NBER WORKING PAPER SERIES

RETRACTIONS

Pierre Azoulay
Jeffrey L. Furman
Joshua L. Krieger
Fiona E. Murray

Working Paper 18499
http://www.nber.org/papers/w18499

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2012

We gratefully acknowledge the financial support of the National Science Foundation through its SciSIP Program (Awards SBE-0738142, SBE-0965364, and SBE-0738394). We thank Heidi Williams and participants in the MIT Labor Lunch and NBER Productivity Lunch for insightful comments. Vivienne Groves and Mikka Rokkanen provided additional research assistance. The project would not have been possible without Andrew Stellman's extraordinary programming skills (http://www.stellman-greene.com/). The usual disclaimer applies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Retractions
Pierre Azoulay, Jeffrey L. Furman, Joshua L. Krieger, and Fiona E. Murray
NBER Working Paper No. 18499
October 2012
JEL No. O33

ABSTRACT

To what extent does "false science" impact the rate and direction of scientific change? We examine the impact of more than 1,100 scientific retractions on the citation trajectories of articles that are close neighbors of retracted articles in intellectual space but were published prior to the retraction event. Our results indicate that following retraction and relative to carefully selected controls, related articles experience a lasting five to ten percent decline in the rate at which they are cited. We probe the mechanisms that might underlie these negative spillovers over intellectual space. One view holds that adjacent fields atrophy post-retraction because the shoulders they offer to follow-on researchers have been proven to be shaky or absent. An alternative view holds that scientists avoid the "infected" fields lest their own status suffers through mere association. Two pieces of evidence are consistent with the latter view. First, for-profit citers are much less responsive to the retraction event than are academic citers. Second, the penalty suffered by related articles is much more severe when the associated retracted article includes fraud or misconduct, relative to cases where the retraction occurred because of honest mistakes.

Pierre Azoulay
MIT Sloan School of Management
100 Main Street, E62-482
Cambridge, MA 02142
and NBER
pazoulay@mit.edu

Jeffrey L. Furman
Boston University - SMG
595 Commonwealth Ave - #653a
Boston, MA 02215
and NBER
furman@bu.edu

Joshua L. Krieger
Massachusetts Institute of Technology
jkrieger@MIT.EDU

Fiona E. Murray
MIT Sloan School of Management
100 Main Street, e62-470
Cambridge, MA 02142
fmurray@mit.edu

1 Introduction

In 2005, South Korean scientist Woo-Suk Hwang and his colleagues published an article in *Science* claiming they had isolated embryonic stem cells from a cloned human embryo via nuclear transfer (Hwang et al. 2005). Immediately following publication, scientists around the world took time and resources to replicate and continue this line of enquiry, thus building on the exciting (albeit controversial) field of human embryonic stem cell production using cloning. Less than a year later, the paper was formally retracted from the literature amidst claims of error and later findings of fraud and embezzlement. In the aftermath, the government of South Korea curtailed investment in stem cell research for five years and, globally, scientists no longer built on the fraudulent Hwang paper; some researchers abandoned the field altogether while others pursued adjacent research lines that built on firmer foundations (Furman, Murray, and Stern 2012). It took several years before researchers started to explore some of the novel hESC production methods proposed in the controversial paper, particularly parthenogenesis. In late 2007 Harvard researchers published definitive results showing that some of the (lesser) claims made by the Korean team were actually useful insights into other methods of hESC production (Kim et al. 2007). Until this new research, the field had been stifled because the retracted paper *"sent a lot of scientists on a wild goose chase and down false paths,"* in the words of a stem cell researcher, Robert Lanza, quoted by the Associated Press (2005).

This dramatic incident illustrates the central questions of our paper: To what extent does "false science" impact the rate and direction of scientific research? To address this question we examine the impact of retractions — publications in the academic literature that are withdrawn by authors or editors — on cumulative knowledge production along retracted research lines and, more importantly, in intellectually related research areas. We do so using a novel approach to characterize the intellectual scope of research fields and their proximity to specific (retracted) papers. Our analysis is timely because "false science," a term we coin to cover a broad range of phenomena, from mistakes to plagiarism to difficulties in replication to systematic fraud, has received considerable recent attention (Fang, Steen, and Casadevall 2012; Lacetera and Zirulia 2009; Pozzi and David 2007). While funders, science policy makers and journal editors have opined on the effectiveness (or lack thereof) of the current retraction system, they do not evaluate the types of evidence that we provide to enable them to consider alternatives to the status quo. For scholars of scientific and technological change, retractions provide an unusual lens to deepen our understanding of the dynamics of cumulative knowledge production, particularly as we seek to move beyond analyses of the determinants of the rate of inventive activity towards an understanding of the factors shaping the choice of research direction.

The spillover effects of retractions on the evolution of research fields is particularly important given the broader welfare implications that arise from scientists shifting their position in "intellectual space" (Aghion et al. 2008, Acemoglu 2012). However, evidence is currently limited. As a starting point, systematic data on journal article retractions shows a strong upward trend in their frequency, but as in the case of criminal activity, the underlying magnitude of scientific mistakes and misdeeds remains poorly established (Martinson, Anderson, and de Vries 2005). In addition, a recent analysis shows that the majority of retractions are caused by misconduct including fraud (Fang et al. 2012). Most recently, Furman, Jensen, and Murray (2012) provide evidence that retraction notices are effective in alerting follow-on researchers to the shaky foundations of a particular paper: citations to retracted papers decline by over 60% in the post-retraction period relative to carefully matched controls. Their analysis, however, focuses on the fate of the retracted papers themselves, not whether and to what extent retractions influence the evolution of adjacent research areas. It also does not distinguish between different types of false science associated with retracted events, although this heterogeneity is of primary importance since the strength of the shoulders provided by retracted articles to follow-on researchers can vary widely. Thus, the challenge for our paper is to elucidate the impact of different types of retractions on retracted research lines and the magnitude of spillovers to research lines in proximate intellectual space.

Our conceptual approach follows Acemoglu (2012), Aghion and co-authors (2009), and others in understanding research as arising through a cumulative process along and across research lines that can be traced out empirically through citations from one publication to another (e.g., Murray and Stern 2007). This approach is grounded in the assumption that knowledge accumulates as researchers take the knowledge in a particular publication and use it as a stepping stone for their follow-on investigations (Mokyr 2004). Although it is a commonplace insight that the process of knowledge accumulation unfolds within an intellectual space (e.g., Hull 1988), it has proved surprisingly difficult for social scientists to gain empirical traction on this concept (see Azoulay, Graff Zivin, and Wang [2010] for a rare exception). We conceptualize retraction events as "shocks" to the structure of the intellectual space around the retracted papers, and implement a procedure to delineate the boundaries of this space in terms of related publications in a way that is scalable and transparent, and with scant reliance on human judgement. We are then interested in studying whether researchers increase or decrease their reliance on related papers following the retraction event. We differentiate this cumulative response across three types of retractions: papers whose results have been clearly shown to be invalid and should not be used as the basis of future research (which we label "absent shoulders" papers), papers where retraction creates doubt about — but does not clearly nullify — the value of the content for follow-on research ("shaky shoulders"), and papers where retraction does not cast aspersions on the validity of the findings ("strong shoulders").

A priori, retraction events could be thought to have two countervailing effects on the intensity of follow-on research direction. On the one hand, researchers may simply substitute away from the specific retracted paper and increase their reliance on other research in the same intellectual field, effectively increasing the prominence of the unretracted papers in that same field. On the other hand, researchers (and/or their funders) may substitute away from the related research line, and not simply from the retracted paper, if broader negative spillover effects are salient. Our results clearly show that the latter effect dominates.

Using the *PubMed Related Citations Algorithm* [PMRA] to delineate the fields surrounding over 1,100 retracted articles in the biomedical research literature, we show that 60,000 related articles experience on average a 6% decline in their citation rate following retraction, relative to the background citation rates for 110,000 control articles that appeared in the same journals and time periods (an empirical approach to controlling for citation trajectories that has been used effectively in prior work on the effect of scientific institutions and governance, e.g., Furman and Stern [2011]). Moreover, this effect is entirely driven by the fate of articles related to retractions with shaky or absent shoulders: There is no broad impact on the field when the retraction occurred because of plagiarism or publisher error. In contrast, mistakes, fake data, and difficulties arising in replication attempts have negative spillover effects on intellectual neighbors. Though the collateral damage (measured in lost citations) is about ten times smaller than the direct penalty suffered by the associated retracted article, we find the effect to be persistent and increasing in magnitude over time.

We then examine the proximate causes of this citation decline. We find evidence that publication rates in the fields affected by a retraction markedly decrease following retraction, relative to control fields. Similarly, we find that NIH funding in these fields declines in an even sharper fashion. What are the mechanisms that lie behind such strong behavioral responses from the scientific community? We attempt to distinguish a *learning* interpretation from one based on *status* concerns. On the one hand, we might simply be observing that retraction events enable scientists to discover that a particular field offers fewer prospects of important findings than was previously believed, leaving them to substitute away from that field onto lines of research that are not directly adjacent to that of the retracted knowledge. Alternatively, scientists in the affected fields might believe that their reputation will be besmirched if they tie their scientific agenda too tightly to a field that has been "contaminated" by a retraction. Status concerns of this kind would just as surely drive away previous (or potential) participants in the field, but these shifts would this time be construed as constituting under-investment in the affected areas from a welfare standpoint.

We find suggestive evidence that the status interpretation accounts for at least part of the damage suffered by retraction-afflicted fields. First, we document that, even in the set of articles related to retractions offering entirely absent shoulders to follow-on researchers, intent matters in modulating the observed citation responses: the penalty suffered by related articles is much more severe when the associated source article was retracted because of fraud or misconduct, relative to cases where the retraction occurred because of "honest mistakes." Second, starting from the premise that status considerations are less likely to drive the citing behavior of scientists employed in industry, relative to that of academic citers, we show that the former are much less responsive to the retraction event than the latter. While a learning story argues in favor of strengthening the retraction system in its current incarnation, the evidence for a status explanation suggests that researchers are overly reactive to retraction notices and challenges us to propose alternative and more effective governance approaches to signal scientific mistakes and "false science."

In the remainder of the paper we examine the institutional context for retractions as the central approach to governing scientific mistakes and misconduct and lay out our broad empirical strategy. We then turn to data, methods and a detailed presentation of our results. We conclude by outlining the implications of our findings for the design of governance mechanisms that could help the "Republic of Science" better cope with the specific challenges posed by the existence of false scientific knowledge.

2 Institutional Context and Empirical Design

Since the seminal work of Arrow, Nelson, and others, economists have recognized the central importance of both the rate and direction of inventive activity in the economy. Nonetheless, traditional analyses (in part as a result of data limitations) have emphasized either the rate of inventive activity or the narrow linear flow from basic to applied research (see for example Aghion et al. 2008). In contrast, recent theoretical work has given a more central role to the concept of broad research lines that constitute a richer intellectual space within which researchers build new knowledge. For instance, Acemoglu (2012) models the costs and benefits of diverse research approaches for the development of alternative technologies; Aghion et al. (2009) develop a model in which early-stage research is constructed from a variety of parallel lines representing a different scientific field or approach. In the limit, the concatenation of these research lines can be thought of as an intellectual continuum across which scientists move as they select the direction of their inquiries.

Whether pursuing a single research line or moving from one intellectual field to another, knowledge accumulation — the process by which follow-on researchers build on ideas developed by prior generations of researchers — is of central importance to scientific progress (Mokyr, 2004). This cumulative process is typically referred to as "standing on the shoulders of giants" in deference to Sir Isaac Newton, but is conceptualized more prosaically as the way in which researchers in one generation learn from and build upon prior research papers. In doing so, each generation moves along or across lines and uses prior contributions to a line as the foundation for their own subsequent discoveries, thus rendering the process of knowledge generation more effective, particularly as the basis for economic growth (Romer 1994).

A variety of institutions and incentives have arisen to support this cumulative process (Dasgupta and David 1994, David 2008). These include the norms of open science (David, 2008), material repositories (Furman and Stern 2011), patent disclosure and licensing policies (Murray and Stern 2007, Aghion et al. 2009, Murray 2010), and information technology (Agrawal and Goldfarb 2008, Ding et al. 2010). To date, however, past scholarship has focused on understanding the role of openness in facilitating knowledge accumulation to the exclusion of other institutional determinants of knowledge production.

Recent instances of large-scale scientific fraud and mistakes have brought to the fore concerns regarding the effectiveness of the institutions that govern the cumulative production of knowledge. The current interest in the costs of scientific errors and fraud reflects an occasional popular interest, usually inspired by episodic, high-profile cases of fraud (cf. Babbage 1830; Weiner 1955; Broad and Wade 1983; LaFollette 1992). In addition, these unfortunate events highlight the economic impact of false science — in terms of wasted government resources and damage to the public image of science — leading high profile journals to editorialize on the issue and attracting the attention of the science-monitoring community, including specialized blogs such as *RetractionWatch*.[1] For economists of science and innovation, debates over false science turn on its impact on cumulative scientific progress: If researchers are unwittingly building on false or shaky research results, their effort is wasted and scientific progress stifled. In spite of this limelight and the importance of the issue, there is only scant descriptive evidence regarding the role of institutions that support the fidelity

[1] http://retractionwatch.wordpress.com/

of scientific knowledge (ORI 2007; Pozzi and David 2007; Lacetera and Zirulia, 2008) and even less exploration of their effectiveness (Furman, Jensen and Murray 2012; Lu, Ginger, Jones and Uzzi 2012). To our knowledge, this study is the first to document systematically how false science shapes the direction of cumulative scientific progress.

2.1 Institutional Context

Very few practices exist to identify and signal research misconduct or error. In the United States, key public funders have created an Office of Research Integrity (ORI) to investigate allegations of fraud or misconduct (Pozzi and David 2007). More broadly applicable is the system of retractions used by journals themselves to alert readers when a research publication is stricken from the scientific literature. Retractions can be made by all or some of the authors of a publication, or by the journal's editor, directly or at the request of the authors' employer. These events can occur for a variety of reasons, as we describe below. Retraction events remain very rare, with unconditional odds of retraction standing at about one per ten thousand, regardless of the data source used to calculate these odds (see Lu et al. 2012 for tabulations stemming from Thomson-Reuters' *Web of Science* database). Figure 1 documents secular increases in the incidence of retractions in PubMed, where this incidence is measured both as a raw frequency and as a proportion relative to total size of the PubMed universe.

While this paper is not focused on the determinants of false science but rather its impact, it is worth noting that the rise in instances of false science (or at least the increase in its documentation via retraction notices) may be linked to a range of factors including the increasingly complex and collaborative organization of the scientific enterprise (Wutchy, Jones and Uzzi 2007) and the growing competition for resources in science.[2] Nonetheless, as a matter of institutional design, the system of retractions treads a treacherous middle ground in managing the integrity of scientific knowledge. At one end of the spectrum, scientific societies and journals could make significant investments in replicating and verifying all

[2]Lacetera and Zirulia (2008) note that competition has ambiguous effects on the incidence of scientific misconduct since scientists can also gain prominence by *detecting* instances of false science.

studies prior to publication, while at the other end, a knowledge registration system with no filtering mechanism could require researchers to expend considerable time and energy on replication and validation. The actual system in existence today relies heavily upon peer-review but provides only limited *ex ante* insurance that published knowledge is of high fidelity. As a result, reputational incentives play an essential role to ensure the integrity of the scientific enterprise (Merton 1973).

Retraction notices, however, are idiosyncratic and vary widely in the amount of information they provide, ranging from a one line sentence to a more elaborated statement of the rationale behind the retraction event. As a result, understanding their effectiveness in policing the scientific community is of central importance to the process of cumulative knowledge production and the allocation of resources, human and financial, within and across scientific fields. A poorly functioning system may be extremely wasteful and distort the location choices of scientists in intellectual space. Our paper explores a number of questions that allow us to provide better and more systematic evidence regarding the effectiveness of retraction practices for the forward movement of science; as a byproduct, our results point to heretofore unexplored avenues to improve upon the current system.

2.2 Empirical Design

Our core research questions require that we overcome two separate data collection challenges. First, we must develop a coding scheme to parse the underlying reasons behind each of the (over 1,100) retractions that serve as "shocks" to the range of intellectual fields we examine. Our coding must also account for the degree to which the retraction leaves intact vs. knocks down the shoulders upon which follow-on researchers may build. Second, we need a credible approach to systematically identify other journal articles that lie in close proximity in intellectual space to the retracted articles as well as a metric to measure their degree of proximity.

Categorizing retraction events. To meet the first challenge, we have developed a detailed taxonomy of retracted articles to capture the differences in the meaning of the retraction events for follow-on researchers, as described in Appendix I. In a second step and taking inspiration from Newton's aphorism, we then systematically assigned the 1,104 retractions in our sample to three mutually exclusive buckets denoted, "Strong Shoulders," "Shaky Shoulders," and "Absent Shoulders," respectively:

- "Strong Shoulders" means the retraction does not in any way degrade the validity of the paper's analysis or claims. This may happen in instances where a publisher mistakenly printed an article twice, when authors published an ostensibly valid study without securing approval to publish the (unchallenged) data , or when an institutional dispute over the ownership of research materials arose.
- "Shaky Shoulders" means that the validity of claims are uncertain or that a fraction of the results are invalid. The typical use of this category concerns instances where results could not be replicated, among other reasons.
- "Absent Shoulders" is the appropriate code for retractions associated with fraudulent results, as well as in cases where a mistake in experimental procedure irretrievably invalidates the paper's results.

In addition, we differentiate between retractions for which the authors intentionally attempted to subvert the scientific truth and those for which the article needed to be retracted because of a genuine error with no indication of foul play. We therefore examined retractions to develop a code for different levels of intentional deception.[3] We use "No Sign of Intentional Deception" to code cases where the authors did not intend to deceive, such as instances of miscommunication, contamination of research materials, or coding error. "Uncertain Intent" applies where fraud is not firmly established, but negligence or unsubstantiated claims raise questions about the authors' motives. The "Intentional Deception" code is reserved for cases where falsification, misconduct, or willful acts of plagiarism and self-plagiarism appear to have occurred and were verified by author admissions or independent reviews of misconduct.

[3]Deception might involve the paper's factual claims (results, materials, or methods), its attribution of scholarly credit through authorship and citations, or the originality of the work.

Delineating research fields. To delineate the boundaries of the research fields affected by retracted articles, we develop an approach based on topic similarity as inferred by the overlap in keywords between each retracted articles and the rest of the (unretracted) scientific literature. Specifically, we use the PubMed Related Citations Algorithm (PMRA) which relies heavily on MeSH keywords; a controlled vocabulary maintained by the National Library of Medicine that provides a very fine-grained partition of the intellectual space spanned by the biomedical research literature. Importantly for our purposes, MeSH keywords are assigned to each scientific publication by professional indexers and not by the authors themselves; the assignment is made without any reference to the literature cited in the article. We then use the "Related Articles" function in PubMed to harvest journal articles that are proximate to the retracted articles, implicitly defining a scientific field as the set of articles whose MeSH keywords overlap with those tagging the ultimately retracted article. As a byproduct, PMRA provides us with both an ordinal and a cardinal dyadic measure of intellectual proximity between each related article and its associated retraction. For the purposes of our analysis, we only consider related articles published prior to the retraction date. We distinguish those published prior to the retracted article and those published in the window between the retracted article's publication date and the retraction event itself. Further, we also exclude related articles with any co-authors in common with the retracted article in order to strip bare our measure of intellectual proximity from any "associational baggage" stemming from collaboration linkages. Finally, we build a set of control articles by selecting the "nearest neighbors"of the related articles, i.e., the articles appearing immediately before or immediately after in the same journal and issue, as in Furman and Stern (2011) and Furman et al. (2012).

Empirical strategy. Together, these advances allow us to estimate the causal impact of retraction events on the vitality of scientific fields. We start by examining the impact of a retraction on the citations to the retracted papers themselves, in a reprise of the earlier study by Furman et al. (2012), but using a more complete sample and carefully differentiating the effect across different types of retractions. Indeed, to the extent that retractions are highly differentiated in the information they impart to follow-on researchers regarding the strength of the shoulders upon which they stand, we would anticipate that this type of variation would powerfully moderate the impact on follow-on citations. We then perform the main exercise of the paper by examining the impact of retraction events on citations to related articles and their controls in a simple difference-in-differences framework. Again, we separately estimate the impact of different types of retractions.

Lastly, we explore the mechanisms that may be at play, focusing on the set of retractions providing "absent shoulders" to follow-on researchers pursuing related intellectual space. We do so by exploring citations to related articles made by authors in academia versus industry, on the assumption that status effects (in comparison to learning effects) are more likely to influence the citing behavior of academic researchers who are more reliant on a complex credit structure than their private-sector counterparts. We also develop an analysis of the rate of production of related articles (rather than citation to these related articles) in the pre- and post-retraction period. Similarly, mapping related articles to NIH funding, we explore how resources devoted to scientific fields are influenced by retractions, comparing again to control fields. Overall, this empirical design advances our ability to examine issues related to the direction of research across scientific fields, and provides a nuanced understanding of the role of retractions in the process of cumulative knowledge production.

3 Data and Sample Construction

This section details the construction of our multilevel, panel dataset.

3.1 Retracted Articles

We begin by extracting from PubMed, the public-access database which indexes the life sciences literature, all original journal articles that were subsequently retracted, provided that these articles were published in 2007 or earlier, and retracted in 2009 at the latest.

After purging from the list a few odd observations,[4] we are left with a sample of 1,104 articles.[5] As detailed in Appendix I, we develop an exhaustive category scheme to code the reasons that explain the retraction event. These reasons are tabulated in Table 1.[6] In our next step, we classify each retraction into one of three categories that denote whether the results contained in the source article can be relied upon for follow-on research. The "strong shoulders" subsample comprises 202 articles retracted for reasons that do not cast any aspersion on the validity of the results contained therein. In contrast, we classify 589 retractions (53.4%) as providing "absent shoulders" for follow-on scientists to stand on, often because of fraudulent data or other types of misconduct. Finally, the "shaky shoulders" category (289 events or 26.2% of the cases) groups those retraction events for which the validity of the results remains shrouded in uncertainty.

Most of our analyses focus on the 589 observations belonging to the "absent shoulders" subsample (Table 2). The papers in this subsample were published between 1973 and 2007, and took an average time of three years to be retracted, though many of the more recent articles were retracted within one year — perhaps because of a higher probability of detection since the dawn of the electronic publishing era. Though this subsample is dominated by instances of fraud or other types of misconduct, 31% of the events appear to be the results of honest mistakes on the part of the investigators involved, with a further 8% for which it is unclear whether the scientists actively subverted the scientific process in the course of performing the research and reporting its results.[7]

[4]These include an article retracted and subsequently unretracted, an erratum that was retracted because of disagreement within the authorship team about whether the original article indeed contained an error, along with a few others.

[5]In comparison, Lu et al. (2012) extract 1,423 retraction events from Thomson Reuters' *Web of Science* over the same period. The *Web of Science* covers a wider cross-section of scientific fields (including the social sciences and engineering), but has shallower coverage than PubMed in the life sciences. By combining the events corresponding to life sciences journals as well as multidisciplinary journals — such as *Science*, *PNAS*, or *Nature* — it appears that the life sciences account for between 60% and 70% of the total number of retractions in the Lu et al. sample.

[6]Despite extensive efforts, we were unable to locate a retraction notice in 24 (2.17%) cases.

[7]This represents an inversion of the relative prevalence of fraud and mistakes, compared to an earlier analysis performed by Nath et al. (2006), but is in line with the recent results reported by Fang et al. (2012).

Regardless of intent, however, it would be a mistake to consider each observation as completely independent from all the others in the sample. Close to sixty percent of the observations can be grouped into "cases" involving more than one retraction event, for example because the same rogue investigator committed fraud in multiple papers, or because the same contaminated research materials were used in multiple published articles. Figure 2 displays the histogram of the distribution of retraction events by retraction case ($N = 334$). The case identifier will play an important role in the econometric analysis since all of our results will report standard errors clustered at the case level of analysis.

3.2 Related Articles

Traditionally, it has been very difficult to assign to individual scientists, or articles, a fixed address in "idea space," and this data constraint explains in large part why bibliometric analyses typically focus on the determinants on the *rate* of scientific progress rather than its *direction*. The empirical exercise in this paper hinges crucially on the ability to relax this constraint in a way that is consistent across retraction events and requires little, if any, human judgement. This challenge is met here by the systematic use of the PubMed Related Citations Algorithm [PMRA], a probabilistic, topic-based model for content similarity that underlies the "related articles" search feature in PubMed.

This database feature is designed to aid a typical user search through the literature by presenting a set of records topically related to any article returned by a PubMed search query.[8] To assess the degree of intellectual similarity between any two PubMed records, PMRA relies crucially on MeSH keywords. MeSH is the National Library of Medicine's [NLM] controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. There are 26,581 descriptors in the 2012 MeSH edition (new terms are added to the dictionary as scientific advances are made). Almost every publication in PubMed is tagged with a set of MeSH terms (between 1 and 103 in the current edition of PubMed, with both the mean

[8] Lin and Wilbur (2007) report that one fifth of "non-trivial" browser sessions in PubMed involve at least one invocation of PMRA.

and median approximately equal to 11). NLM's professional indexers are trained to select indexing terms from MeSH according to a specific protocol, and consider each article in the context of the entire collection (Bachrach and Charen 1978; Névéol et al. 2010). What is key for our purposes is that the subjectivity inherent in any indexing task is confined to the MeSH term assignment process, which occurs upstream of the retraction event and does not involve the articles' authors.

Using the MeSH keywords as input, PMRA essentially defines a distance concept in idea space such that the proximity between a source article and any other PubMed-indexed publication can be assessed. The algorithm focuses on the smallest neighborhood in this space that includes 100 related records.[9] Given our set of source articles, we delineate the scientific fields to which they belong by focusing on the set of articles returned by PMRA that satisfy five additional constraints: (i) they are original articles (as opposed to editorials, comments, reviews, etc.); (ii) they were published up to the year that precedes the calendar year of the underlying retraction event; (iii) they appear in journals indexed by the Web of Science (so that follow-on citation information can be collected); (iv) they do not share any author with the source, and (v) they are cited at least once by another article indexed by the Web of Science in the period between their publication year and 2011. Figure 2 runs through a specific example in the sample to illustrate the use of PMRA.[10] Appendix II illustrates through an example how PMRA processes MeSH keyword information to delineate the boundaries of research fields.

For the set of 589 retractions with absent shoulders, the final dataset comprises 32,699 related articles that can be ordered by relatedness using both an ordinal measure (the rank

[9] However, the algorithm embodies a transitivity rule as well as a minimum distance cutoff rule, such that the effective number of related articles returned by PMRA varies between 4 and 2,642 in the larger sample of 1,104 retractions, with a mean of 172 records and a median of 121.

[10] To facilitate the harvesting of PubMed-related records on a large scale, we have developed an open-source software tool that queries PubMed and PMRA and stores the retrieved data in a MySQL database. The software is available for download at `http://www.stellman-greene.com/FindRelated/`.

returned by PMRA) as well as a cardinal measure which we normalize such that a score of 100% corresponds to the first "non-trivial" related record.[11]

As a result of these computational and design choices, the boundaries of the fields we delineate are derived from semantic linkages to the exclusion of other considerations such as backward and forward citation relationships, or coauthorships. Judgement and subjectivity is confined to the initial indexing task which assigns keywords to individual articles. The individuals performing these tasks are trained in a consistent way, draw the keywords from a controlled vocabulary which evolves only slowly over time, and do not have any incentives to "window-dress" the articles they index with terms currently in vogue in order to curry attention from referees, editors, or members of funding panels. Of course, the cost of this approach is that it may result in boundaries between fields that might only imperfectly dovetail with the contours of the scientific communities with which the authors in our sample would self-identify. The main benefit, however, is that it makes it sensible to use citation information to evaluate whether the narrow fields around each retracted article atrophy or expand following each retraction event.

3.3 Identification Strategy and Nearest-Neighbor Controls

A natural starting point to identify the spillovers of retraction events on their associated fields is to examine changes in citations received by the set of related articles after the retraction, relative to before, using a simple related article fixed effects specification. Since the retraction effect is mechanically correlated with the passage of time as well as with a paper's vintage, our specifications must include age and calendar year effects, as is the norm in empirical studies of scientific productivity (Levin and Stephan 1991). In this framework, the control group that pins down the counterfactual age and calendar time effects for articles related to a current retraction is comprised of other related articles whose associated retraction occurred in earlier periods or will occur in future periods. This approach may be problematic in

[11] A source article is always trivially related to itself. The relatedness measures are based on the raw data returned by PMRA, and ignore the filters applied to generate the final analysis dataset, e.g., eliminating reviews, etc.

our setting. First, related articles observed after their associated retraction event are not appropriate controls if the event affected the trend in the citation rate; Second, the fields from which retractions are drawn might not represent a random cross-section of all scientific fields, but rather might be subject to idiosyncratic life cycle patterns, with their productive potential first increasing over time, eventually peaking, and thereafter slowly declining. If this is the case, fixed effects will overestimate the true effect of the retraction effect, at least if we rely on articles treated in earlier or later periods as as an "implicit" control group.

To mitigate these threats to identification, our preferred empirical strategy relies on the selection of matched controls for each related — i.e., "treated" — article. In concrete terms, we select as controls for each related article their "nearest neighbors" in the same journal, volume and issue, i.e., the two articles that immediately precede and follow the treated article. When the related article is first or last in the particular issue of the journal considered, we select a single control. The final dataset corresponding to the "Absent Shoulders" subsample comprises 65,461 such controls.

One potential concern with this control group is that its members may also be affected by the retraction treatment, since they are drawn from the same set of journals as the related articles. In what follows, we will ignore this threat to identification for three separate reasons. First, the fields identified by PMRA are extremely thin slices of intellectual space, and their boundaries do not depend on journal or date of publication information (see Appendix II). Second, in the extremely rare cases in which one of these nearest neighbor controls also happens to be related to a retraction through the algorithm, we select instead the article that is "twice removed" in the table of contents from the focal related article. Finally, as can be observed in Table 3, the rate at which the controls cite the retraction with which they are indirectly associated is almost two orders of magnitude smaller than the rate of citation that links the retractions with the "treated" (i.e., related) articles.

Citation data. PubMed does not contain citation data but we were able to retrieve this information from the *Web of Science* (up to the end of 2011) using a perl script. We further process these data to make them amenable to statistical analysis. First, we eliminate all self-citations, where self-citation is inferred by overlap between any of the cited authors with any of the citing authors (an author name is the combination formed by the last name and the first initial for the purpose of this filter). Second, we parse the citing article data to distinguish between the institutional affiliations of citers, in particular by flagging the citing articles for which at least one of the addresses recorded by the *Web of Science* is a corporate address, which we infer from the presence of abbreviations such as Inc, Corp, GmbH, Ltd, etc. We then aggregate this information at the cited article-year level of analysis. In other words, we can decompose the total number of citations flowing to individual articles at a given point in time into a "private" and a "public" set, where public citations should be understood as stemming from academic scientists, broadly construed (this will also include scientists employed in the public sector as well as those employed by non-profit research institutes).

Descriptive Statistics. Table 3 provides basic information about the matched sample. By construction, control and treated articles are matched on year of publication and journal, and they appear to match very closely on the length of the authorship roster. Because in many cases, retraction occurs relatively quickly after publication, only 30% of the related articles in the data are published after the publication of the source article, and only 8% of these articles cite the soon-to-be-retracted source. Conversely, only 6.1% of the articles at risk of being cited by the source (because they were published before its publication) are in fact cited by it.

Table 3 indicates that related articles and their nearest neighbors differ slightly in the total number of citations received at baseline (the calendar year preceding the retraction event), with related articles having received 1.8 citations more on average than the controls. Figure 4 compares the distributions of cumulative baseline citations for control and related articles, respectively. The controls appear to be slightly more likely to have received zero or one citation at baseline. This is not necessarily surprising, if, as mentioned above, articles related to retractions are drawn from fields that draw more attention from the scientific community in the years leading up to the retraction event. Nonetheless, these small differences in the *level* of citations at baseline could prove problematic for our identification strategy if they translate into preexisting *trends* in citations for treated, relative to control articles in the pre-retraction period. We will carefully document below that such pre-trends are extremely small in magnitude and undetectable from a statistical standpoint, buttressing the core assumption that underlies our empirical strategy.

3.4 Field-level Analyses

To examine the proximate causes of the spillover effects of retractions on their fields, we study whether patterns of entry into these fields, or the funding that accrues to active researchers in these same fields, is altered by the retraction event. To do so, we create a second dataset that collapses the related article-level data onto a retracted article-level panel dataset.

As previously, we view scientific fields as isomorphic to the set of articles related (through PMRA) to a given source article. In contrast to the previous section, however, we make use of the related articles published after a retraction event (as well as before). A "field" is born in the year during which the oldest related article was published. Starting from the set of 589 retractions in the "absent shoulders" subsample, we eliminate 24 observations for which this oldest related article is "too young" — it appeared less than five years before the retraction event. This ensures that all the fields in the dataset have at least a five year time series before its associated retraction event; each field defined in this way is followed up to the end of 2011. We then select 1,076 "nearest neighbor" articles that appear in the same journal and issue as the retracted articles, allowing us to delineate 1,076 control fields in an analogous fashion.

It is then straightforward to compute yearly "entry rates" into treated and control fields by counting the number of related articles published in the field in each year. Capturing funding information at the field level is slightly more involved. PubMed systematically records NIH grant acknowledgements using grant numbers, but without referencing the particular grant cycle to which the publication should be credited. We adopt the following procedure. For each related publication, we identify the closest preceding year in a three-year window during which funding was awarded through either a new award or a competitive renewal. We then sum all the funding in the grant year that ultimately generates publications in the focal field.

The descriptive statistics for the field-level analyses are displayed on Table 4. The number of observations across the publication frequency dataset and the funding dataset differ because (i) the funding data is available only until 2007, whereas the publication data is available until the end of our observation period (2011); and (ii) we drop from the funding analysis the fields for which there is not a single publication acknowledging NIH funding for the entire 1970-2007 period.

4 Results

The exposition of the econometric results proceeds in four stages. After a brief exposition of the main econometric issues, we present descriptive statistics and results pertaining to the effect of retractions on the rate of citations that accrue to the retracted articles. Second, we examine the extent of the retraction effect on the set of related articles. Third, we study whether the retraction events altered patterns of entry and funding into the scientific fields associated with the retracted articles. Fourth, we explicate the mechanism(s) underlying the results.

4.1 Econometric Considerations

Our estimating equation relates the number of citations that are received by related article j in year t to characteristics of j and of retracted article i:

$$E\left[CITES_{jt}|X_{ijt}\right] = exp\left[\beta_0 + \beta_1 RLTD_j \times AFTER_{it} + f(AGE_{jt}) + \delta_t + \gamma_{ij}\right]$$

where $AFTER$ denotes an indicator variable that switches to one the year after the retraction, $RLTD$ denotes an indicator variable that is equal to one for related articles and zero for control articles, $f(AGE_{jt})$ corresponds to a flexible function of article j's age, the δ_t's stand for a full set of calendar year indicator variables, and the γ_{ij}'s correspond to source article/related article (or control) fixed effects, consistent with our approach to analyze *changes* in j's rate of citations following the retraction of source article i.

The fixed effects control for many individual characteristics that could influence citation rates, such as journal status. To flexibly account for article-level life cycle effects, $f(AGE)$ consists of thirty two age indicator variables, where age measures the number of years elapsed since the article was published.[12]

Estimation. The dependent variable of interest is extremely skewed. For example, 40.33% of the artcle-year observations in the data correspond to years in which the related articles/controls receive zero citations. Following a long-standing tradition in the study of scientific and technical change, we present conditional quasi-maximum likelihood estimates based on the fixed-effect Poisson model developed by Hausman et al. (1984). Because the Poisson model is in the linear exponential family, the coefficient estimates remain consistent as long as the mean of the dependent variable is correctly specified (Gouriéroux et al. 1984).[13]

Inference. QML standard errors are robust to arbitrary patterns of serial correlation (Wooldridge 1997), and hence immune to the issues highlighted by Bertrand et al. (2004) concerning inference in DD estimation. We cluster the standard errors around retraction cases in the results presented below.

Dependent Variables. Our primary outcome variable is an article's number of citations in a given year. Secondary outcomes include the number of related articles (either to retracted papers or their nearest neighbors) published before and after the retraction event, as well as the amount of NIH funding (in millions of 2007 dollars) flowing to scientists who subsequently publish related articles (either to retracted papers or their nearest neighbors). Though the

[12]The omitted category corresponds to articles in their year of publication, i.e., articles' birth year. It is not possible to separately identify calendar year effects from age effects in the within-article dimension of our panel dataset in a completely flexible fashion, because one cannot observe two articles at the same point in time that have the same age but different vintages (Hall et al. 2007). In our specifications, the indicator variable corresponding to articles in their thirty-first year also absorbs the subsequent age dummies.

[13]In unreported specifications, we find that OLS yields qualitatively similar findings.

funding measure is distributed over the positive real line, the Hausman et al. estimator can still be used in this case (Santos Silva and Tenreyro 2006).

4.2 Effect of Retraction on Retracted Papers

Table 5 reports the results from simple difference-in-differences analyses for the sample of 1,037 retractions and 1,922 nearest neighbors in the journals in which the retracted articles appeared.[14] Column 1 reports the estimate of the retraction effect for the baseline specification. The result implies that, relative to the controls, retracted papers lose 69% of their citations in the post-retraction period. The magnitude of the effect is in line with the 60% decline estimated by Furman et al. 2012 in a smaller sample of PubMed indexed retractions. Column 2 shows that the effect is barely affected when we drop from the sample those observations corresponding to retracted articles for which the retraction reason is missing.

Column 3 includes in the specifications the main effect of the retraction treatment as well as two interactions with the "shaky shoulders" and "absent shoulders" indicator variables. In this model, the main effect implicitly captures the post-retraction fate of the retracted papers that still maintain "strong shoulders." While this effect is negative and statistically significant (with an implied decrease in the citation rate equal to 38%) its magnitude is markedly smaller than that of the effect corresponding to the "shaky shoulders" retractions (66%) and smaller still than the effect for the the "absent shoulders" category (73%). Dropping the "strong shoulders" group from the sample increases the magnitude of the retraction effect in absolute value (to 72%, column 4), while focusing on only the earliest retraction event in each case slightly lowers the estimated effect (66%, column 5). Finally, the results in column 6 indicate that, within the sample of "absent shoulders" retractions, the cases corresponding to serious fraud suffer an additional penalty, though this additional effect is not estimated particularly precisely.

In short, our results confirm the earlier findings of Furman et al. 2012. In addition, the results in column 3 provide important empirical validation for the coding exercise detailed

[14]Sixty seven retracted articles needed to be dropped from the estimation sample because they appeared in journals not indexed by the Web of Science.

in Appendix I. Although the coefficients in this specification are not statistically different from each other, their magnitudes are ordered in an intuitive way, with the post-retraction penalty decreasing monotonically with the strength of the shoulders provided to follow-on researchers.

4.3 Effect of Retraction on Related Papers

We now turn to the core of the empirical analysis, examining the effect of retraction effects on the citation outcomes for the related articles identified by the PubMed Related Citations Algorithm. The first set of results appears in Table 6, which is structured analogously to Table 5. Column 1 reports the difference-in-difference estimate for the entire sample. We find that related articles experience a precisely estimated 5.45% decline in the rate at which they are cited in the post-retraction period, relative to the control articles. Column 2 shows that the estimate does not change after dropping the articles related to retractions for which we were unable to find the underlying reason. Column 3 parses the retraction effect according to our "shoulders" coding. A clear difference emerges between the fate of articles related to "strong shoulders" retraction and the fate of those related to either "shaky shoulders" or "absent shoulders" retractions. The articles related to "strong shoulders" retractions are essentially immune to the retraction event (in fact the estimated effect is positive, but also small in magnitude and not statistically different from zero). In contrast, the implied elasticities for the articles related to "shaky shoulders" and "absent shoulders" retractions are 8.70% and 6.20%, respectively (the corresponding estimates are not statistically different from each other). In other words, we find evidence of negative spillovers of the retraction event onto the adjacent research area, but only in the cases for which the underlying cause of the retraction suggests that follow-on researchers should proceed with caution (if proceeding at all) before building on the retracted paper's results.

By eliminating from the estimation sample the observations associated with "strong shoulders" retractions, Column 4 further documents that the negative spillovers stemming from the retraction event are of comparable magnitudes for articles related to both "shaky shoulders" and "absent shoulders" retractions. Column 5 only retains the first retraction event across retraction cases. Though the magnitude of the treatment effect shrinks somewhat, it remains negative and precisely estimated. Finally, column 6 restricts the estimation sample even further, retaining only articles related to retractions offering "absent shoulders" and their nearest neighbor controls. In this sample, multiple-retraction fraud cases appear to account for almost all of the post-retraction penalty, with an estimated average yearly citation discount equal to ten percent.

The rest of our analysis focuses on the "absent shoulders" subsample of 589 retractions and 98,160 related or control articles. Figure 5 provides a way of scaling the negative spillovers of retraction events onto their related fields by comparing the post-retraction penalty experienced by related articles with the post-retraction penalty experienced by the retracted articles themselves. In both cases, the penalty is measured by differencing the log number of cumulative citations between 2011 and the year of the retraction event (using instead a fixed two-year window starting in the year of the retraction yields very similar results). The slope of the regression line is very close to .1, indicating that related articles lose, on average, only one tenth of the citations lost by the retraction. We note that this ratio dovetails with that of the elasticities estimated in Tables 5 and 6, respectively. Moreover, with an average of 60 related papers per retracted article, the *aggregate* citation consequences of the retraction events for the scientific fields involved are not trivial.

Dynamics of the treatment effect. We also explore the dynamics of the effects uncovered in Table 6. We do so in Figure 6 by estimating a specification in which the treatment effect is interacted with a set of indicator variables corresponding to a particular year relative to the retraction year, and then graphing the effects and the 95% confidence interval around them. Two features of the figure are worthy of note. First, there is no discernible evidence of an effect in the years leading up to the retraction, a finding that validates ex post our identification strategy.[15] Second, after the retraction, the treatment effect increases mono-

[15]This finding is also reassuring as it suggests that retractions are not endogenous to the exhaustion of a particular intellectual trajectory, i.e., it does not appear as if researchers resort to the type of misconduct that yields retractions after uncovering evidence that their field is on the decline.

tonically in absolute value with no evidence of recovery — the effect of the retraction event on the vitality of the field surrounding the retracted article appears permanent.

Exploring heterogeneity in the effect of retractions. We explore a number of factors that could modulate the magnitude of the retraction effect on intellectual neighbors' citation rates. Table 7 reports the results of five specifications that include interaction terms between the retraction treatment effect and characteristics of either the retracted article or the retracted/related article dyad. Column 1 simply replicates the simple specification of Table 6, column 1, though the estimation sample is limited to the articles related to the 589 retractions in the "absent shoulders" group and their associated control articles.

Column 2 examines whether the geographic origin of the retracted article's reprint author matters for the magnitude of the retraction response (the interaction corresponding to US-based reprint authors is omitted). There does not seem to be much of a differentiated response, though the articles related to Asian retractions appear relatively immune to the retraction event, in contrast to effects corresponding to the other geographic regions.

Column 3 focuses on heterogeneity with respect to the impact of the retracted article, measured by the cumulative number of citations garnered by the retracted article up to the end of the year in which the retraction event occurred. We create four interaction terms, the first pertaining to source articles in the second quartile of the baseline cumulative citation distribution, the second pertaining to the source articles in the third quartile, the third mapping into source articles with a citation count placing them between the 75^{th} and 95^{th} percentiles, and the fourth interaction capturing the highly-cited retracted articles (in the top ventile of the baseline cumulative citation distribution).[16] We find that more highly-cited retractions result in a larger negative spillovers on their neighboring field, though the contrast is really between low-impact (bottom quartile) retractions and the rest of the sample.

[16]The interaction between the retraction effect and a level of citations at baseline placing the source article in the bottom quartile is the omitted category. The universe of articles used to generate the cumulative citation distribution is restricted to the set of 589 articles in the "absent shoulders" retraction subsample. Because not all articles are retracted with the same speed (see Table 2), our measure gives some of the retractions a longer time span to accumulate citations. While an alternative is to make our quantile measures vintage-specific, we have found that for the earlier retraction years, the retracted article data is too sparse to generate meaningful variation using vintage-adjusted measures of retracted article pre-retraction impact.

Columns 4 and 5 examine whether citation linkages between the related and retracted articles moderate the magnitude of the retraction treatment effect. Recall that relatedness in the PMRA sense does not take into account citation information, but only semantic proximity as inferred from MeSH keywords. Related articlea can be published before the underlying source — in which case they are at risk of being *cited* by it — or after the source's publication (but before its retraction) — in which case they are at risk of *citing* the soon-to-be retracted publication. In column 4, we limit the estimation sample to the articles published after the retracted piece but before the retraction. In this subsample, we find that the negative retraction response to be especially pronounced (−13.15%) for the 6.1% of articles that were directly building on the retracted articles (as inferred by a citation link). Column 5, in contrast, restricts the estimation sample to the set of related articles (and their controls) that appeared before the retracted articles were published. We find that the related articles that are also cited by the retraction experience a 5.97% boost in the citation rate following the retraction event. This result is consistent with the idea that the researchers who continue to work in the field in spite of the retraction event choose to build instead on prior, unretracted research. The overall effect on the field can still be negative since only a small fraction (7.9%) of articles related to the source are also cited by the source.

Figure 7 explores the extent to which the age of a related article at the time of the retraction event influences the magnitude of the treatment effect. In this figure, each circle corresponds to the coefficient estimates stemming from a specification in which the citation rates for related articles and their controls are regressed onto year effects, article age indicator variables, as well as interaction terms between the treatment effect and the vintage of each related articles at the time of the retraction. Since related articles in the sample are published between one and ten years before their associated retraction event, there are ten such interaction terms.[17] The results show that only recent articles (those published one, two, or three years before the retraction) experience a citation penalty in the post-retraction period, whereas older articles are relatively immune to the retraction event.

[17]The 95% confidence interval (corresponding to robust standard errors, clustered around case codes) are denoted by the blue vertical bars.

Finally, Figures 8 and 9 investigate the extent to which "relatedness" (in the sense of PMRA) exacerbates the magnitude of the response. In Figure 8, we use the ordinal measure of relatedness, namely the relatedness rank received by a focal article in the list returned by PMRA for a specific source article. We create 22 interaction variables between the retraction effect and the relatedness rank: Top 5, Top 6-10,. . . , Top 95-100, 100 and above. The results show that lower-ranked (i.e., more closely related) articles bear the brunt of the negative citation response in the post-retraction event. Figure 9 is conceptually similar, except that it relies on the cardinal measure of relatedness. We create one hundred variables interacting the retraction effect with each percentile of the relatedness measure, and estimate the baseline specification of Table 7, column 1 in which the main retraction effect has been replaced by the 100 corresponding interaction terms. Figure 9 graphs the estimates along with the 95% confidence interval around them. The results are a bit noisy, but here, too, closely related articles (those for which the relatedness measure is above the 80^{th} percentile) appear to experience a sharper drop in citations post-retraction.

4.4 Effect on Entry and Funding at the Field Level

So far, the empirical exercise has examined cumulative knowledge production building on ideas that originated before the retraction event, allowing us to hold the quality of these ideas constant over the entire observation window. In order to understand the proximate causes of the negative spillovers documented above, we must examine whether the retraction event influenced the *production* of new ideas in the affected fields, and assess the extent to which these same events altered the distribution of funding across scientific fields.

Table 8 reports the results. Columns 1a, 1b, and 1c report our estimate of the treatment effect for entry into the retraction-relevant fields, whereas columns 2a and 2b report the treatment effect for funding. A number of interesting patterns emerge. First, the response is consistently negative, indicating that both funding and publication activity decrease in the affected fields following the first relevant retraction event and relative to the patterns observed in control fields. Second, the magnitude of the treatment effect increases when we define the boundaries around fields in a stricter fashion. Third, the funding response is always larger in magnitude than the publication response. Figure 10 provides event study graphs for both the publication intensity effect (Panel A) and funding effect (Panel B) using the same approach as that followed in Figure 6. In both cases, the magnitude of the retraction effect increases over time without evidence of a reversal.

To summarize, these results help explain why we observe downward movement in the citations received by related articles highlighted earlier: there are fewer papers being published in these fields and also less funding available to write such papers. While these effects constitute the proximate causes of the negative spillovers that are the central finding of the paper, they beg the question of what the underlying mechanisms are. What explains the flight of resources away from these fields?

4.5 Underlying Mechanisms of the Retraction Effect

We exploit the fine-grained level of detail in the data to distinguish between two potential mechanisms that could explain the retraction effect. Does the relative decrease in attention attendant to the retraction event reflect mostly scientists learning about the limited potential for follow-on research in retraction-afflicted fields, or does it instead proceed from a fear (rational or not) of having one's reputation besmirched through intellectual association with the "contaminated" fields? These two broad classes of explanations are not mutually exclusive, but ascertaining their relative importance matters because their welfare implications differ. If scientists simply redirect their efforts away from retraction-rich fields because they are less fertile after the retraction event, there would be little for the social planner to deplore. But if status considerations, divorced from cumulative knowledge production, drive the direction of scientific inquiry, the risk exists that the negative spillovers we documented earlier constitute underinvestment from a social welfare standpoint.

We begin by examining whether the retracted authors' intent influences the citation response to related articles written before the retraction event. Limiting the estimation sample to the set of retractions offering "absent shoulders" to follow-on researchers, we include in the benchmark specification two additional variables corresponding to the interaction of the retraction effect with, respectively, the "uncertain intent" and "intentional deception" indicators mentioned earlier (Table 9, column 1b). The evidence clearly shows that the post-retraction penalty is larger when there is clear evidence of malicious intent. It is possible that retractions associated with misconduct are, even in this restricted sample, more consequential for the field than are retractions associated with "honest mistakes." However, the same effect is observed in column 1c, where we also control for retraction "size" by including in the specification interaction terms between the retraction effect and the quartiles of post-retraction penalty at the retracted article level (coefficients not reported).

Our second attempt at disentangling mechanisms focuses on heterogeneous citer responses. We start from the premise that scientists employed by profit-seeking firms would persist in investigating topics that university-based scientists (and NIH study sections) frown upon, as long as the possibility of developing a commercial product remained. We parse the forward citation data to separate the citations that stem from private firms (mostly pharmaceutical and biotechnology firms, identified by suffixes such as *Inc.*, *Corp.*, *LLC*, *Ltd.*, *GmbH*, etc.) from those that originate in academia (broadly defined to include non-profit research institutes and public research institutions as well as universities). Even though we classify as "private" any citing article with a mix of private and academic addresses, almost 90% of the citations in our sample are "academic" according to this definition. In columns 2a and 2b, we find that academic and private citers do not differ at all in the extent to which they penalize the *retracted articles.* Conversely, columns 2c and 2d indicate that private citers hardly penalize *related articles*, whereas academic citers do to the extent previously documented.[18] The difference between the coefficients is statistically significant at the 95^{th} level ($p < 0.01$).

Considered as a whole, these findings are consistent with the view that the retraction-induced spillovers we have documented stem, at least in part, from scientists' concern that their peers will hold them in lower esteem if they remain within an intellectual field whose

[18]The estimation sample is limited to the set of related articles and their controls that receive at least one citation of each type over the observation period.

reputation has been tarnished by retractions, even though these scientists were neither coauthors on the retracted article itself nor building directly upon it. In short, there is an empirical basis to claim that retractions lead to underinvestment in the affected fields, thus challenging science policy makers to design an appropriate policy response.

5 Conclusions

This paper constitutes the first investigation of the effect of "false science" on the direction of scientific progress. We find that scientific misconduct stifle scientists' pursuit of specific research lines, as we would anticipate if retraction events provide new signals of the fidelity of scientific knowledge. More centrally, our findings show that scientific misconduct and mistakes, as signaled to the scientific community through retractions, cause a relative decline in the vitality of neighboring intellectual fields. These spillovers in intellectual space are significant in magnitude and persistent over time. In other words, there is clear evidence of negative spillovers in instances of "false science" to broader swaths of the intellectual field in which they take place.

Of course, an important limitation of our analytical approach is that, though we can document that retraction events cause a decrease in the rate of citations to related articles, we cannot pinpoint exactly where the missing citations go, or more precisely, in which direction scientists choose to redirect their inquiries after the event. Nonetheless, the empirical evaluation has a number of interesting implications. Through the coding scheme we have developed to understand the particular circumstance of each retraction event, we highlight the limitations of the institutional practices that are supposed to ensure the fidelity of scientific knowledge. In particular, the analysis brings systematic evidence to bear on the heightened attention devoted to the topic of scientific misconduct in science policy circles. Some analysts suggest that the scientific reward system has been corrupted and is in need of wholesale, radical reform (Fang et al. 2012). Others retort that a system of retractions is precisely what the "Republic of Science" needs — a mechanism through which mistakes and misconduct can be swiftly and easily signaled (Furman et al. 2012). The validity of the more optimistic view hinges crucially on what is signaled by a retraction notice and on how scientists in the affected fields process this information and act upon it. Our results suggest that retractions do have the desired effect on the particular paper in question, but also lead to spillover effects onto the surrounding intellectual fields, which become less vibrant.

If these negative spillovers simply reflected the diminished scientific potential of the affected fields, then the "collateral damage" induced by retractions would not be a cause for concern and would reinforce the belief that the retraction process is a relatively effective way to police the scientific commons (Furman et al. 2012). However, our evidence indicates that broad perceptions of legitimacy are an important driver of the direction of scientific inquiry. Unfortunately, retraction notices often obfuscate the underlying reason for retraction, which diminishes the information content of the signal they provide to follow-on researchers. As a result, there could be high returns to developing a standardized coding approach for retractions that journals and scientific societies could draw upon to help the scientific community update their beliefs regarding the nature and scope of false science. While journal editors may understandably balk at the suggestion that it is incumbent upon them to make clear determinations regarding the underlying causes of retractions, a clearly-articulated schema would increase the incentives of authors to report problems emerging after the publication of an article and provide a more nuanced context within which universities themselves (as well as funding bodies) might investigate and adjudicate instances of false science.[19]

A second issue raised by our paper relates to our understanding of what constitutes an intellectual field. As we noted in the introduction, economists have devoted considerably more time and attention to the study of the rate of inventive activity than to its direction. This gap has arisen in part because of the empirical challenges associated with delineating the boundaries among intellectual fields. Our approach relaxes the data constraint through the systematic use of keyword information. The same approach could also prove itself useful to explore more generally the ways in which researchers, through their publications, choose

[19] Alternative mechanisms — such as "replication rings" — have been proposed to counteract the negative spillovers in intellectual space associated with retraction events (Kahneman 2012). Whether "local" responses of this type can be implemented successfully is questionable, in light of the costs they would impose on researchers active in retraction-affected fields.

positions in intellectual space, and change these positions over time. At the same time, economists' conceptual grasp of intellectual landscapes remains in its infancy, with a near exclusive focus on vertical "research lines" (cf. Aghion et al. 2008). We hope that our empirical results will prove useful to economists seeking to understand movement across research lines and the consequences of these movements for cumulative knowledge production and, ultimately, economic growth.

References

Acemoglu, Daron. 2012. "Diversity and Technological Progress." In Josh Lerner, and Scott Stern (Eds.), *The Rate & Direction of Inventive Activity Revisited*, pp. 319-356. Chicago, IL: University of Chicago Press.

Aghion, Philippe, Mathias Dewatripont, Fiona Murray, Julian Kolev, and Scott Stern. 2009. "Of Mice and Academics: Examining the Effect of Openness on Innovation." NBER Working Paper #14819.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. "Academic Freedom, Private Sector Focus, and the Process of Innovation." *RAND Journal of Economics* **39**(3): 617-635.

Agrawal, Ajay K., and Avi Goldfarb. 2008. "Restructuring Research: Communication Costs and the Democratization of University Innovation." *American Economic Review* **98**(4): 1578-1590.

Associated Press. 2005. "Spectacular fraud shakes stem cell field." December 23, 2005. Accessed from http://www.msnbc.msn.com/id/10589085/ on 10/18/2012.

Azoulay, Pierre, Joshua Graff Zivin, and Jialan Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics* **125**(2): 549-589.

Babbage, Charles. 1830. *Reflections on the Decline of Science in England, and on Some of Its Causes*. London, UK: B. Fellowes and J. Booth.

Bachrach, C. A., and Thelma Charen. 1978. "Selection of MEDLINE Contents, the Development of its Thesaurus, and the Indexing Process." *Medical Informatics (London)* **3**(3): 237-254.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* **119**(1): 249-275.

Broad, William, and Nicholas Wade. 1983. *Betrayers of the Truth: Fraud and Deceit in the Halls of Science*. New York, NY: Simon & Schuster.

Dasgupta, Partha, and David. Paul. 1994. "Towards a New Economics of Science." *Research Policy* **23**(5): 487-521.

David, Paul A. 2008. "The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution." *Capitalism and Society* **3**(2): Article 5.

Ding, Waverly W., Sharon G. Levin, Paula E. Stephan, and Anne E. Winkler. 2010. "The Impact of Information Technology on Scientists' Productivity, Quality and Collaboration Patterns." *Management Science* **56**(9): 1439-1461.

Fang, Ferric C., R. Grant Steen, and Arturo Casadevall. 2012. "Misconduct Accounts for the Majority of Retracted Scientific Publications." *Proceedings of the National Academy of Science Early Edition*.

Furman, Jeffrey L., and Scott Stern. 2011. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Knowledge Production." *American Economic Review* **101**(5): 1933-1963.

Furman, Jeffrey L., Kyle Jensen, and Fiona Murray. 2012. "Governing Knowledge in the Scientific Community: Exploring the role of Retractions in Biomedicine." *Research Policy* **41**(2): 276-290.

Furman, Jeffrey, Fiona Murray, and Scott Stern. 2012. "Growing Stem Cells: The Impact of US Policy on the Geography and Organization of Scientific Discovery." *Journal of Policy Analysis and Management* **31**(3): 661-705.

Gouriéroux, Christian, Alain Montfort, and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica* **53**(3): 701-720.

Hall, Bronwyn H., Jacques Mairesse, and Laure Turner. 2007. "Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists." *Economics of Innovation and New Technology* **16**(2): 159-177.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* **52**(4): 909-938.

Hull, David L. 1988. *Science as a Process*. Chicago, IL: University of Chicago Press.

Hwang, Woo Suk, Sung Il Roh, Byeong Chun Lee, Sung Keun Kang, Dae Kee Kwon, Sue Kim, Sun Jong Kim, Sun Woo Park, Hee Sun Kwon, Chang Kyu Lee, Jung Bok Lee, Jin Mee Kim, Curie Ahn, Sun Ha Paek, Sang Sik Chang, Jung Jin Koo, Hyun Soo Yoon, Jung Hye Hwang, Youn Young Hwang, Ye Soo Park, Sun Kyung Oh, Hee Sun Kim, Jong Hyuk Park, Shin Yong Moon, and Gerald Schatten. 2005. "Patient-Specific Embryonic Stem Cells Derived from Human SCNT Blastocysts." *Science* **308**(5729): 1777-1783.

Kim, Kitai, Paul Lerou, Akiko Yabuuchi, Claudia Lengerke, Kitwa Ng, Jason West, Andrew Kirby, Mark J. Daly, and George Q. Daley. 2007. "Histocompatible Embryonic Stem Cells by Parthenogenesis." *Science* **315**(5811): 482-486.

Lacetera, Nicola, and Lorenzo Zirulia. 2009. "The Economics of Scientific Misconduct." *Journal of Law, Economics, and Organization* **27**(3): 568-603.

LaFollette, Marcel C. 1992. *Stealing Into Print: Fraud, Plagiarism, and Misconduct in Scientific Publishing*. Berkeley, CA: University of California Press.

Levin, Sharon G., and Paula E. Stephan. 1991. "Research Productivity over the Life Cycle: Evidence for Academic Scientists." *American Economic Review* **81**(1): 114-32.

Lin, Jimmy, and W. John Wilbur. 2007. "PubMed Related Articles: A Probabilistic Topic-based Model for Content Similarity." *BMC Bioinformatics* **8**(423).

Lu, Susan Feng, Ginger Jin, Benjamin Jones, and Brian Uzzi. 2012. "Protecting Truth in Science: The Prior Publication Penalty." Working Paper, Northwestern University.

Martinson, Brian C., Melissa S. Anderson, and Raymond de Vries. 2005. "Scientists Behaving Badly." *Nature* **435**(7043): 737-738.

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

Murray, Fiona. 2010. "The Oncomouse that Roared: Hybrid Exchange Strategies as a Source of Productive Tension at the Boundary of Overlapping Institutions." *American Journal of Sociology* **116**(2): 341-388.

Murray, Fiona, and Scott Stern. 2007. "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?" *Journal of Economic Behavior and Organization* **63**(4): 648-687.

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

Nath, Sara B., Steven C. Marcus, and Benjamin G. Druss. 2006. "Retractions in the Research Literature: Misconduct or Mistakes?" *Medical Journal of Australia* **185**(3): 152-154.

Névéol, Aurélie, Rezarta Islamaj Dogan, and Zhiyong Lu. 2010. "Author Keywords in Biomedical Journal Articles." *AMIA Symposium Proceedings* 537-541.

NIH Office of Research Integrity. 2007. *Annual Report.* Rockville, MD: Department of Health and Human Services.

Pozzi, Andrea, and Paul A. David. 2007. "Empirical Realities of Scientific Misconduct in Publicly Funded Research: What Can Be Learned from the Data?" in *ESF-ORI First World Conference on Scientific Integrity—Fostering Responsible Research*, held at the Gulbenkian Foundation, Lisbon, Portugal.

Santos Silva, J.M.C., and Silvanna Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics* **88**(4): 641-658.

Weiner, J.S. 1955. *The Piltdown Forgery.* New York, NY: Oxford University Press.

Wooldridge, Jeffrey M. 1997. "Quasi-Likelihood Methods for Count Data." In M. Hashem Pesaran, and Peter Schmidt (Eds.), *Handbook of Applied Econometrics*, pp. 352-406. Oxford: Blackwell.

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

Table 1: Reasons for Retractions

	All Cases		"Strong Shoulders" Subsample		"Shaky Shoulders" Subsample		"Absent Shoulders" Subsample	
Plagiarism	90	8.15%	78	38.61%	11	3.81%	1	0.17%
Duplicated Publication	92	8.33%	90	44.55%	2	0.69%	0	0.00%
Publisher Error	13	1.18%	8	3.96%	5	1.73%	0	0.00%
Faulty/Absent IRB Approval	9	0.82%	5	2.48%	4	1.38%	0	0.00%
Not Enough Information To Classify	42	3.80%	0	0.54%	36	12.46%	6	1.05%
Questions About Validity	35	3.17%	0	0.00%	31	10.73%	4	0.68%
Author Dispute	33	2.99%	5	2.48%	28	9.69%	0	0.00%
Miscellaneous	24	2.17%	15	7.43%	8	2.77%	1	0.17%
Did Not Maintain Proper Records	3	0.27%	0	0.00%	3	1.04%	0	0.00%
Fake Data	361	32.70%	0	0.00%	14	4.84%	347	58.91%
Error/Mistake	271	24.55%	1	0.50%	62	21.45%	208	35.31%
Could Not Replicate	92	8.33%	0	0.00%	78	26.99%	14	2.38%
Fake Data & Plagiarism	15	1.36%	0	0.00%	7	2.45%	8	1.36%
Missing	24	2.17%	0	0.00%	0	0.00%	0	0.00%
Total	1,104	100.00%	202	100.00%	289	100.00%	589	100.00%

Note: Retraction reasons for a set of 1,104 original articles indexed by PubMed, published between 1973 and 2008, and retracted before the end of 2009. This sample is further broken down into three subsamples. The "strong shoulders" subsample comprises 202 articles retracted for typically innocuous reasons, or at least reasons that do not cast doubt on the veracity of the results contained therein. The "shaky shoulders" subsample comprises 289 retracted articles for which either the retraction notice or information retrievable on the world-wide web cast some doubt on the extent the results should be built upon by follow-on researchers. Finally, the "absent shoulders" subsample contains 589 retracted articles that will be the source sample for the bulk of the analysis. For these cases, we could ascertain with substantial certainty that the results are not to be relied upon for future research. This can occur because of intentional misconduct on the part of the researchers involved, or because of mistakes on their part. The comprehensive spreadsheet listing of these retracted articles – complete with the references used to code retraction reasons – can be downloaded at http://jkrieger.scripts.mit.edu/retractions/.

.

Table 2: Descriptive Statistics for 589 Retracted Source Articles ["Absent Shoulders" Subsample]

	Mean	Median	Std. Dev.	Min.	Max.
Publ. Year for Retracted Article	1997.606	2000	7.848	1973	2007
Retraction Year	2000.844	2004	7.821	1977	2009
Retraction Speed (years)	3.238	2	2.893	0	16
Nb. of Related Articles	59.205	43	64.021	1	627
Part of a Multiple Retractions Case	0.625	1	0.485	0	1
Intentional Deception	0.611	1	0.488	0	1
Uncertain Intent	0.081	0	0.274	0	1
No Sign of Intentional Deception	0.307	0	0.462	0	1
Part of a Multiple Retractions Fraud Case	0.458	0	0.499	0	1
Cumulative Citations [as of 7/2012]	45.100	21	70.493	0	728
US-based Reprint Author	0.533	1	0.499	0	1

Note: These 589 retractions can be grouped into 334 distinct cases – a case arises because a researcher, or set of researchers, retracts several papers for related reasons, e.g., because of repeated fraud.

Table 3: Descriptive Statistics for Related Articles and "Nearest-Neighbor" Controls ["Absent Shoulders" Subsample]

		Mean	Median	Std. Dev.	Min.	Max.
NN Controls (N=65,461)	Article Publication Year	1999.098	2001	7.018	1970	2008
	Number of Authors	5.134	5	2.951	1	78
	PubMed Matching Score	0.575	1	0.164	0	1
	Article Relatedness Ranking	157.012	91	201.024	1	1,987
	Article Age at time of Retraction	3.988	4	2.427	1	10
	Published After Retracted Article	0.300	0	0.458	0	1
	Baseline Stock of Cites	12.020	4	35.497	0	3,064
	Baseline Stock of Cites from Private Firms	1.157	0	3.809	0	230
	Cites Retracted Piece (N=19,617)	0.001	0	0.031	0	1
	Cited by Retracted Piece (N=34,190)	0.001	0	0.028	0	1
Related Articles (N=32,699)	Article Publication Year	1999.244	2001	6.969	1970	2008
	Number of Authors	5.119	5	2.726	1	51
	PubMed Matching Score	0.568	1	0.161	0	1
	Article Relatedness Ranking	163.157	94	204.430	1	1,987
	Article Age at time of Retraction	3.960	4	2.421	1	10
	Published After Retracted Article	0.300	0	0.458	0	1
	Baseline Stock of Cites	13.811	4	38.534	0	3,713
	Baseline Stock of Cites from Private Firms	1.272	0	4.203	0	368
	Cites Retracted Piece (N=9,804)	0.079	0	0.270	0	1
	Cited by Retracted Piece (N=17,047)	0.061	0	0.240	0	1

Note: The set of related articles is composed of journal articles linked to the 589 retracted articles of Table 2 through PubMed's "related articles" algorithm (see Figure 1) and downloaded using the open source FindRelated software [`http://www.stellman-greene.com/FindRelated/`]. We exclude from the raw data (i) articles that do not contain original research, e.g., reviews, comments, editorials, letters; (ii) articles published outside of a time window running from ten years before the retraction event to one year before the retraction event; (iii) articles that appear in journals indexed by PubMed but not indexed by Thompson-Reuters' *Web of Science;* (iv) articles that we fail to match to *Web of Science;* (v) articles that we do match to *Web of Science*, but receive zero forward citations (exclusive of self-citations) from their publication year up until the end of 2011; and (vi) articles for which at least one author also appears on the authorship roster of the corresponding retracted article. For each related article, we select as controls its "nearest neighbors" in the same journal and issue – i.e., the articles that immediately precede and/or immediately follow it in the issue. By convention, the controls inherit some of the properties of their treated neighbor, including relateness ranking and score.

Table 4: Descriptive Statistics for the Entry and Funding Samples

		Article Frequencies [1975-2011]						Funding [1975-2007]				
		Nb. of Related Articles		Nb. of Closely Related Articles (rank 20 or lower)		Nb. of Closely Related Articles (80% score or higher)			Nb. of Grants		$ Amounts	
		Mean	Std. Error	Mean	Std. Error	Mean	Std. Error		Mean	Std. Error	Mean	Std. Error
Control	N=1,076	4.64	7.49	0.31	0.88	0.22	0.55	N=778	1.21	2.43	$5,587,872	$20,433,959
Retracted	N=565	3.99	7.36	0.24	0.80	0.15	0.46	N=411	1.16	2.64	$5,077,185	$16,683,593
Total	N=1,1641	4.42	7.45	0.29	0.86	0.20	0.52	N=1,189	1.19	2.50	$5,413,844	$19,239,559

Note: We compute entry rates into the field surrounding a retracted article (or one of its nearest neighbor) by counting the number of PubMed-related articles in a particular year. We measure NIH funding for the same fields by summing the grant amounts awarded in a particular year that yields at least one publication over the next three years that is related to either a retracted article or one of their nearest-neighbor controls. The means and standard deviations are computed over all observations in the resulting retracted article/year panel dataset (N×T=53,451 for related article frequencies; N×T=42,524 for funding).

Table 5: Effects of Retraction on Citations to Retracted Articles, by Retraction Reason

	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample	Excludes Missing Rtrct. Reasons	Excludes Missing Rtrct. Reasons	Further Excludes "Strong Shoulders" Retractions	Only earliest retraction event in each case	Only Includes "Absent Shoulders" Retractions
After Retraction	-1.171**	-1.172**	-0.472**	-1.210**	-1.081**	-1.128**
	(0.099)	(0.100)	(0.099)	(0.104)	(0.066)	(0.115)
After Retraction × Shaky Shoulders			-0.609**			
			(0.141)			
After Retraction × Absent Shoulders			-0.809**			
			(0.152)			
After Retraction × Multiple Fraud Case						-0.196†
						(0.111)
Nb. of Retraction Cases	720	705	705	552	552	552
Nb. of Retracted/Control Articles	2,959	2,915	2,915	2,431	1,570	2,393
Nb. of Article-Year Obs.	39,469	38,925	38,925	34,735	20,513	33,920
Log Likelihood	-62,620	-62,182	-62,054	-57,596	-34,611	-57,019

Note: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable is the total number of forward citations (exclusive of self-citations) received by each retracted article (or its nearest neighbor controls) in a particular year. All models incorporate a full suite of calendar year effects as well as 31 article age indicator variables (age zero is the omitted category). Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities. For example. the estimates in column (1) imply that retracted articles suffer on average a statistically significant (1-exp[-1.171])=68.99% yearly decrease in the citation rate after the retraction event.

QML (robust) standard errors in parentheses, clustered around retraction cases.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 6: Effects of Retractions on Citations to Related Articles, by Retraction Reason

	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample	Excludes Missing Rtrct. Reasons	Excludes Missing Rtrct. Reasons	Further Excludes "Strong Shoulders" Retractions	Only earliest retraction event in each case	Only Includes "Absent Shoulders" Retractions
After Retraction	-0.059**	-0.059**	0.040	-0.085**	-0.038*	-0.013
	(0.013)	(0.013)	(0.030)	(0.030)	(0.016)	(0.026)
After Retraction × Shaky Shoulders			-0.131**			
			(0.044)			
After Retraction × Absent Shoulders			-0.104**	0.028		
			(0.037)	(0.038)		
After Retraction × Multiple Fraud Case						-0.094*
						(0.037)
Nb. of Retraction Cases	770	747	747	573	572	334
Nb. of Source Articles	1,104	1,080	1,080	878	580	589
Nb. of Related/Control Articles	169,741	167,306	167,306	137,969	90,167	98,160
Nb. of Article-Year Obs.	2,094,725	2,064,465	2,064,465	1,800,425	1,066,306	1,261,713
Log Likelihood	-2,747,714	-2,714,047	-2,713,760	-2,398,154	-1,457,463	-1,686,237

Note: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable is the total number of forward citations (exclusive of self-citations) received by each related article in a particular year. All models incorporate a full suite of calendar year effects as well as 31 article age indicator variables (age zero is the omitted category). Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities. For example. the estimates in column (1) imply that related articles suffer on average a statistically significant (1-exp[-0.059])=5.73% yearly decrease in the citation rate after the retraction event.

QML (robust) standard errors in parentheses, clustered around retraction cases.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 7: Exploring Heterogeneity in the Magnitude of the Retraction Effect "Absent Shoulders" Subsample

	(1)	(2)	(3)	(4)	(5)
After Retraction	-0.058** (0.017)	-0.077** (0.027)	0.087 (0.080)	-0.019 (0.024)	-0.077** (0.018)
After Retraction × Asian Reprint Author		0.092* (0.044)			
After Retraction × European Reprint Author		0.014 (0.049)			
After Retraction × RoW-based Reprint Author		-0.112 (0.142)			
After Retraction × Source in btw. 25% and 50% of cit. distrib.			-0.130 (0.084)		
After Retraction × Source in btw. 50% and 75% of cit. distrib.			-0.155† (0.084)		
After Retraction × Source in btw. 75% and 95% of cit. distrib.			-0.193* (0.086)		
After Retraction × Source in btw. 95% and 100% of cit. distrib.			-0.098 (0.130)		
After Retraction × Cites Retracted Piece				-0.141* (0.058)	
After Retraction × Cited by Retracted Piece					0.135* (0.054)
Nb. of Retraction Cases	334	334	334	204	325
Nb. of Source Articles	589	589	589	384	551
Nb. of Related/Control Articles	98,160	98,160	98,160	29,421	51,237
Nb. of Article-Year Obs.	1,261,713	1,261,713	1,261,713	334,424	720,165
Log Likelihood	-1,686,525	-1,686,366	-1,686,162	-434,755	-973,570

Note: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable is the total number of forward citations (exclusive of self-citations) received by each related article in a particular year. All models incorporate a full suite of calendar year effects as well as 31 article age indicator variables (age zero is the omitted category). Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities. For example. the estimates in column (1) imply that related articles suffer on average a statistically significant (1-exp[-0.058])=5.64% yearly decrease in the citation rate after the retraction event.

QML (robust) standard errors in parentheses, clustered around retraction cases.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 8: Effect of Retraction on Publication Frequency and NIH Funding

	(1a)	(1b)	(1c)	(2a)	(2b)
	Nb. of Related Articles	Nb. of Closely Related Articles (rank 20 or lower)	Nb. of Closely Related Articles (80% score or higher)	Nb. of Grants	$ Amounts
After Retraction	-0.354**	-0.522**	-0.465**	-1.145**	-1.312**
	(0.088)	(0.143)	(0.117)	(0.106)	(0.130)
Nb. of Retraction Cases	333	333	333	331	331
Nb. of Retracted/Control Articles	1,641	1,499	1,609	1,499	1,499
Nb. of Article-Year Obs.	53,451	48,829	52,451	42,524	42,524
Log Likelihood	-170,399	-27,034	-23,828	-50,394	-243,240

Note: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable is the total number of related articles published in a particular source/year (columns 1a, 1b, and 1c) or the total dollar amount of NIH funding awarded in a particular year that yields at least one publication over the next three years that is related to either a retracted article or one of their nearest-neighbor controls (columns 2a, and 2b). All models incorporate a full suite of calendar year effects.

QML (robust) standard errors in parentheses, clustered around retraction cases.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 9: Interpreting Citation Behavior for Articled Related to "Absent Shoulders" Retractions

	Does Intent Matter for Within-Field Spillovers?			**Heterogeneous Citer Behavior**			
				Retracted Papers		**Related Papers**	
	(1a) Benchmark Specification	(1b) Interaction with Retracted Authors' Intent	(1c)	(2a) Academic Citations Only	(2b) Private-Firms Citations Only	(2c) Academic Citations Only	(2d) Private-Firms Citations Only
		w/o controls for retraction "size"	w/ controls for retraction "size"				
After Retraction	-0.058**	0.016	0.008	-1.293**	-1.309**	-0.054**	-0.006
	(0.017)	(0.039)	(0.043)	(0.154)	(0.188)	(0.017)	(0.023)
After Retraction × Uncertain Intent		-0.099	-0.094				
		(0.065)	(0.064)				
After Retraction × Intentional Deception		-0.097*	-0.099*				
		(0.046)	(0.045)				
Nb. of Retraction Cases	334	334	334	304	304	334	334
Nb. of Source Articles	589	589	589	1,089	1,089	589	589
Nb. of Related/Control Articles	98,160	98,160	98,160			62,205	62,205
Nb. of Article-Year Obs.	1,261,713	1,261,713	1,261,713	15,711	15,711	807,203	807,203
Log Likelihood	-1,686,525	-1,686,298	-1,686,141	-30,568	-8,234	-1,366,136	-402,337

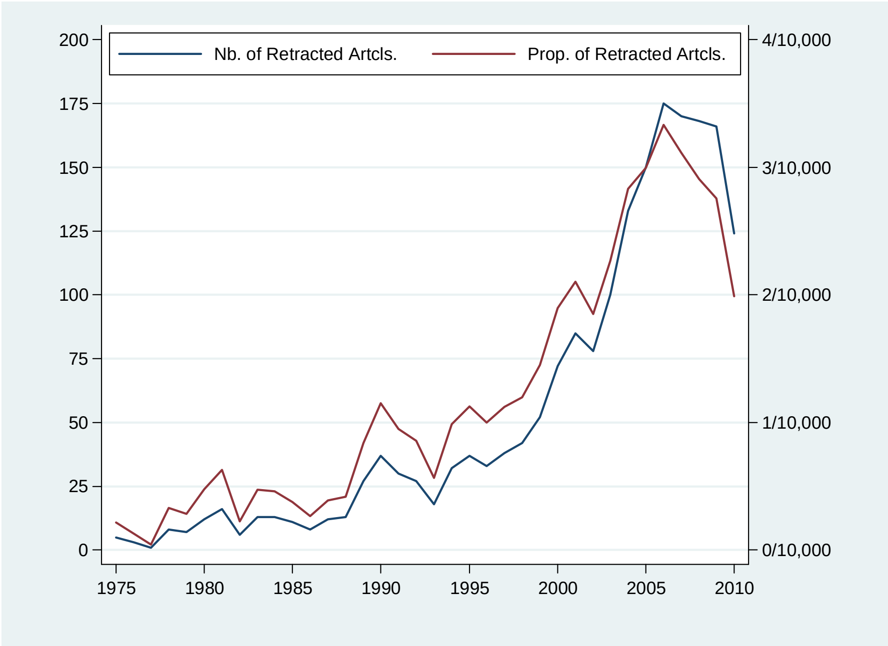
Note: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable is the total number of forward citations (exclusive of self-citations) received by each related article in a particular year. All models incorporate a full suite of year effects as well as 31 article age indicator variables (age zero is the omitted category). Exponentiating the coefficients and differencing from one yields numbers interpretable as elasticities.

Retraction "size" refers to four covariates capturing the magnitude of the post-retraction change in citations experienced by the underlying retracted articles – btw. the 25th and 50th percentile, btw. the 80th and 75th percentile, btw. the 75th and 95th percentile, and above the 95th percentile (the omitted category corresponds to post-retraction citation change in the bottom quartile; the size covariates are included in column [1c] but the associated coefficients/standard errors are not reported).

In columns (2a) and (2b), the estimation sample is limited to those related articles and controls that receive at least one "private firm" citation between their year of publication and 2011. For this analysis, a citation is said to emanate from a private firm when at least one address listed by the *Web of Science* includes a suffix such as `Inc.`, `Corp.`, `LLC`, `Ltd.`, `GmbH`, etc.

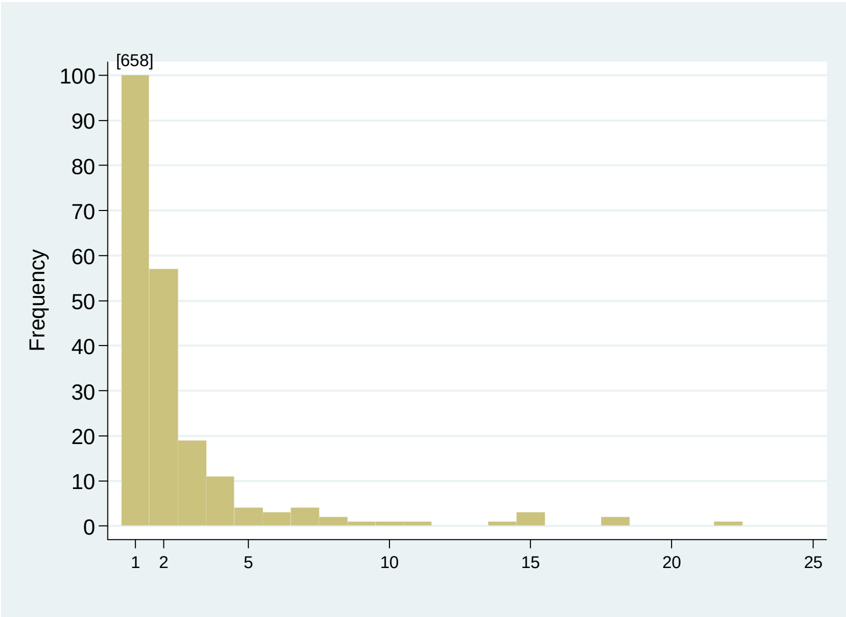
QML (robust) standard errors in parentheses, clustered around retraction cases.
$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure 1: Incidence of PubMed-Indexed Retractions



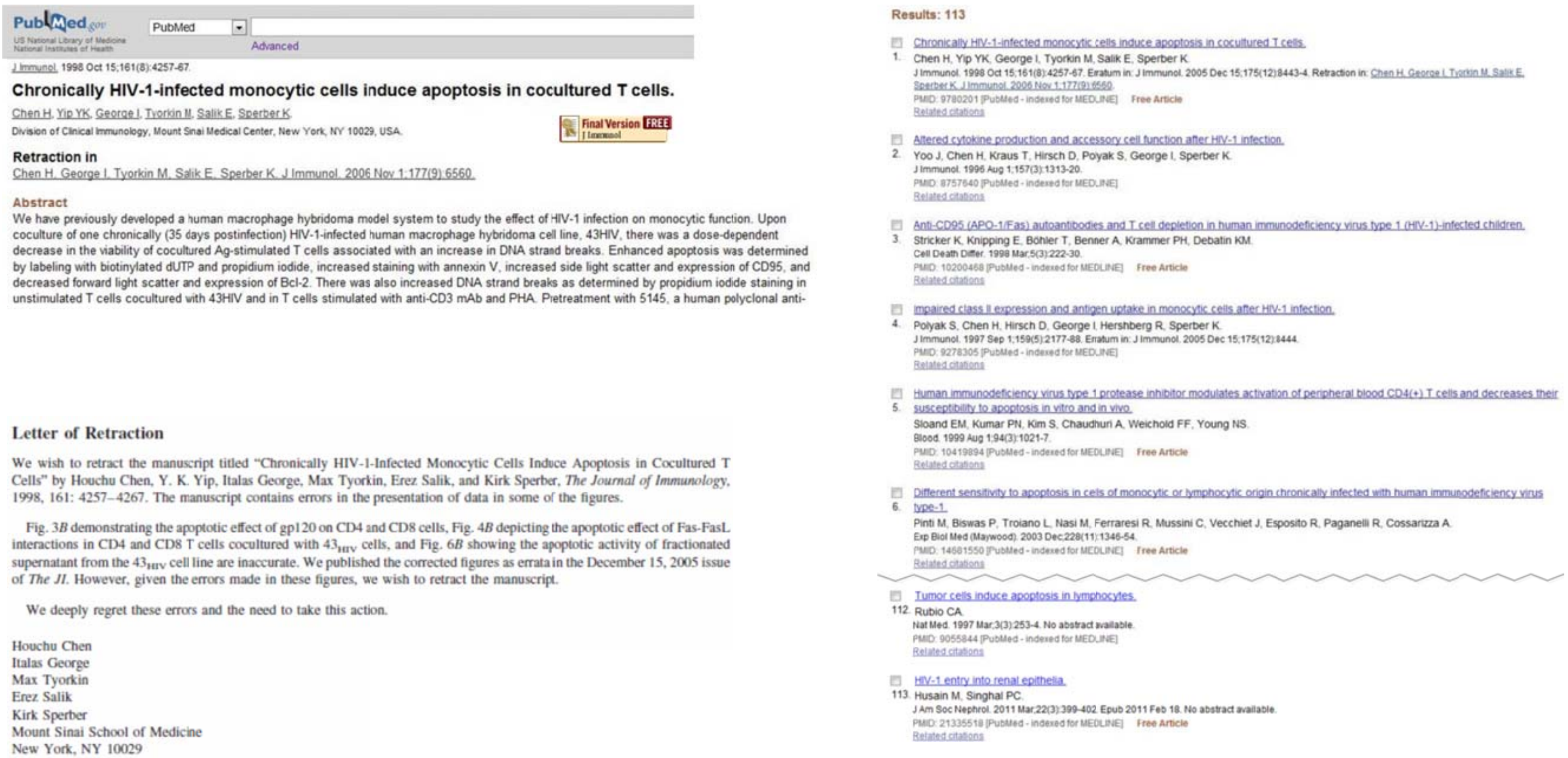
Note: The solid blue line displays the yearly frequency of retraction events in PubMed as a whole, all retraction reasons included. The solid red line displays the yearly retraction rate, where the denominator excludes PubMed-indexed articles that are not original journal articles (e.g., comments, editorials, reviews, etc.)

Figure 2: Distribution of Retraction Events by Retraction Case



Note: The left-most bar in this histogram, corresponding to single-retraction cases, has been truncated. These singleton cases comprise 658 retraction events (59.60% of the sample).

Figure 3: Retracted & Related Articles



PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed
Advanced

J Immunol. 1998 Oct 15;161(8):4257-67.

Chronically HIV-1-infected monocytic cells induce apoptosis in cocultured T cells.

Chen H, Yip YK, George I, Tyorkin M, Salik E, Sperber K.

Division of Clinical Immunology, Mount Sinai Medical Center, New York, NY 10029, USA.

Final Version FREE J Immunol

Retraction in

Chen H, George I, Tyorkin M, Salik E, Sperber K. J Immunol. 2006 Nov 1;177(9):6560.

Abstract

We have previously developed a human macrophage hybridoma model system to study the effect of HIV-1 infection on monocytic function. Upon coculture of one chronically (35 days postinfection) HIV-1-infected human macrophage hybridoma cell line, 43HIV, there was a dose-dependent decrease in the viability of cocultured Ag-stimulated T cells associated with an increase in DNA strand breaks. Enhanced apoptosis was determined by labeling with biotinylated dUTP and propidium iodide, increased staining with annexin V, increased side light scatter and expression of CD95, and decreased forward light scatter and expression of Bcl-2. There was also increased DNA strand breaks as determined by propidium iodide staining in unstimulated T cells cocultured with 43HIV and in T cells stimulated with anti-CD3 mAb and PHA. Pretreatment with 5145, a human polyclonal anti-

Letter of Retraction

We wish to retract the manuscript titled "Chronically HIV-1-Infected Monocytic Cells Induce Apoptosis in Cocultured T Cells" by Houchu Chen, Y. K. Yip, Italas George, Max Tyorkin, Erez Salik, and Kirk Sperber, *The Journal of Immunology*, 1998, 161: 4257–4267. The manuscript contains errors in the presentation of data in some of the figures.

Fig. 3*B* demonstrating the apoptotic effect of gp120 on CD4 and CD8 cells, Fig. 4*B* depicting the apoptotic effect of Fas-FasL interactions in CD4 and CD8 T cells cocultured with 43_{HIV} cells, and Fig. 6*B* showing the apoptotic activity of fractionated supernatant from the 43_{HIV} cell line are inaccurate. We published the corrected figures as errata in the December 15, 2005 issue of *The JI*. However, given the errors made in these figures, we wish to retract the manuscript.

We deeply regret these errors and the need to take this action.

Houchu Chen
Italas George
Max Tyorkin
Erez Salik
Kirk Sperber
Mount Sinai School of Medicine
New York, NY 10029

Results: 113

1. Chronically HIV-1-infected monocytic cells induce apoptosis in cocultured T cells.
Chen H, Yip YK, George I, Tyorkin M, Salik E, Sperber K.
J Immunol. 1998 Oct 15;161(8):4257-67. Erratum in: J Immunol. 2005 Dec 15;175(12):8443-4. Retraction in: Chen H, George I, Tyorkin M, Salik E, Sperber K. J Immunol. 2006 Nov 1;177(9):6560.
PMID: 9780201 [PubMed - indexed for MEDLINE] Free Article
Related citations

2. Altered cytokine production and accessory cell function after HIV-1 infection.
Yoo J, Chen H, Kraus T, Hirsch D, Polyak S, George I, Sperber K.
J Immunol. 1996 Aug 1;157(3):1313-20.
PMID: 8757640 [PubMed - indexed for MEDLINE]
Related citations

3. Anti-CD95 (APO-1/Fas) autoantibodies and T cell depletion in human immunodeficiency virus type 1 (HIV-1)-infected children.
Stricker K, Knipping E, Böhler T, Benner A, Krammer PH, Debatin KM.
Cell Death Differ. 1998 Mar;5(3):222-30.
PMID: 10200468 [PubMed - indexed for MEDLINE] Free Article
Related citations

4. Impaired class II expression and antigen uptake in monocytic cells after HIV-1 infection.
Polyak S, Chen H, Hirsch D, George I, Hershberg R, Sperber K.
J Immunol. 1997 Sep 1;159(5):2177-88. Erratum in: J Immunol. 2005 Dec 15;175(12):8444.
PMID: 9278305 [PubMed - indexed for MEDLINE]
Related citations

5. Human immunodeficiency virus type 1 protease inhibitor modulates activation of peripheral blood CD4(+) T cells and decreases their susceptibility to apoptosis in vitro and in vivo.
Sloand EM, Kumar PN, Kim S, Chaudhuri A, Weichold FF, Young NS.
Blood. 1999 Aug 1;94(3):1021-7.
PMID: 10419894 [PubMed - indexed for MEDLINE] Free Article
Related citations

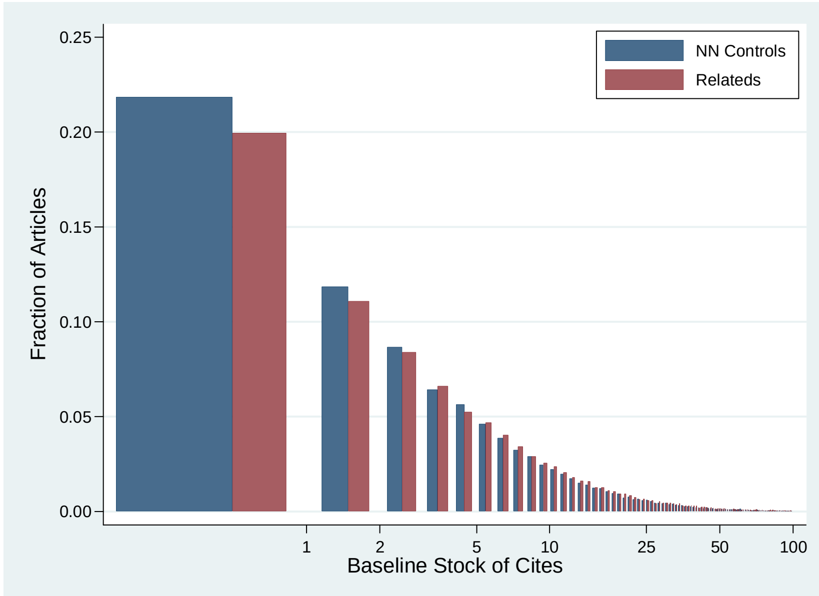
6. Different sensitivity to apoptosis in cells of monocytic or lymphocytic origin chronically infected with human immunodeficiency virus type-1.
Pinti M, Biswas P, Troiano L, Nasi M, Ferraresi R, Mussini C, Vecchiet J, Esposito R, Paganelli R, Cossarizza A.
Exp Biol Med (Maywood). 2003 Dec;228(11):1346-54.
PMID: 14681550 [PubMed - indexed for MEDLINE] Free Article
Related citations

112. Tumor cells induce apoptosis in lymphocytes.
Rubio CA.
Nat Med. 1997 Mar;3(3):253-4. No abstract available.
PMID: 9055844 [PubMed - indexed for MEDLINE]
Related citations

113. HIV-1 entry into renal epithelia.
Husain M, Singhal PC.
J Am Soc Nephrol. 2011 Mar;22(3):399-402. Epub 2011 Feb 18. No abstract available.
PMID: 21335518 [PubMed - indexed for MEDLINE] Free Article
Related citations

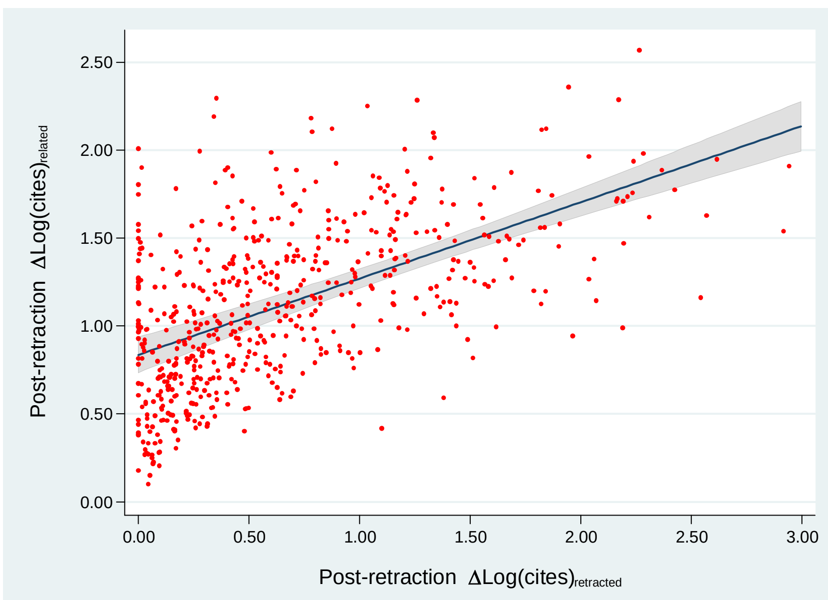
Note: We illustrate the retracting process and that of identifying the related articles through the use of an example. Kirk Sperber, a researcher at Mount Sinai School of Medicine engaged in falsification of research data which resulted in three articles being retracted, including the 1998 Journal of Immunology paper (`pmid 9780201`) referenced above, which was retracted in 2006 (`pmid 17056588`). While the retracting notice argues that the authors were simply guilty of an honest mistake, the investigation performed by NIH's Office of Research Integrity (ORI) concluded that Sperber clearly engaged in scientific misconduct [`http://www.gpo.gov/fdsys/pkg/FR-2008-10-08/pdf/E8-23820.pdf`]. As a result, this observation belongs to the set of 589 retraction in the "absent shoulders" subsample, and we further classify it as one for which the author(s) intended to subvert the scientific process. On the right-hand side panel, one sees that PubMed identifies 112 related articles related to this pmid, but our analysis includes only 77 of these records, since some are not original articles, others are published in 2006 or thereafter, and for yet others, we cannot find a corresponding record in the Web of Science from which we could harvest citation information.

Figure 4: Cumulative Citations at Baseline for Related Articles and their "Nearest-Neighbor" Controls



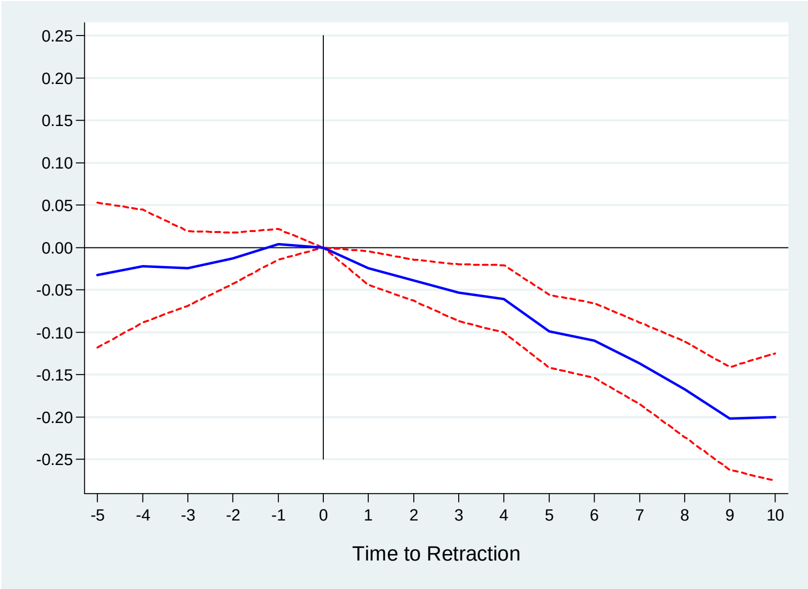
Note: We compute the cumulative number of citations, up to the year that immediately precedes the year of retraction, between 32,699 treated (i.e., related) articles and 65,461 control articles in the "absent shoulders" subsample.

Figure 5: Post-Retraction Period Scatterplot of Changes in Citation Rates for Related Articles and their Associated Retracted Articles



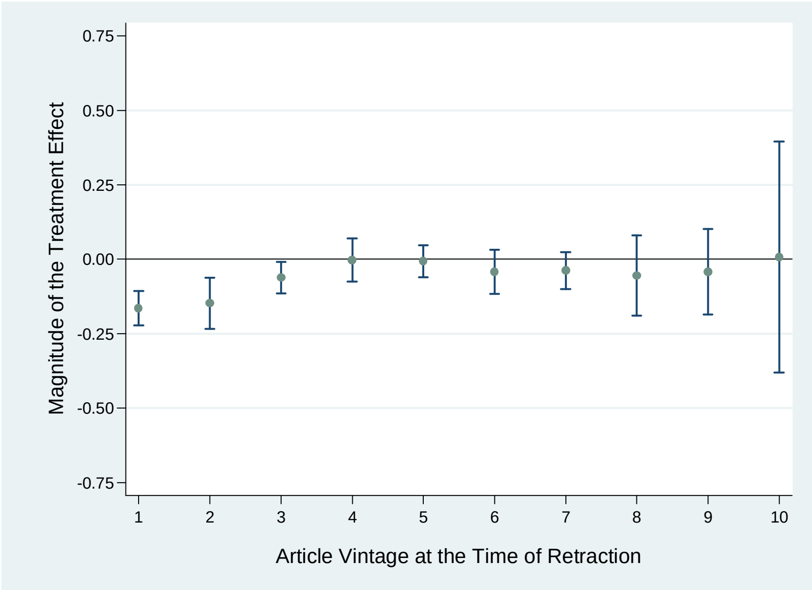
Note: The figure explores the relationship between the post-retraction citation "penalty" suffered by retracted articles and the average change in citation experienced by the set of articles that are related in intellectual space to the retracted articles. The post-retraction period refers to the years between the year of retraction and 2011 (using a two-year fixed window instead of this variable window yields very comparable results). The citation changes are computed by forming the difference in the logs of one plus the number of citations received by each article up until the beginning and the end of the post-retraction window, respectively. The slope of the retraction line is about 0.1, i.e., for every ten citations "lost" by a retracted articles, related articles suffer a penalty of about one citation.

Figure 6: Dynamics of the Retraction Effect on Forward Citation Rates



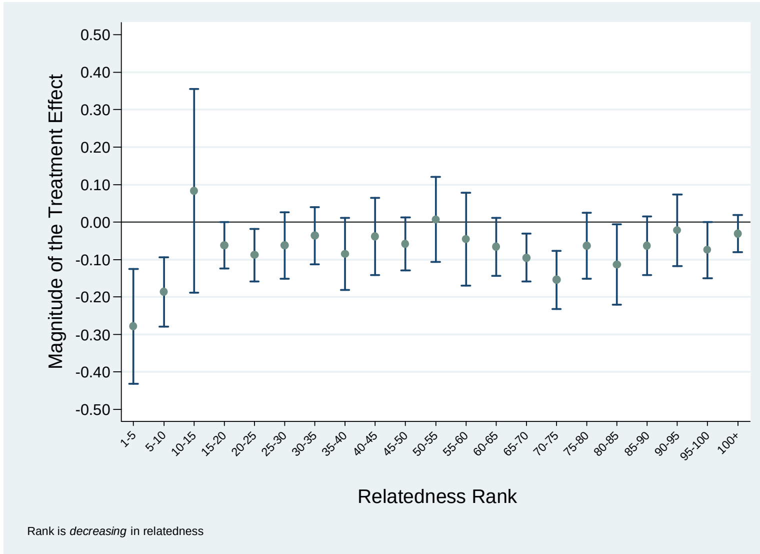
Note: The solid blue lines in the above plot correspond to coefficient estimates stemming from conditional fixed effects quasi-maximum likelihood Poisson specifications in which the citation rates for related articles and their controls are regressed onto year effects, article age indicator variables, as well as 20 interaction terms between treatment status and the number of years before/elapsed since the retraction event (the indicator variable for treatment status interacted with the year of retraction itself is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around case codes) around these estimates is plotted with dashed red lines.

Figure 7: Interaction between the Post-Retraction Treatment Effect and Related Article Vintage at the Time of Retraction



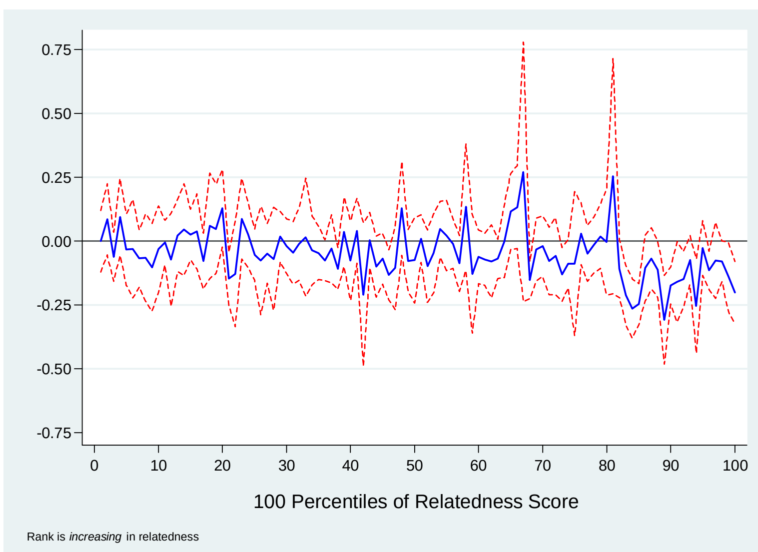
Note: The green circles in the above plot correspond to coefficient estimates stemming from conditional fixed effects QML Poisson specifications in which the citation rates for related articles and their controls are regressed onto year effects, article age indicator variables, as well as interaction terms between the treatment effect and the vintage of each related articles at the time of the retraction. Since related articles in the sample are published between one and ten years before their associated retraction event, there are ten such interaction terms. The 95% confidence interval (corresponding to robust standard errors, clustered around case codes) are denoted by the blue vertical bars.

Figure 8: Interaction between the Post-Retraction Treatment Effect and Relatedness Rank as per PubMed's "Related Article" Algorithm



Note: The green circles in the above plot correspond to coefficient estimates stemming from conditional fixed effects QML Poisson specifications in which the citation rates for related articles and their controls are regressed onto year effects, article age indicator variables, as well as interaction terms between the treatment effect and indicator variables for the relatedness ranking btw. the related article and its associated retraction (as per PubMed's "Related Articles" algorithm). Each circle correspond to five consecutive ranks (e.g., Top 5, Top 6-10, etc.) with all articles receiving a rank above one hundred grouped together in the same bin. The 95% confidence interval (corresponding to robust standard errors, clustered around case codes) are denoted by the blue vertical bars and their caps.

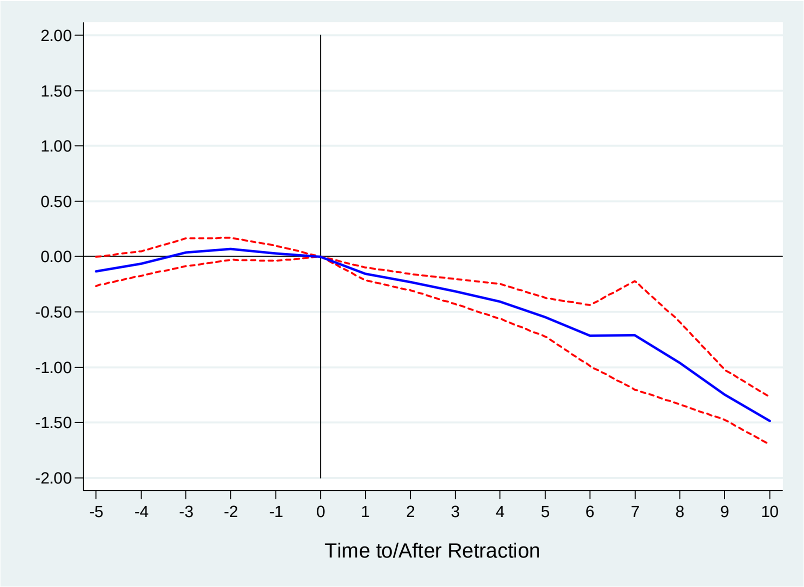
Figure 9: Interaction between the Post-Retraction Treatment Effect and 100 Percentiles of the Relatedness Score



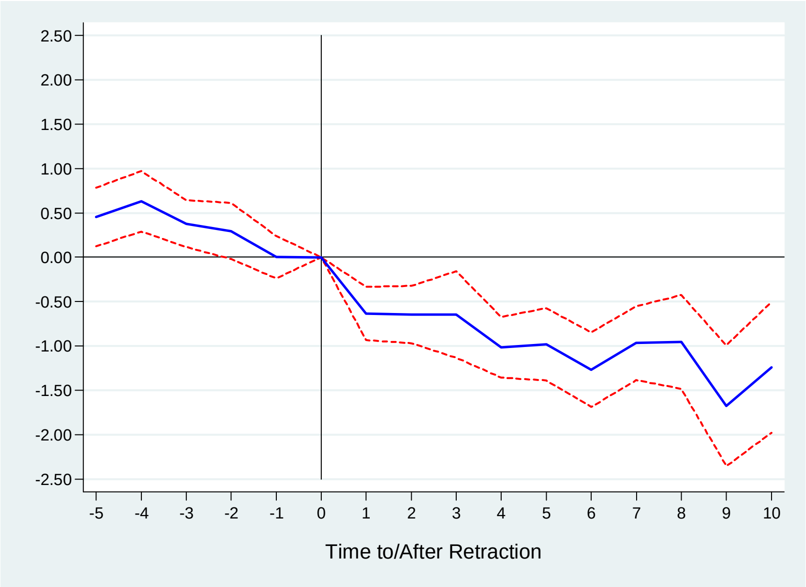
Note: The solid blue lines in the above plot correspond to coefficient estimates stemming from conditional fixed effects QML Poisson specifications in which the citation rates for related articles and their controls are regressed onto year effects, article age indicator variables, as well as 100 interaction terms between the treatment effect and indicator variables for each percentile of the relatedness score between the related article and its associated retraction (as per PubMed's "Related Articles" algorithm). The 95% confidence interval (corresponding to robust standard errors, clustered around case codes) around these estimates is plotted with dashed red lines.

Figure 10
Field-level Dynamics

A. Article Frequency



B. NIH Funding



Notes: The solid blue lines in the above plot correspond to coefficient estimates stemming from conditional fixed effects quasi-maximum likelihood Poisson specifications in which the number of related publications (Panel A) and NIH funding in millions of 2007 dollars (Panel B) associated with a particular source article are regressed onto year effects as well as 20 interaction terms between treatment status and the number of years before/elapsed since the retraction event (the indicator variable for treatment status interacted with the year of retraction itself is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around retraction cases) around these estimates is plotted with dashed red lines; Figure 10A corresponds to a dynamic version of the specification in column (1a) of Table 8, while Figure 10B corresponds to a dynamic version of the specification in column (2b) in the same table.

Appendix I:
Coding of Retraction Reasons

The purpose of this document is to describe the retractions coding scheme that forms the basis of the analysis implemented in the main body of the paper, as well as to provide a method for classification of future retractions. The goal is to reconcile two contradictory objectives: one the one hand, group retractions into a small number of mutually exclusive categories; on the other hand, capture in a meaningful way the inherent heterogeneity in retraction reasons.

The coding scheme has been developed by the authors solely for the purpose of scholarly academic research. The coding of each individual retraction is based on a range of public information sources, ranging from the notice of retraction itself, to entries in the "Retraction Watch" blog, to results of Google searches. No additional information has been gathered from the authors of the retracted papers or others involved in these cases. As such the coding represents an informed judgment of the context in which each retraction event took place, rather than the outcome of a formal investigation. The list of retractions, article characteristics, and reasons can be downloaded from the internet at the following URL: `http://jkrieger.scripts.mit.edu/retractions/`.

Methods Summary: Analysis of retractions indexed by PubMed, published between 1973 and 2008, and retracted before the end of 2009, yielded 13 mutually exclusive "reasons" categories (see list below). In a first step, we assign one of these reasons to each retracted article solely based off the information contained in the retraction notice. In a second step, we assigned a reason to each retracted article based off information in the notice as well as any additional information found through internet sleuthing (e.g., news articles, blogs, press releases, etc.):

We also code each retraction observation based on its validity as a foundation for future research. These "shoulders" categories are ***Strong Shoulders***, ***Shaky Shoulders***, and ***Absent Shoulders. Strong Shoulders*** means that the retraction does not cast doubt on the validity of the paper's underlying claims. A publisher mistakenly printing an article twice, an author plagiarizing someone elses description of a phenomenon, or an institutional dispute about the ownership of samples are all examples where the content of the retracted paper is not in question. ***Shaky Shoulders*** means that the validity of claims is uncertain or that only a portion of the results are invalidated by the retraction. ***Absent Shoulders*** is the appropriate code in fraud cases, as well as in instances where the main conclusions of the paper are compromised by an error.

Lastly, we attempt to discern the level of intentional deceit involved in each case. Deception might involve the paper's actual claims (results, materials method), its attribution of scholarly credit through authorship and citations, or the originality of the work. We use ***No Sign of Intentional Deception*** to code instances where the authors did not intend to deceive, such as in the case of "honest mistakes" or miscommunications. ***Uncertain Intent*** applies where fraud is not firmly established, but negligence or unsubstantiated claims raise questions about an author's motives. The ***Intentional Deception*** code covers cases where falsification, intentional misconduct or willful acts of plagiarism appear to have occured. The "intent" and "shoulders" coding are inherently more subjective than that of the underlying retraction reasons. In fact, there is no simple mapping of the latter into reasons into the former: each shoulders or intent code is assigned based on a thorough review of the available evidence in each case and according to the guidelines below.

The reasons categories capture a combination of context, validity and intent, while the "shoulders" code only pertains to the validity of the article's content, and the "intent" code relates only to intent.[i]

[i]For each category, we mention a small number of PubMed IDs of notices that can serve as good illustrations of the coding choice.

Retraction Reasons:

1. Fake Data. This reason matches with an intuitive definition of scientific fraud. These cases may include the manipulation and misrepresentation of measurements and calculations, as well as the complete fabrication of patients, samples and results. Oftentimes, these retractions will involve an author admitting wrongdoing, or an institutional investigation concluding that the author(s) engaged in falsification of records or results. The existence of an investigation alone does not satisfy the criteria of the "Fake Data" reason code, since it is substantive conclusions of an investigation that distinguish these cases from the cases for which "Questions about Validity" is the more appropriate code (see below). The default "shoulders" code is ***Absent Shoulders*** unless the notice or other sources explicitly communicate that falsification only concerns minor results in the paper. The default "intent" code is ***Intentional Deception*** as intent distinguishes this reason from the "Error/Mistake" category.

2. Error/Mistake. This reason applies where an inaccuracy of claims is central to the retraction notice or case, but there is no evidence of intentional deception or falsification. Contaminated reagents, erroneous interpretations of experimental results, and mislabeled figures are common explanations for the "Error/Mistake" coding. Vague retraction notices that cite "irregularities" and "inaccuracies" in the paper also fall into this category unless we have further evidence linking the authors to suspicion of misconduct. The "shoulders" coding is highly dependent on the context of the error. If the error impacts the main findings of the paper (as is the case when samples or reagents are contaminated), then we assign the ***Absent Shoulders*** code. If the error/mistake only pertains to a minor finding, or if the notice maintains support for the key conclusions, then we use ***Shaky Shoulders*** code. ***Strong Shoulders*** is only appropriate when the mistake clearly has no impact on the veracity of the claims (see #15354845, a letter in which the author refers to the "NSW Companion Animal Registry" rather than the proper name of "Central Animal Records"). The appropriate "intent" code is usually usually ***No Sign of Intentional Deception***, unless the retraction notice explicitly refers to gross negligence (#12949529) or an especially suspicious explanation is given for the error (as in the case where researchers administered primates methamphetamine instead of MDMA — #17176514, #12970544).

3. Could Not Replicate. A common explanation in retracting notices is that the authors (or other researchers) were unable to reproduce the findings of the retracted paper. Some of these notices are vague and give no further insight into the reproducibility issues (#8704228), while others offer vague conjectures regarding the source of the problem without identifying its root cause (#8999116). The "shoulders" coding defaults to ***Shaky Shoulders*** because of the uncertainty surrounding the validity of the original findings. These cases sometimes warrant a ***Absent Shoulders*** classification if the original results are clearly incorrect and central to the paper's claims, even if the authors have not identified the source of the problem. ***No Sign of Intentional Deception*** is the standard code for "Could Not Replicate" retractions. Exceptions might result in the ***Uncertain Intent*** code when a single author is conspicuously left off the retraction notice (#9508700), or the lack or reproducibility seems linked to the work of a single author (#1364942).

4. Plagiarism. Plagiarism cases are usually easy to spot in retraction notices. Copying or closely imitating text, or using someone else's figures or images without assigning the appropriate credit are typical examples. In some cases, the notice may not explicitly accuse the authors of plagiarism, but will highlight that "copyright infringement" (#19021584) or "close resemblance" with another paper is the reason for retraction. ***Strong Shoulders*** characterizes most retractions in this category because the offense is copying rather than mistake or falsification. However, ***Shaky Shoulders*** may be appropriate when the results section of the paper contains plagiarized content, calling into question the accuracy of the claims (#19264925). Intentionality is usually assumed, though cases of carelessness (#11023382), language issues (#14667944), or miscommunication may warrant an ***Uncertain Intent*** designation.

5. Fake Data & Plagiarism. This category covers cases where fraud involved both fake data and plagiarism, as independently defined above. These cases will likely involve an investigation that finds the author guilty of falsification and plagiarism (#12411512, #12833069, #19575288). If a retraction meets the criteria of "Fake Data & Plagiarism" then ***Absent Shoulders*** and ***Intentional Deception*** are the logical complementary codes.

6. Duplication. The important criterion for "Duplication" is that the authors copied from themselves. Most of the articles in this category already appeared in another journal before the second journal realized that the entire article is an exact duplicate or virtually identical to an article by the same authors in a different journal (#12589830). Some of the "Duplication" cases are not entirely republished articles, but will reproduce important content, such as data, charts and conclusions (#15580694). As with plagiarism cases, these cases are assigned the ***Strong Shoulders*** code by default, but may fall into the ***Shaky Shoulders*** bucket when meaningful differences exist between the duplicated article and its original version (#1930642). The "intent" coding follows a similar logic, with ***Intentional Deception*** being the primary classification. Yet, ***Uncertain Intent*** sometimes is the more logical choice when duplication resulted from an apparent miscommunication (#17047133, #16683328).

7. Questions about Validity. This category captures retraction cases associated with vague misconduct allegations (#118049464), suspicious "irregularities" (#118560433), and "questionable" data (#118951275). The hallmark of these retraction notices is that they obfuscate the nature of the misconduct. The vague nature of this category's notices makes ***Shaky Shoulders*** and ***Uncertain Intent*** the frequent choice for complementary codes.

8. Author Dispute. These cases involve disagreements between authors about content, credit, and permission. Often, these different types of disputes will be combined (#14723797, #19727599). Paper submission without the consent of coauthors is the most common underlying reason for this code. Unless warranted by information gained through sleuthing, ***Shaky Shoulders*** is the appropriate code for "Author Dispute" cases — most disputes stem from conflicts surrounding credit attribution and the verification of results, rather than outright fraud. ***Intentional Deception*** is the prevalent intent code in "Author Dispute" cases, though exceptions do exist (#17081259; #16003050).

9. Lack of Consent/IRB Approval. This category includes cases where the authors did not get IRB approval or did not secure patient informed consent before conducting their study. Ambiguous cases of "ethics violations" (#19819378, #18774408) also fall into the "Lack of Consent/IRB Approval" category. The default "shoulders" code is ***Shaky Shoulders***. ***Strong Shoulders*** may be appropriate if there is evidence indicating that the authors believed they had IRB approval (#14617761), or that the paper's results are devoid of fraud/deception (#16832233). Determining the level of intent is less straightforward for this category. In general, ethics violations count as ***Intentional Deception***, but uncertainty about author intent may warrant other coding choices. For example, the authors may have erroneously thought they received IRB approval (#14617761), or approval may have been officially obtained only after the authors completed the study (#16842490).

10. Did Not Maintain Proper Records. Although the dataset only has three retractions that fall into this category, we include it as distinct retracting reason. The defining characteristic of this category is absence of proper data records. With proper records, the scientific community could better determine the reliability of the claims contained in these papers. ***Shaky Shoulders*** and ***Uncertain Intent*** are the proper complementary codes.

11. Publisher Error. Retractions occasionally stem from publisher mistakes rather than author misconduct or error. The associated notices establish that the publisher is solely responsible for the error, which is usually a duplicate publication (#17452723, #15082607) or printing of an earlier draft (#19662582, #15685781). ***Strong Shoulders*** is a natural fit for publisher errors resulting from duplicates, while ***Shaky Shoulders*** is appropriate when the journal prints the wrong draft. By definition, the proper intent coding is ***No Sign of Intentional Deception.***

12. NOT ENOUGH INFORMATION TO CLASSIFY or MISSING. The essential difference between these two categories is that we have a notice for the former and do not have a notice for the latter. “Not Enough Information to Classify” implies that the notice is so vague that we cannot assign another code. Such retractions will usually take on the form of a simple statement such as *“This article has been withdrawn at the request of the authors”* (#19785092). The default “shoulders” code is ***Shaky Shoulders***, but ***Absent Shoulders*** may be preferable when the notice mentions inaccuracies (#9786782, #7566837). The lack of information in these cases makes ***Uncertain Intent*** the proper intent code.

Appendix II: PubMed Related Citations Algorithm [PMRA]

The following paragraphs were extracted from a brief description of PMRA:[ii]

> *The neighbors of a document are those documents in the database that are the most similar to it. The similarity between documents is measured by the words they have in common, with some adjustment for document lengths. To carry out such a program, one must first define what a word is. For us, a word is basically an unbroken string of letters and numerals with at least one letter of the alphabet in it. Words end at hyphens, spaces, new lines, and punctuation. A list of 310 common, but uninformative, words (also known as stopwords) are eliminated from processing at this stage. Next, a limited amount of stemming of words is done, but no thesaurus is used in processing. Words from the abstract of a document are classified as text words. Words from titles are also classified as text words, but words from titles are added in a second time to give them a small advantage in the local weighting scheme. MeSH terms are placed in a third category, and a MeSH term with a subheading qualifier is entered twice, once without the qualifier and once with it. If a MeSH term is starred (indicating a major concept in a document), the star is ignored. These three categories of words (or phrases in the case of MeSH) comprise the representation of a document. No other fields, such as Author or Journal, enter into the calculations.*
>
> *Having obtained the set of terms that represent each document, the next step is to recognize that not all words are of equal value. Each time a word is used, it is assigned a numerical weight. This numerical weight is based on information that the computer can obtain by automatic processing. Automatic processing is important because the number of different terms that have to be assigned weights is close to two million for this system. The weight or value of a term is dependent on three types of information: 1) the number of different documents in the database that contain the term; 2) the number of times the term occurs in a particular document; and 3) the number of term occurrences in the document. The first of these pieces of information is used to produce a number called the global weight of the term. The global weight is used in weighting the term throughout the database. The second and third pieces of information pertain only to a particular document and are used to produce a number called the local weight of the term in that specific document. When a word occurs in two documents, its weight is computed as the product of the global weight times the two local weights (one pertaining to each of the documents).*
>
> *The global weight of a term is greater for the less frequent terms. This is reasonable because the presence of a term that occurred in most of the documents would really tell one very little about a document. On the other hand, a term that occurred in only 100 documents of one million would be very helpful in limiting the set of documents of interest. A word that occurred in only 10 documents is likely to be even more informative and will receive an even higher weight.*
>
> *The local weight of a term is the measure of its importance in a particular document. Generally, the more frequent a term is within a document, the more important it is in representing the content of that document. However, this relationship is saturating, i.e., as the frequency continues to go up, the importance of the word increases less rapidly and finally comes to a finite limit. In addition, we do not want a longer document to be considered more important just because it is longer; therefore, a length correction is applied.*
>
> *The similarity between two documents is computed by adding up the weights of all of the terms the two documents have in common. Once the similarity score of a document in relation to each of the other documents in the database has been computed, that document's neighbors are identified as the most similar (highest scoring) documents found. These closely related documents are pre-computed for each document in PubMed so that when one selects Related Articles, the system has only to retrieve this list. This enables a fast response time for such queries.*

We illustrate the use of PMRA with an example taken from our sample. Amitav Hajra is a former University of Michigan graduate student who falsified data in three papers retracted in 1996. One of Hajra's retracted papers (PubMed ID #7651416) appeared in the September 1995 issue of *Molecular and Cellular Biology* and lists 27 MeSH terms. Its 10^{th} most related paper (PubMed ID #8035830), according to the PMRA algorithm, appeared in the same journal in August 1994 and has 23 MeSH terms, 10 of which overlap with the Hajra article. These terms include common terms such as "Mice" and "DNA-Binding Proteins/genetics" as well as more specific keywords including "Core Binding Factor Alpha Subunits," "Neoplasm Proteins/metabolism," and "Transcription Factor AP-2." In contrast, one of the nearest neighbor to the related article (PubMed ID

[ii] Available at `http://ii.nlm.nih.gov/MTI/related.shtml`

#8035831) is tagged by 17 MeSH terms, of which only two terms ("Base Sequence" and "Molecular Sequence Data") overlap with those listed by PubMed for the retraction. Even though all three articles came from the same journal, the overlap in MeSH terms stronly suggests that the related paper is closer in intellectual space to the retraction than is its nearest neighbor control.

PMRA and MeSH Terms Overlap — An Example

Retracted Article PMID #7651416	Related Article PMID #8035830	Related Article Nearest Neighbor PMID #8035831
3T3 Cells	Animals	Amino Acid Sequence
Animals	Base Sequence	Base Sequence
Base Sequence	Binding Sites	Biological Clocks*
Cell Transformation, Neoplastic/genetics*	Cloning, Molecular	Cell Cycle*
Chromosome Inversion	Core Binding Factor Alpha 1 Subunit	Cell Size
Chromosomes, Human, Pair 16/genetics*	Core Binding Factor alpha Subunits	Fungal Proteins/genetics*
Core Binding Factor Alpha 2 Subunit	Core Binding Factor beta Subunit	Gene Expression Regulation, Fungal*
Core Binding Factor alpha Subunits	Core Binding Factors	Genes, Fungal*
Core Binding Factor beta Subunit	DNA-Binding Proteins/genetics*	Glycine
DNA-Binding Proteins/genetics	Gene Expression Regulation, Enzymologic*	Molecular Sequence Data
DNA-Binding Proteins/metabolism*	Leukocyte Elastase	RNA, Fungal/genetics
Gene Expression Regulation, Neoplastic	Leukocytes/enzymology*	RNA, Messenger/genetics
Humans	Mice	Repetitive Sequences, Nucleic Acid
Leukemia, Myeloid/etiology	Molecular Sequence Data	Restriction Mapping
Leukemia, Myeloid/genetics*	Neoplasm Proteins*	Saccharomyces cerevisiae/cytology
Mice	Nuclear Proteins/metabolism	Saccharomyces cerevisiae/genetics*
Models, Biological	Oligodeoxyribonucleotides/chemistry	Threonine
Molecular Sequence Data	Pancreatic Elastase/genetics*	
Mutation	Peroxidase/genetics*	
Neoplasm Proteins/genetics	Promoter Regions, Genetic*	
Neoplasm Proteins/metabolism*	RNA, Messenger/genetics	
Proto-Oncogene Proteins*	Transcription Factor AP-2	
Transcription Factor AP-2	Transcription Factors/genetics*	
Transcription Factors/genetics		
Transcription Factors/metabolism*		
Transcriptional Activation		
Transfection		