research Policy* 46(9): 1552-1569..

Bacon, Francis (1620). *Novum Organum*.

Bar-Isaac, Heski (2007). “Something to Prove: Reputation in Teams” *RAND Journal of*

Economics, 38, 495-511.

Becker, Gary S. and Kevin M. Murphy (1992). “The Division of Labor, Coordination Costs, and

Knowledge,” *Quarterly Journal of Economics*, 107 (4), 1137-1160.

- Bikard, Michael, Fiona Murray, and Joshua Gans (2015). “Exploring Trade-offs in the Organization of Scientific Work: Collaboration and Scientific Reward,” *Management Science*, 61(7): 1473–1495.
- Cohen and Bailey (1997). “What Makes Teams Work: Group Effectiveness Research from the Shop Floor to the Executive Suite,” *Journal of Management*, 23(3): 239-290.
- Fang, Ferric, R. Grant Steen, and Arturo Casadevall (2012). “Misconduct accounts for the majority of retracted scientific publications” *PNAS*.
- Furman, Jeffrey L. and Scott Stern (2011). “Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research” *American Economic Review* 101: 1933-1963.
- Furman, Jeffrey L., K. Jensen, and Fiona Murray (2012). “Governing Knowledge in the Scientific Community: Exploring the Role of Retractions in Biomedicine,” *Research Policy* 41 (2).
- Hirsch, J. E. (2005). “An Index to Quantify an Individual’s Scientific Research Output,” *Proceedings of the National Academy of Sciences* 102 (46) 16569-16572.
- Holmstrom, Bengt (1982). “Moral Hazard in Teams,” *Bell Journal of Economics* 324-340.
- Hsu, David H. (2008). “Technology-based Entrepreneurship”, *The Handbook of Technology and Innovation Management*.
- Jones, Benjamin F. (2009). “The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?” *Review of Economic Studies*. 76(1).
- Jones, Benjamin F. (2010). “As Science Evolves, How Can Science Policy?” *NBER Innovation Policy and the Economy* 11.
- Lacetera, Nicola and Lorenzo Zirulia (2011) “The Economics of Scientific Misconduct”, *Journal of Law, Economics, and Organization*, Volume 27, Issue 3, Pages 568–603,

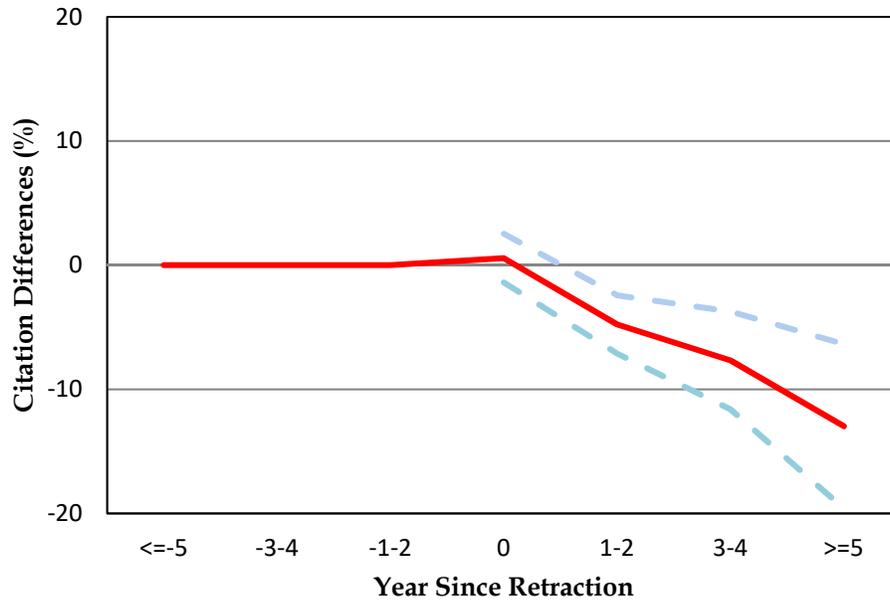
- Lu, Susan Feng, Ginger Jin, Benjamin Jones, Brian Uzzi (2013). “The Retraction Penalty: Evidence from the Web of Science,” *Scientific Reports*, Article Number: 3146.
- Merton, Robert (1968). “The Matthew Effect in Science”, *Science* 159 (3810): 56-63.
- Rao, Hayagreeva (1994). “The Social Construction of Reputation: Certification Contests, Legitimation, and the Survival of Organizations in the American Automobile Industry: 1895-1912,” *Strategic Management Journal*, 15: 29-44.
- Rayo, Luis (2007). “Relational Incentives and Moral Hazard in Teams,” *Review of Economic Studies*, 74(3): 937-963.
- Reagans, Ray., Ezra Zuckman and Bill McEvily (2004) “How to Make the Team: Social Networks vs. Demography as Criteria for Designing Effective Teams”, *Administrative Science Quarterly*, 49(1): 101-133.
- Shapiro, Carl (1983). “Premiums for High Quality Products as Returns to Reputations” *The Quarterly Journal of Economics* 98(4): 659-679.
- Simcoe, Tim and Dave Waguespack (2011). “Status, Quality, and Attention: What’s in a (Missing) Name?” *Management Science* 57, 274-290.
- Smith, Adam (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*.
- Stephan, Paula (2012). *How Economics Shapes Science*, Harvard University Press: Cambridge, MA.
- Wageman, Ruth and George Baker (1997). “Incentives and Cooperation: The Joint Effects of Task and Reward Interdependence on Group Performance,” *Journal of Organizational Behavior*, 18(2): 139-158.

Welbourne, Theresa, David Balikin, and Luis Gomez-Mejia (1995). “Gainsharing and Mutual Monitoring: A Combined Agency-Organizational Justice Interpretation,” *Academy of Management Journal*, 38(3): 881-899.

Woodman, Richard, John Sawyer, and Ricky Griffin (1993). “Toward a Theory of Organizational Creativity,” *Academy of Management Review*, 18(2): 293-321.

Wuchty, Stefan, Benjamin F Jones and Brian Uzzi (2007). “The Increasing Dominance of Teams in the Production of Knowledge” *Science*. 316(5827): 1036-1039.

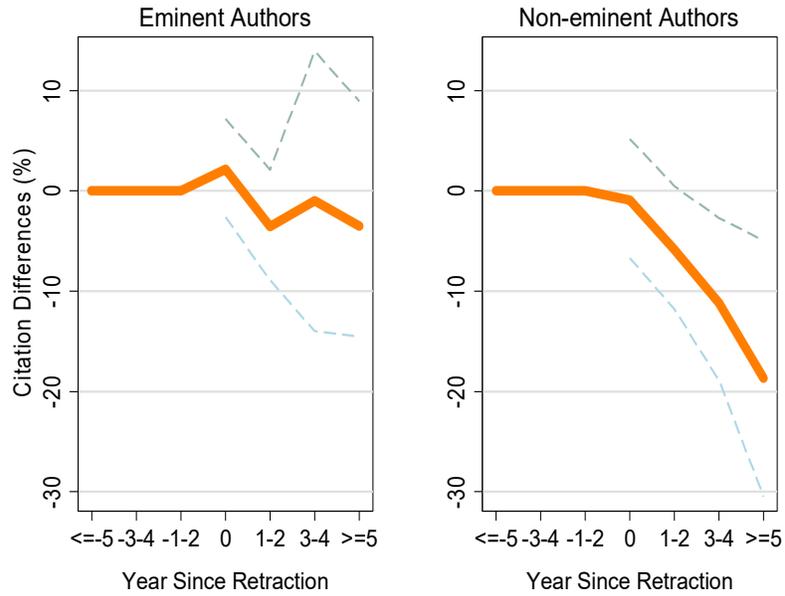
Figure 1: Citations to an Author's Prior Publications, Compared to Control Papers, by Years since Retraction Event



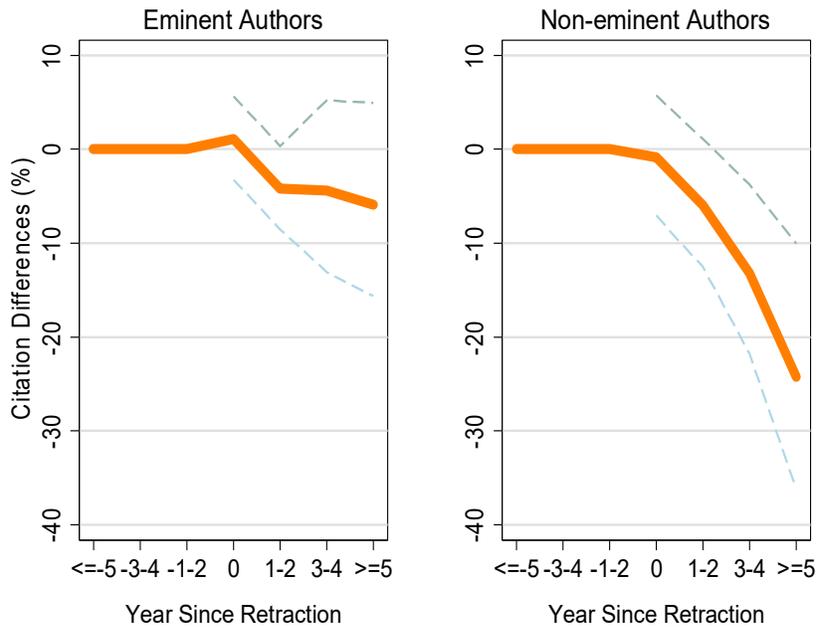
Note: This figure follows Lu et al. (2013) but restricts analysis to retraction events where the retracted paper was team-authored.

Figure 2: Citation Losses by Author Standing

Panel A: Absolute Standing



Panel B: Relative Standing



Notes: In Panel A, authors are divided into two groups based on their absolute standing, where eminent authors are defined as being in the upper 10th percentile by h-index (and non-eminent authors are everyone else). In Panel B, authors are divided according to relative standing within the team, where eminent authors are the individual with the highest h-index (and non-eminent authors are everyone else).

Table 1: Summary Statistics

Panel A: Unit of observation = author, treated only

| Absolute Measures of Standing | Definition | Obs | Mean | | | SD | Min | Max |
|-------------------------------|------------------------------|-----|------|---------|----------|------|-----|-------|
| | | | All | Eminent | Ordinary | | | |
| Prior Publications | total prior papers | 732 | 24 | 136 | 13 | 46 | 1 | 452 |
| Prior Citations | total prior citations | 732 | 1071 | 8209 | 364 | 3570 | 0 | 67946 |
| Prior h-index | prior h-index | 732 | 10 | 44 | 6 | 14 | 0 | 132 |
| Career Age | academic age till retraction | 732 | 10 | 27 | 9 | 9 | 1 | 51 |

Notes: The eminent/ordinary authors are classified by prior h-index. We define an author as an eminent author if his or her prior h-index is among the top 10 percentile and 0 otherwise.

Panel B: Unit of observation = paper, treated only

| | Retracted Papers | Prior Work |
|---------------------------------------|------------------|------------|
| Paper Counts | 276 | 10,209 |
| % Published in 2000s | 86.2% | 45.5% |
| % Published in 1990s | 13.8% | 40.0% |
| % Published in 1980s | 0% | 14.5% |
| Yearly Mean Citation | | |
| Count ^(a) | 3.9 | 3.0 |
| Mean Age Since | | |
| Publication ^(b) | 5.3 | 11.6 |
| Mean Age at Retraction ^(c) | 2.2 | 8.5 |
| Mean Authors per Paper | 5.9 | 5.4 |

Notes: (a) Mean citation rate is the rate in years prior to the retraction event (b) Age since publication is the difference between 2009 (the end of our sample) and the publication year; (c) Age at retraction is the difference between the year of the retraction event and the publication year. Note that control papers, by construction of the matching process, have exactly the same publication year, mean citation counts and dynamics prior to retraction, and age at retraction.

Table 2: Effect of Retraction on Citations to Prior Work, by Absolute Standing of the Author at Time of Retraction

| Absolute Standing of the treated author | Standing Measures | | |
|--|-------------------------|----------------------------|----------------------|
| | Total # of prior papers | Total # of prior citations | H-index |
| | (1) | (2) | (3) |
| Treated*Post($t \geq 1$) | -0.093** (0.039) | -0.101*** (0.034) | -0.114*** (0.040) |
| Author Standing*Treated*Post($t \geq 1$) | 0.040 (0.036) | 0.030** (0.012) | 0.029** (0.015) |
| Author-Paper Fixed Effects | Y | Y | Y |
| Year Since Publication Dummies | Y | Y | Y |
| Observations | 419,239 | 419,239 | 419,239 |
| Number of unique papers | 34,562 | 34,562 | 34,562 |

Notes: Author standing refers to the noted empirical measure of eminence for a treated author in the year prior to retraction, standardized by sample mean and standard deviation. All regressions report coefficients from maximum likelihood

estimation of a Poisson count model, errors clustered by each retraction event. Standard errors in parentheses, *** $p < 0.01$,

** $p < 0.05$, * $p < 0.1$.

Table 3: Effect of Retraction on Citations to Prior Work, by Author Standing Relative to Coauthors at Time of Retraction

| Standing of a treated author relative to the coauthors within the team | Top 1 in | | | Top 2 in | Top 2 in | |
|--|------------|------------|----------|------------|-----------|----------|
| | Top 1 in | Total # of | Top 1 | Total # of | Total # | Top 2 |
| | Total # of | prior | in h- | prior | of prior | in h- |
| | prior work | citations | index | work | citations | index |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated*Post(t>=1) | -0.114** | -0.119*** | 0.119*** | -0.175*** | -0.151*** | 0.154*** |
| | (0.044) | (0.045) | (0.045) | (0.046) | (0.055) | (0.052) |
| Author | | | - | | | - |
| Standing*Treated*Post(t>=1) | 0.065 | 0.074* | 0.072* | 0.121*** | 0.095* | 0.097* |
| | (0.042) | (0.043) | (0.043) | (0.046) | (0.056) | (0.053) |
| Author-Paper Fixed Effects | Y | Y | Y | Y | Y | Y |
| Year Since Publication Dummies | Y | Y | Y | Y | Y | Y |
| Observations | 419,239 | 419,239 | 419,239 | 419,239 | 419,239 | 419,239 |

| | | | | | | |
|-------------------------|--------|--------|--------|--------|--------|--------|
| Number of unique papers | 34,562 | 34,562 | 34,562 | 34,562 | 34,562 | 34,562 |
|-------------------------|--------|--------|--------|--------|--------|--------|

Notes: See notes for Table 2. The difference here is that author standing is now a dummy for whether a treated author had the highest standing (“Top 1”) within the team or is among the two individuals with highest standing (“Top 2”) in the team.

Table 4: Effect of Retraction on Citations to Prior Work, by Own and Coauthor Standing

| Team configurations in the retracted paper | All Authors | | |
|---|--------------------------|--------------------|---------------------|
| | Total # of | | |
| | Total # of prior work | prior citations | Prior h-index |
| | (1) | (2) | (3) |
| Treated*Post($t \geq 1$) | -0.016 (0.037) | -0.059 (0.076) | 0.009 (0.029) |
| Self is eminent and Co-author is ordinary *Treated*Post($t \geq 1$) | -0.029 (0.061) | -0.002 (0.093) | -0.056 (0.060) |
| Self is ordinary and Co-author is eminent *Treated*Post($t \geq 1$) | -0.123* (0.067) | -0.126 (0.097) | -0.165** (0.082) |
| Self is ordinary and Co-author is ordinary *Treated*Post($t \geq 1$) | -0.063 (0.064) | 0.009 (0.089) | -0.101* (0.057) |

| | | | |
|--------------------------------|---------|---------|---------|
| Author-Paper Fixed Effects | Y | Y | Y |
| Year Since Publication Dummies | Y | Y | Y |
| Observations | 419,239 | 419,239 | 419,239 |
| Number of papers | 34,562 | 34,562 | 34,562 |

Notes: We classified the authors into four groups using dummy variables indicating whether (1) own standing is ordinary and the highest-standing coauthor is ordinary, (2) own standing is ordinary but a coauthor is eminent, (3) own standing is eminent and the highest-standing coauthor is ordinary, and (4) own standing and a coauthor are both eminent (the omitted category in the regression). Author standing is measured in the year prior to retraction. All regressions report coefficients from maximum likelihood estimation of a Poisson count model, errors clustered by each retraction event. All regressions include all one-way and two-way interactions terms; we do not report those coefficients for brevity. Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Effect of Retraction on Citations to Prior Work, Accounting for Rookie Coauthors

| Measure of Coauthor Status | h-index | | |
|-------------------------------------|----------------------|-------------------|----------------------|
| | Full | | |
| | Sample | Eminent | Ordinary |
| | (1) | (2) | (3) |
| Treated*Post(t>=1) | -0.121*** (0.038) | -0.045 (0.038) | -0.117*** (0.041) |
| Author | | | |
| Status*Treated*Post(t>=1) | 0.026** (0.013) | | |
| % No | | | |
| Prior*Treated*Post(t>=1) | 0.073*** (0.025) | 0.042 (0.037) | 0.106*** (0.033) |
| Author-Paper Fixed Effects | Y | Y | Y |

Table 5: Effect of Retraction on Citations to Prior Work, Accounting for Rookie Coauthors

| Measure of Coauthor Status | h-index | | |
|-------------------------------------|----------------------|-------------------|----------------------|
| | Full | | |
| | Sample | Eminent | Ordinary |
| | (1) | (2) | (3) |
| Treated*Post(t>=1) | -0.121*** (0.038) | -0.045 (0.038) | -0.117*** (0.041) |
| Author | | | |
| Status*Treated*Post(t>=1) | 0.026** (0.013) | | |
| % No | | | |
| Prior*Treated*Post(t>=1) | 0.073*** (0.025) | 0.042 (0.037) | 0.106*** (0.033) |
| Author-Paper Fixed Effects | Y | Y | Y |