Does Science Advance One Funeral at a Time?

By Pierre Azoulay, Christian Fons-Rosen, and Joshua S. Graff Zivin*

We examine how the premature death of eminent life scientists alters the vitality of their fields. While the flow of articles by collaborators into affected fields decreases after the death of a star scientist, the flow of articles by non-collaborators increases markedly. This surge in contributions from outsiders draws upon a different scientific corpus and is disproportionately likely to be highly cited. While outsiders appear reluctant to challenge leadership within a field when the star is alive, the loss of a luminary provides an opportunity for fields to evolve in new directions that advance the frontier of knowledge.
JEL: I23, J24, O31.
Keywords: economics of science, scientific fields, superstars, invisible college, cumulative knowledge production.

> "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."
>
> Max Planck
> *Scientific Autobiography and Other Papers*

Whether manna from heaven or the result of the purposeful application of research and development, technological advances play a foundational role in all modern theories of economic growth (Solow 1957, Romer 1990, Aghion and Howitt 1992). Only in the latter part of the nineteenth century, however, did technological progress start to systematically build upon scientific foundations (Mokyr

* Azoulay: MIT Sloan School of Management and NBER, 100 Main Steet—E62-487, Cambridge, MA 02142 (e-mail: pazoulay@mit.edu); Fons-Rosen: UC Merced and CEPR, Department of Economics, 5200 N. Lake Road, Merced, CA 95343 (e-mail: cfons-rosen@ucmerced.edu); Graff Zivin: UCSD and NBER, School of Global Policy & Strategy, 9500 Gilman Drive, La Jolla, CA 92093 (e-mail: jgraffzivin@ucsd.edu). Azoulay and Graff Zivin acknowledge the financial support of the National Science Foundation through its SciSIP Program (Award SBE-1460344). Fons-Rosen acknowledges financial support from the Spanish Ministry of Economy and Competitiveness through a grant (ECO-2014-55555-P) and through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563). The project would not have been possible without Andrew Stellman's extraordinary programming skills (www.stellman-greene.com). Mikka Rokkanen provided additional research assistance. We thank Heidi Williams, Xavier Jaravel, Danielle Li, Sameer Srivastava, Scott Stern, Bruce Weinberg, and seminar audiences at the NBER, UC Berkeley, National University of Singapore, and Stanford University for useful discussions. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The usual disclaimer applies.

1992, 2002). Economists—in contrast to philosophers, historians, and sociologists (Kuhn 1962, Shapin 1996, Merton 1973)—have devoted surprisingly little effort to understanding the processes and institutions that shape the evolution of science.[1] How do researchers identify problems worthy of study and choose among potential approaches to investigate them?

Presumably these choices are driven by a quest for recognition and scientific glory, but the view that scientific advances are the result of a pure competition of ideas—one where the highest quality insights inevitably emerge as victorious—has long been considered a Panglossian but useful foil (Kuhn 1962; Akerlof and Michaillat 2017). Indeed, the provocative quote from Max Planck in the epigraph of this paper underscores that even the most celebrated scientist of his era understood that the pragmatic success of a scientific theory does not entirely determine how quickly it gains adherents, or its longevity.

Can the idiosyncratic stances of individual scientists do much to alter, or at least delay, the course of scientific advance? Perhaps for the sort of scientific revolutions that Planck—the pioneer of quantum mechanics—likely had in mind, but the proposition that established scientists are slower than novices in accepting paradigm-shifting ideas has received little empirical support whenever it has been put to the test (Hull, Tessner, and Diamond 1978; Gorham 1991; Levin, Stephan, and Walker 1995). Paradigm shifts are rare, however, and their very nature suggests that once they emerge, it is exceedingly costly to resist or ignore them. In contrast, "normal" scientific advance—the regular work of scientists theorizing, observing, and experimenting within a settled paradigm or explanatory framework—may be more susceptible to political jousting. The absence of new self-evident and far-reaching truths means that scientists must compete in a crowded intellectual landscape, sometimes savagely, for the supremacy of their ideas (Bourdieu 1975).

In this paper, we use a difference-in-differences setup to test "Planck's Principle" in the context of academic biomedical research, an enormous domain which has been the province of normal scientific change ever since the "central dogma" of molecular biology (Crick 1970) emerged as a unifying description of the information flow in biological systems. Specifically, we examine how the premature death of 452 eminent scientists alter the vitality (measured by publication rates and funding flows) of subfields in which they actively published in the years immediately preceding their passing, compared to matched control subfields. In contrast with prior work that focused on collaborators (Azoulay, Graff Zivin, and Wang 2010; Oettl 2012; Jaravel, Petkova, and Bell 2018; Mohnen 2018), our work leverages new tools to define scientific subfields which allows us to expand our focus to the response by scientists who may have similar intellectual interests with the deceased stars without ever collaborating with them.

To our surprise, it is not competitors from within a subfield that assume the

[1] A notable exception is the theoretical model of scientific revolutions developed by Bramoullé and Saint-Paul (2010).

mantle of leadership, but rather entrants from other fields that step in to fill the void created by a star's absence. Importantly, this surge in contributions from outsiders draws upon a different scientific corpus and is disproportionately likely to be highly cited. Thus, consistent with the contention by Planck, the loss of a luminary provides an opportunity for fields to evolve in novel directions that advance the scientific frontier. The rest of the manuscript is dedicated to elucidating the mechanisms responsible for this phenomenon.

It does not appear to be the case that stars use their influence over financial or editorial resources to block entry into their fields, but rather that the very prospect of challenging a luminary in the field serves as a deterrent for entry by outsiders. Indeed, most of the entry we see occurs in those fields that lost a star who was especially accomplished. Even in those fields that have lost a particularly bright star, entry can still be regulated by key collaborators left behind. We find suggestive evidence that this is true in fields that have coalesced around a narrow set of techniques or ideas or where collaboration networks are particularly tight-knit. We also find that entry is more anemic when key collaborators of the star are in positions that allow them to limit access to funding or publication outlets to those outside the club that once nucleated around the star.

To our knowledge, this manuscript is the first to examine the dynamics of scientific evolution using the standard empirical tools of applied microeconomics.[2] We conceptualize the death of eminent scientists as shocks to the structure of the intellectual neighborhoods in which they worked several years prior to their death, and implement a procedure to delineate the boundaries of these neighborhoods in a way that is scalable, transparent, and does not rely on *ad hoc* human judgment. The construction of our dataset relies heavily on the *PubMed Related Citations Algorithm* [PMRA], which groups scientific articles into subfields based on their intellectual content using abstract words, title words, and very detailed keywords drawn from a controlled vocabulary thesaurus curated by the National Library of Medicine. As such, we are able to delineate circumscribed areas of scientific inquiry whose boundaries are not defined by shared training, collaboration, or citation relationships.

In addition to providing evidence regarding a central question for scholars studying the scientific process, our paper is among the very few economic studies that attend to the ways in which scientists position themselves in intellectual space (cf. Borjas and Doran [2015a, 2015b] and Myers [2018] for other notable examples). As such, our work can be understood as integrating the traditional concerns of economists—understanding how incentives and institutions influence the rate of

[2]Considerable work outside of economics has examined the evolution of scientific fields through network and community detection techniques (e.g., Rosvall & Bergstrom 2008; Börner, Chen, and Boyack 2003; cf. Fortunato and Hric (2016) for a review of this fast-evolving research area). These approaches rely on collaboration or citation links to define the vertices of the knowledge network used to partition a scientific space into subfields. While social scientists have utilized these techniques to explain a wide range of phenomena (e.g., Foster, Rzhetsky, and Evans 2015), these approaches are less well-suited to our setting where citation and collaboration are among the primary outcomes of interest.

knowledge production or diffusion—with those of cognate disciplines such as sociology and philosophy, who have traditionally taken the *direction* of scientific change as the central problem to be explained.

The rest of the paper proceeds as follows. In the next section, we examine the institutional context and lay out our broad empirical strategy. In section II, we then turn to data, methods and descriptive statistics. We report the results in section III. Section IV concludes by outlining the implications of our findings for future work.

I. Institutional Context and Empirical Design

Our empirical analyses are centered on the academic life sciences. The merits of this focus are several fold. First, the field has been an important source of scientific discovery over the past half century. Many modern medical therapies can trace their origins to research conducted in academic laboratories (Sampat and Lichtenberg 2011; Azoulay, Li, and Sampat 2017). These discoveries, in turn, have generated enormous health and welfare gains for economies around the world.

Second, the life science research workforce is large and specialized. The Faculty Roster of the Association of American Medical Colleges lists more than 200,000 faculty members employed in U.S. medical schools and academic medical centers in 2015.[3] Moreover, scientific discoveries over the past half-century have greatly expanded the knowledge frontier, necessitating increasing specialization by researchers and a greater role for collaboration (Jones 2009). If knowledge and techniques remain at least partially tacit long after their initial discovery, tightly-knit research teams may be able to effectively control entry into intellectual domains. The size and maturity of this sector, including its extensive variety of narrowly-defined subfields, makes it an ideal candidate for an inquiry into the determinants of the direction of scientific effort in general, and how it is influenced by elite scientists in particular.

Third, the academic research setting also offers the practical benefits of an extensive paper trail of research inputs, outputs, and collaboration histories. On the input side, reliance of researchers on one agency for the majority of their funding raises the possibility that financial gatekeeping by elite scientists could be used to regulate entry into scientific fields. Data on NIH funding at the individual level, as well as membership in "study sections" (the peer-review panels that evaluate the scientific merits of grant applications) will allow us to examine such concerns directly. Most importantly for our study, the principal output of researchers—publications—are all tagged by a controlled vocabulary of keywords managed by the National Library of Medicine. This provides the raw material that helps delineate scientific subfields without appealing to citation linkages or

[3]This figure excludes life science academics employed in graduate schools of arts and science or other non-medical school settings such as MIT, Rockefeller University, The Salk Institute, UC Berkeley, the intramural campuses of NIH, etc.

collaborative relationships (the specifics of this process are described in detail in Section II.B and Appendix C in the supplementary online material).

These many virtues, however, may come at the expense of generalizability. While the life sciences span a wide range of research styles—from small-team data-driven epidemiology, to medium-size laboratories under the helm of a single principal investigator, to large-scale multi-institution clinical trials—most biomedical researchers cluster topically and socially in small, quasi-independent subfields. This broad domain seldom features exceedingly small research teams (as in pure mathematics) or "big science" efforts where capital needs are so extensive and specialized as to fully consolidate the field into a single or a handful of large authorship teams (as in high-energy particle physics, e.g., Aad et al. 2015). As such, one should refrain from applying our findings to other fields of science where the structure of collaborative efforts and the degree of intellectual clustering are likely to generate different patterns of succession, compared to those observed in the life sciences.

Accounts by practicing scientists indicate that collaboration plays a large role in both the creation and diffusion of new ideas (Reese 2004), and historians of science have long debated the role of controversies and competition in shaping the direction of scientific progress and the process through which new subfields within the same broad scientific paradigm are born and grow over time (Hull 1988; Morange 1998; Shwed and Bearman 2010). Our study presents a unique opportunity to test some of their insights in a way that is more systematic and can yield generalizable insights on the dynamics of field evolution.

II. Empirical Design, Data, and Descriptive Statistics

Below, we provide a detailed description of the process through which the matched scientist/subfield dataset used in the econometric analysis was assembled. We begin by describing the criteria used to select our sample of superstar academics, with a particular focus on "extinction events"; the set of subfields in which these scientists were active prior to their death and the procedure followed to delineate their boundaries. Finally, we discuss the matching procedure implemented to identify control subfields associated with eminent scientists who did not pass away but are otherwise similar to our treatment group.

A. Superstar sample

Our basic approach is to rely on the death of "superstar" scientists as a lever to estimate the extent to which the production of knowledge in the fields in which they were active changes after their passing. The study's focus on the scientific elite can be justified both on substantive and pragmatic grounds. The distribution of publications, funding, and citations at the individual level is extremely skewed (Lotka 1926; de Solla Price 1963) and only a tiny minority of scientists contribute, through their published research, to the advancement of science (Cole and Cole

1972). Stars also leave behind a corpus of work and colleagues with a stake in the preservation of their legacy, making it possible to trace back their careers, from humble beginnings to wide recognition and acclaim.

The elite academic life scientist sample includes 12,935 individuals, which corresponds to roughly 5% of the entire relevant labor market. In our framework, a scientist is deemed elite if they satisfy at least one of the following criteria for cumulative scientific achievement: (1) highly funded scientists; (2) highly cited scientists; (3) top patenters; and (4) members of the National Academy of Sciences or of (5) the National Academy of Medicine. Since these criteria are based on extraordinary achievement over an entire scientific career, we augment this sample using additional criteria to capture individuals who show great promise at the early and middle stages of their scientific careers (so-called "shooting stars"). These include: (6) NIH MERIT awardees; (7) Howard Hughes Medical Investigators; and (8) early career prize winners. Appendix A provides additional details regarding these metrics of "superstardom" and explores the sensitivity of our core set of results to the type of scientists ("cumulative stars" vs. "shooting stars") included in the sample.

For each scientist, we reconstruct their career from the time they obtained their first position as independent investigators (typically after a postdoctoral fellowship) until 2006. Our dataset includes employment history, degree held, date of degree, gender, and department affiliations as well as complete list of publications, patents and NIH funding obtained in each year by each scientist.[4]

The 452 scientists who pass away prematurely, and who are the particular focus of this paper, constitute a subset of this larger pool of 12,935. To be included in our sample, their deaths must intervene between 1975 and 2003 (this allows us to observe at least three years' worth of scientific output for every subfield after the death of a superstar scientist). Although we do not impose any age cutoff, the median and mean age at death is 61 with 85% of these scientists having passed away before the age of 70 (we explore the sensitivity of our results to the age at death in Appendix E). We also require evidence, in the form of published articles and/or NIH grants, that these scholars were still in a scientifically active phase of their career in the period just preceding their death (this is the narrow sense in which we deem their deaths to have occurred prematurely).

Within this sample, 229 (51%) of these scientists pass away after a protracted illness, whereas 185 (41%) die suddenly and unexpectedly. We were unable to ascertain the particular circumstances of 37 (8.20%) death events.[5] Table 1 provides descriptive statistics for the deceased superstar sample. The median star received her degree in 1957 and died at the age of 61. 40% of the stars hold an

[4]Appendix B details the steps taken to ensure that the list of publications is complete and accurate, even in the case of stars with frequent last names. Though we apply the term of "star" or "superstar" to the entire group, there is substantial heterogeneity in intellectual stature within the sample (see Table 1).

[5]Table A3 in Appendix A provides the full list of deceased superstars, together with their year of birth and death, cause of death, institutional affiliation at the time of their passing, and a short description of their research expertise.

MD degree (as opposed to a PhD or MD/PhD), and 90% of them are male. On the output side, the stars each received an average of roughly 16.6 million dollars in NIH grants, and published 138 papers that garnered 8,341 citations over the course of their careers (as of 2015).

B. Delineating Research Fields

The source of the publication data is *PubMed*, an online resource from the National Library of Medicine that provides fast, free, and reliable access to the biomedical research literature. *PubMed* indexes more than 40,000 journals within the life sciences.

To delineate the boundaries of the research fields in which each deceased star was active, we develop an approach based on topic similarity between each article where the star appeared as a last author in a window of five years prior to her death, and the rest of the scientific literature.[6] Specifically, we use the *PubMed Related Citations Algorithm* (Lin and Wilbur 2007) which relies heavily on Medical Subject Headings (MeSH), but not in any way on citation or collaboration linkages.

MeSH terms constitute 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 publication by professional indexers who focus solely on their scientific content. That said, the *PubMed* Related Citations Algorithm (hereafter PMRA) also uses title and abstract words as inputs, which are selected by the authors, and may reflect their aspirations. While this raises the possibility that our subfield definitions are not impervious to social influences, it does offer one advantage, namely that our subfield boundaries can quickly reflect the emergence of new terms whose inclusion in the official MeSH thesaurus will occur with some lag.[7] Regardless, as will become clear in the next section, our difference-in-differences design alleviates the concern that idiosyncratic features of PMRA might affect our conclusions, since these would influence treatment and control subfields in a symmetric fashion.

We then use the "Related Articles" function in *PubMed* to harvest journal articles that are intellectually proximate to the star scientists' own papers in the last five years of her life.[8] Appendix C describes the algorithm in more detail and

[6] A robust social norm in the life sciences systematically assigns last authorship to the principal investigator, first authorship to the junior author who was responsible for the conduct of the investigation, and apportions the remaining credit to authors in the middle of the authorship list, generally as a decreasing function of the distance from the extremities (Zuckerman 1968; Nagaoka and Owan 2014). Only in the case of last authorship can we unambiguously associate the star with a subfield.

[7] Importantly, defining subfields as isomorphic to the set of articles related (in a PMRA-sense) to a source article does not imply a fixed number of articles per subfield. On the contrary, PMRA-generated subfields can be of arbitrary large size. In Appendix C, we document the variation in subfield size and explore the sensitivity of our results to alternate subfield definitions, including those that exclude potentially endogenous intellectual linkages.

[8] To facilitate the harvesting of *PubMed*-related records on a large scale, we have developed an

performs extensive robustness checks. In particular, we verify that the cutoff rules used by PMRA to generate a set of intellectual neighbors for a given source article do not induce treated subfields to exhibit idiosyncratic truncation patterns—from above or from below—compared to control subfields. Using a tunable version of PMRA, we also assess the robustness of our core results to manipulations of these cutoff rules. Reassuringly, our results are qualitatively similar regardless of the rule employed.

To fix ideas, consider "The transcriptional program of sporulation in budding yeast" [PubMed ID 9784122], an article published in the journal *Science* in 1998 originating from the laboratory of Ira Herskowitz, an eminent UCSF biologist who died in 2003 from pancreatic cancer. As can be seen in Appendix Figure C4, PMRA returns 72 original related journal articles for this source publication. Some of these intellectual neighbors will have appeared before the source to which they are related, whereas others will have only been published after the source. Some will represent the work of collaborators, past or present, of Herskowitz's, whereas others will represent the work of scientists he may never have come in contact with during his life, much less collaborated with. The salient point is that nothing in the process through which these related articles are identified biases us towards (or away from) articles by collaborators, frequent citers of Herskowitz's work, or co-located researchers.

Consider now the second most-related article to Herskowitz's *Science* paper listed in Figure C4, "Phosphorylation and maximal activity of *Saccharomyces cerevisiae* meiosis-specific transcription factor Ndt80 is dependent on Ime2." Figure C5 in Appendix C displays the MeSH terms that tag this article along with its source. As a byproduct, PMRA also provides a cardinal dyadic measure of intellectual proximity between each related article and its associated source article. In this particular instance, the relatedness score of "Phosphorylation..." is 94%, whereas the relatedness score for the most distant related article in Figure C4, "Catalytic roles of yeast..." is only 62%.

In the five years prior to his death (1998-2002), Herskowitz was the last author on 12 publications, the publications most closely associated with his position as head of a laboratory. For each of these source publications, we treat the set of publications returned by PMRA as constituting a distinct subfield, and we create a subfield panel dataset by counting the number of related articles in each of these subfields in each year between 1975 and 2006. An important implication of this data construction procedure is that the subfields we delineate are quite limited in scope. One window into the degree of intellectual breadth for subfields is to gauge the overlap between the articles that constitute any pair of subfields associated with the same star. In the sample, the 452 deceased stars account for 3,076 subfields, and 21,661 pairwise combination of subfields (we are only considering

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/`. Prior research leveraging the intellectual linkages between articles generated by PMRA include Azoulay et al. (2015), Azoulay et al. (2019), and Myers (2018).

pairs of subfields associated with the same individual star). Appendix Figure C6 displays the histogram for the distribution of overlap, which is extremely skewed. A full half of these pairs exhibit exactly zero overlap, whereas the mean of the distribution is 0.06. To find pairs of subfields that display substantial amounts of overlap (for example, half of the articles in subfield 1 also belong in subfield 2), one must reach far into the right tail of the distribution, specifically, above the 98^{th} percentile.

As such, the subfields we delineate are relatively self-contained. Performing the analysis at the level of the subfield—rather than lumping together all the subfields of an individual star—will provide us with an opportunity to exploit variation in the extent of participation of the star within each of her subfields. We will also check the validity of the main results when rolling the data up from the subfield level to the star level in Appendix F. Finally, since even modest amounts of overlap entail that the observations corresponding to the subfields of individual stars will not be independent in a statistical sense, we will cluster standard errors at the level of the star scientist.[9]

C. Identification Strategy

Given our interests in the effect of superstar death on entry into scientific subfields, our empirical strategy is focused on changes in published research output after the superstar passes away, relative to when she was still alive. To ensure that we are estimating the effect of interest and not some other influence that is correlated with the passage of time, our specifications include age and period effects, as is the norm in studies of scientific productivity (Levin and Stephan 1991). These temporal controls are tantamount to using subfields that lost a superstar in earlier or later periods as an implicit control group when estimating entry into subfields that currently experienced the death of a superstar. If the death of a superstar only represented a one-time shift in the level of entry into the relevant subfields, this would not be problematic. But if these unfortunate events affect trends—and not simply levels—of scientific activity, this approach may not suffice to filter out the effect of time-varying omitted variables, even when flexible age and calendar time controls are included in the econometric specification. One tangible concern about time-varying effects relates to the life cycle of subfields, where productive potential may initially increase over time before peaking and then slowly declining.

To mitigate this threat to identification, our preferred empirical strategy relies on the selection of a matched scientist/subfield for each treated scientist/subfield. These control observations are culled from the universe of subfields in which superstars who do not die are active (see Section II.A and Appendix D). Combining the treated and control samples enables us to estimate the effect of superstar death

[9]The compactness of these subfields likely reflect the technology of research within the life sciences, a similar exercise performed in a different domain of science, particularly those characterized by large collaborative projects, might well result in subfields with substantially more overlap.

in a difference-in-differences framework. Appendix Figure D1 illustrates the procedure used to identify control subfields in the particular case of the Herskowitz publication highlighted above.

We begin by looking at all the articles that appeared in the same journal and in the same year as the treated source articles. From this set of articles, we keep only those that have one of the still-living superstars in the last authorship position. Then, using a "coarsened exact matching" procedure detailed in Appendix D, the control source articles are selected such that (1) the number of authors in the treated and control are approximately similar; (2) the age of the treated and control superstars differ by no more than five years; and (3) the number of citations received by the treated and source article are similar. For the Herskowitz/"sporulation in budding yeast" pair, we can select 10 control articles in this way. All of these controls were also published in *Science* in 1998, and have between five and seven authors. One of these controls is "Hepatitis C Viral Dynamics in Vivo...," whose last author is Alan Perelson, a biophysicist at Los Alamos National Lab. Perelson and Herskowitz obtained their PhD only a year apart. The two papers had received 514 and 344 citations respectively by the end 2003. Though this is a large difference, this places both well above the 99^{th} percentile of the citation distribution for 5-year old articles published in 1998.

One potential concern with the addition of this "explicit" control group is that control subfields could be affected by the treatment of interest. What if, for instance, a control source article happens to be related (in a PMRA sense) with the treated source? Because the subfields identified by PMRA are narrow, this turns out to be very infrequent. Nonetheless, we remove all such instances from the data. We then find all the intellectual neighbors for these control source articles using PMRA; a control subfield is defined by the set of related articles returned by PMRA, in a manner that is exactly symmetric to the procedure used to delineate treated subfields. When these related articles are parsed below to distinguish between those published by collaborators and non-collaborators of the star, or between those by intellectual outsiders and insiders, covariates for treated and control observations will always be defined with perfect symmetry.

D. Descriptive Statistics

The procedure described above yields a total of 34,218 distinct subfields; 3,076 subfields correspond to one of the 452 dead scientists, whereas 31,142 subfields correspond to one of 5,809 still-living scientists. Table 2 provides descriptive statistics for control and treated subfields in the baseline year, i.e., the year of death for the deceased scientist.[10]

Covariate balance. In the list of variables displayed in Table 2, a number of covariates are balanced between treated and control subfields solely by virtue of

[10]We can assign a counterfactual year of death for each control subfield, since each control subfield is associated with a particular treated subfield through the matching procedure described above.

the coarsened exact matching procedure—for instance, (star) investigator year of degree, the source article number of authors, or the source article number of citations at baseline. However, there is nothing mechanical to explain the balance between treated and control subsamples with respect to the stock of our main outcome variable: the number of articles in the star's field. Figure 1 compares the distributions of the cumulative number of articles published in our sample of subfields up to the year of death, broken down by treatment status. Overall, one can observe a great deal of overlap between the two histograms; the means and medians are virtually identical. Of course, balance in the levels of the outcome variable is not technically required for the validity of the empirical exercise.[11] Yet, given the *ad hoc* nature of the procedure used to identify control subfields, this degree of balance is reassuring.

Another happy byproduct of our matching procedure is that treated and control scientists also appear quite similar in the extent of their eminence at the time of (counterfactual) death, whether such eminence is measured through NIH funding, the number of articles published, or the number of citations these articles received.

Collaborators vs. non-collaborators. One critical aspect of the empirical analysis is to distinguish between collaborators and non-collaborators of the star when measuring publishing activity in a subfield. It is therefore crucial to describe how this distinction can be made in our data. Information about the superstars' colleagues stems from the Faculty Roster of the Association of American Medical Colleges (AAMC), to which we secured licensed access for the years 1975 through 2006, and which we augmented using NIH grantee information (cf. Azoulay, Graff Zivin, and Wang [2010] for more details).

An important implication of our reliance on these sources of data is that we can only identify authors who are faculty members in U.S. medical schools, or recipients of NIH funding. We cannot systematically identify scientists working for industrial firms, or scientists employed in foreign academic institutions.[12] The great benefit of using AAMC data, however, is that they ensure we have at our disposal both demographic and employment information for every individual in the relevant labor market: their (career) age, type of degree awarded, place of employment, gender, and research output, whether measured by publications or NIH grants.

To identify authors, we match the authorship roster of each related article in one of our subfields with the AAMC roster.[13] We tag as a collaborator any author who appeared as a co-author of the star associated with the subfield on any publication prior to the death. Each related article is therefore assigned to

[11] What is required is that the trends in publication activity be comparable between treated and control subfields up until the death of the treated scientist. We verify that this is the case below.

[12] We can identify trainees who later go on to secure a faculty position, but not those who do not stay in academia.

[13] We limit ourselves to authors with relatively infrequent names. Though this may create some measurement error, there is no reason to suspect that the wrongful attribution of articles to authors will impact treated and control subfields in a differential way.

one of two mutually-exclusive bins: the "collaborator" bin comprises the set of publications with at least one identified author who coauthored with the star prior to the year of death (or counterfactual death); the "non-collaborator" bin comprises the set of publications with no identified author who coauthored with the star prior to the year of death (or counterfactual death).[14] As can be seen in Table 2, roughly 11% of the publication activity at baseline can be accounted for by collaborators. Moreover, this proportion is very similar for control and treated subfields.[15]

A first look at subfield activity. Figure E1 in Appendix E confirms that the treated and control subfields are on similar trajectories in publication activity up to the time of superstar death (though they diverge after the death event). This provides suggestive evidence for the validity of our research design, and is notable since the coarsened exact matching procedure that generated the sample of control subfields did not make any use of these outcomes. Moreover, the absence of differential trends can be observed for overall activity, for activity restricted to collaborators of the star, and for the publishing activity of non-collaborators.

More boldly, we can use these averages in the raw data to examine changes in outcomes after the death. For both treated and control subfields, the curves exhibit a pronounced inverted U-shaped pattern, with activity first increasing until it reaches a peak roughly two years before the death of the star (or counterfactual death for the control subfields and their associated stars). Activity then decreases steadily, but the slope of the decrease appears more pronounced for control subfields, relative to treated subfields (Panel A). This pattern is flipped when examining activity due to collaborators (Panel B): the relative decline is much more pronounced for treated subfields, which is consistent with the results in Azoulay, Graff Zivin, and Wang (2010). Panel C, which focuses on subfield activity limited to non-collaborators, provides the first non-parametric evidence that the downward-sloping part of the activity curve is less steep for treated subfields.

Figure E1 provides a transparent illustration of subfield publication activity over time, which proceeds directly from averaging the raw data, but the evidence it provides should be handled with an abundance of caution. First, it conflates calendar time and experimental time, when in actuality the death events in the data occur at varying frequencies between the years 1975 and 2003. Second, covariates like field age are not perfectly balanced across the treated and control groups, since the number of control subfields is not identical across treated subfields. Finally, it abstracts away from robust inference, and particularly from clustering: one would expect the subfield outcomes associated with an identical star to be

[14] We identify the publications in the subfield for which the superstar is an author and eliminate them from these calculations. As a result, any decrease in activity within the subfield cannot be ascribed to the mechanical effect of its star passing away.

[15] We define collaboration status by looking at the authorship roster for the entire corpus of work published by the star before or in the year of death, and not only with respect to the articles of the star that belong to the focal subfield.

correlated. Our econometric framework, described below, addresses these limitations and as a result provides a more solid foundation for the estimation of the causal effect of star death on the dynamics of subfield activity.

III. Results

The exposition of the econometric results proceeds in stages. After a review of methodological issues, we provide results that pertain to the main effect of superstar death on subfield growth, measured by publication rates and funding flows. Next, we attempt to elucidate the mechanism (or set of mechanisms) at work to explain our most robust finding, that of relative subfield growth in the wake of a star's passing, a growth entirely accounted for by contributions from non-collaborators. We do so by examining the characteristics of the articles published by non-collaborators, before turning to the characteristics of their authors. We also explore heterogeneity in the treatment effect through the interaction of the post-death indicator variable with various attributes of the stars and the subfields.

A. *Econometric Considerations*

Our estimating equation relates publication or funding activity in subfield i in year t to the treatment effect of losing a superstar:

$$E\left[y_{it}|X_{it}\right] = exp\Big[\beta_0 + \beta_1 AFTER\_DEATH_{it} + \beta_2 AFTER\_DEATH_{it} \times TREAT_i + f(AGE_{it}) + \delta_t + \gamma_i\Big] \tag{1}$$

where y is a measure of subfield activity, $AFTER\_DEATH$ denotes an indicator variable that switches to one in the year after the superstar associated with i passes away, $TREAT$ is an indicator variable for treated subfields, $f(AGE_{it})$ corresponds to a flexible function of the field's age, the δ_t's stand for a full set of calendar year indicator variables, and the γ_i's correspond to subfield fixed effects, consistent with our approach to analyze *changes* in activity within subfield i following the passing of a superstar.[16]

The subfield fixed effects control for many time-invariant characteristics that could influence research activity, such as the need for capital equipment or the extent of disease burden (e.g., for clinical fields). A pregnant metaphor for the growth of scientific knowledge has been that of biological evolution (Hull 1988; Chavalarias and Cointet 2013): a field is born when new concepts are introduced, resulting in an accelerating production of "offspring" (articles), until the underlying scientific community loses its thematic coherence, ushering in an era of decline

[16]To avoid confusion, we have suppressed any subscript for the superstars. This is without loss of generality, since each subfield is uniquely associated with a single star.

(or alternatively, splitting or merging events). To flexibly account for such life cycle effects, we include subfield age indicator variables (where subfield age is computed as the number of years since the year of publication for the source article). The calendar year effects filter out the effects of the general expansion of the scientific enterprise as measured by the number of journals and articles published each year.[17]

We follow Jaravel et al. (2018) in including in our specification an indicator for the timing of death that is common to treated and control subfields (whose effect will be identified by the coefficient β_1) in addition to the effect of interest, an interaction between $AFTER\_DEATH$ and $TREAT$ (whose effect will be identified by the coefficient β_2). The effects of these two variables are separately identified because (i) death events are staggered across our observation period and (ii) control subfields inherit a counterfactual date of death because they are uniquely associated with a treated subfield through the matching procedure described in section II.C. The inclusion of the common term addresses the concern that age, calendar year, and subfield fixed effects may not fully account for shifts in subfield activity around the time of the star's passing. If this is the case, $AFTER\_DEATH$ will capture the corresponding transitory dynamics, while $AFTER\_DEATH \times TREAT$ will isolate the causal effect of interest. Empirically, we find that in some specifications, the common term has substantial explanatory power, though its inclusion does not radically alter the magnitude of the treatment effect.

Estimation. The dependent variables of interest, including publication counts and NIH grants awarded, are skewed and non-negative. For example, 31.40% of the subfield/year observations in the data correspond to years of no publication activity; the figure climbs to 56.70% if one focuses on the count of NIH grants awarded. Following a long-standing tradition in the study of scientific and technical change, we present quasi-maximum likelihood (hereafter QML) estimates based on the conditional fixed effects Poisson model developed by Hausman, Hall, and Griliches (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, Monfort, and Trognon 1984).

Inference. QML (i.e., "robust") standard errors are consistent even if the underlying data generating process is not Poisson. In fact the Hausman et al. estimator can be used for any non-negative dependent variables, whether integer or continuous (Santos Silva and Tenreyro 2006), as long as the variance/covariance matrix is computed using the outer product of the gradient vector (and therefore does

[17] It is not possible to separately identify calendar year effects from age effects in the "within subfield" dimension of a panel in a completely flexible fashion, because one cannot observe two subfields at the same point in time that have the same age but were born in different years (Hall, Mairesse, and Turner 2007).

not rely on the Poisson variance assumption). Further, QML standard errors are robust to arbitrary patterns of serial correlation (Wooldridge 1997), and hence immune to the issues highlighted by Bertrand, Duflo, and Mullainathan (2004) concerning inference in DD estimation. We cluster the standard errors around superstar scientists in the results presented below.[18]

Dependent Variables. Our primary outcome variable is publication activity in a subfield. However, we go beyond this raw measure by assigning the related articles that together constitute the subfield into a variety of bins. For instance, we can decompose publication activity in the subfield into two mutually exclusive subfields: articles with a superstar on the authorship roster vs. articles without a superstar; etc. Articles in each bin can then be counted and aggregated up to the subfield/year level.

Capturing funding flows at the field level is slightly more involved. *PubMed* systematically records NIH grant acknowledgements using grant numbers. Unfortunately, these grant numbers are often truncated and omit the grant cycle information that could enable us to pin down unambiguously the particular year in which the grant was awarded. When it is missing, we impute the award year using the following rule: for each related publication that acknowledges NIH funding, we identify the latest year in the three-year window that precedes the publication during which funding was awarded through either a new award or a competitive renewal. To measure funding activity in a subfield, we create a count variable that sums all the awards received in particular year, where these awards ultimately generate publications in the focal subfield.

B. Main effect of superstar death

Table 3 and Figure 2 present our core results. Overall, we find that publication activity increases slightly following the death of a star scientist who was an active contributor to it, but the magnitude of the effect is modest (about 5.2%) and imprecisely estimated (column 1). Yet, this result conceals a striking pattern that is uncovered when we distinguish between publications by collaborators and non-collaborators. The decline in publication activity accounted for by previous collaborators of the star is large, on the order of 20.7% (column 2). This evidence is consistent with previous findings, which showed that coauthors of superstar scientists who die suffer a drop in output, particularly if their non-collaborative work exhibited strong keyword overlap with the star, i.e., if they were intellectually connected in addition to being coauthors (Azoulay, Graff Zivin, and Wang 2010, Table VI, column 2).

[18]Knowledge spillovers and scientific breakthroughs, including the adoption of research tools, could encourage innovation across related fields. This possibility is not entirely dealt with by clustering inference at the star level, since spatial dependence in knowledge space could occur between any pair of subfields, whereas clustering only allows for dependence among the subfields associated with the same star. As it turns out, the Poisson conditional fixed effects estimator also provides a consistent estimator of the variance in the presence of time-invariant patterns of spatial auto-correlation (Bertanha and Moser 2016).

A limitation of the previous work focusing on the fate of collaborators after the loss of an eminent scientist always lied in the failure to distinguish between social and intellectual channels of influence, since every treated scientist was by definition a collaborator, even if merely a casual one. In this study, we can relax this constraint, and when we do, we find that relative publication activity by non-collaborators in the subfield increases by a statistically significant $100 \times (e^{0.082} - 1) = 8.6\%$ (column 3).[19]

We also explore the dynamics of the effects uncovered in Table 3. We do so 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 superstar's death, and then graphing the effects and the 95% confidence interval around them (Panels A, B, and C of Figure 2 correspond to columns 1, 2, and 3 in Table 3).[20]

Two features of the figure are worthy of note. First, the dynamics amplify the previous results in the sense that we see the effects increasing (in absolute value) monotonically over time—there is no indication that the effects we estimated in Table 3 are merely transitory. Five years after a star's death, the relative increase in publication activity by non-collaborators is large enough in magnitude to fully offset the decline in activity by collaborators. Second, there is no discernible evidence of an effect in the years leading up to the death, a finding that validates *ex post* our identification strategy.

Nevertheless, the case for the exogeneity of death events with respect to the course of knowledge growth and decline within a subfield is stronger for sudden causes of deaths than for anticipated causes of death. Figure E2 in Appendix E provides a version of Figure 2, Panel C (event study graphs for non-collaborators) broken down by causes of death (anticipated vs. sudden). While there is more variability in the estimated path of outcomes in the years leading up to the death event in the anticipated case (Panel A) than in the sudden case (Panel B), it is imprecisely estimated and non-monotonic. In both panels, however, one can observe a slow but steady increase after the event in the rate of contributions by non-collaborators in treated subfields, relative to control subfields. The distinction between sudden and anticipated events is explored further in section III.D.

The last three columns of Table 3 focus on funding flows from the National Institutes of Health (NIH) rather than publication flows. More precisely, the outcome variable in columns 4, 5, and 6 is the number of distinct NIH awards that acknowledge a publication in the subfield in the three-year window before the year of publication for the related article (summing the financial total of grant amounts, as opposed to the number of grants, yields similar results). The patterns are very similar to those obtained in the case of publication activity,

[19] The number of observations varies ever so slightly across columns because the conditional fixed effects specification drops observations corresponding to subfields for which there is no variation in activity over the entire observation period. This is true as well for the results reported in Tables 4 through 7.

[20] In these specifications, the $AFTER\_DEATH$ term which is common to treated and control subfields is also interacted with a complete series of lags and leads relative to the year of death or counterfactual death.

both in terms of magnitudes and in terms of statistical significance.

C. Subfield growth patterns

In the remainder of the manuscript, we seek to characterize the kind of contribution, and the type of investigators that give rise to the novel empirical regularity we uncovered: that of relative growth for subfields following the death of their superstar anchor, a phenomenon entirely accounted for by research activity undertaken by scientists who never collaborated with the star while alive. As a consequence, all the results below pertain to contributions by non-collaborators; any article with even one author who collaborated with the star is excluded from the count of articles that constitute the dependent variable.

The impact and direction of new research. What characterizes the additional contributions that together lead to increased activity in a subfield after a star has passed on? Are these in fact important contributions to the subfield? Do they continue to focus on mainstream topics within the subfield, or should they be understood as taking the intellectual domain in a novel direction? Tables 4 and 5 explore these issues.

In Table 4, we parse every related article in the subfields to assign them into one of six mutually exclusive bins, based on their vintage-specific long-run citation impact: articles that fall in the bottom quartile of the citation distribution; in the second quartile; in the third quartile; articles that fall above the 75^{th} percentile, but below the 95^{th} percentile; articles that fall above the 95^{th} percentile, but below the 99^{th} percentile; articles that fall above the 99^{th} percentile of the citation distribution.[21] Each column in Table 4 (with the exception of the first which simply replicates the effect for all papers, regardless of impact, that was previously displayed in Table 3, column 3) reports the corresponding estimates. A startling result is that the magnitude of the treatment effect increases sharply and monotonically as we focus on the rate of contributions with higher impact. In contrast, the number of lower-impact articles contributed by non-collaborators contracts slightly, though the effect is not precisely estimated.[22]

[21] A vintage is comprised of all the articles published in a given year. When we are referring to the vintage-specific, article-level distribution of citations, the relevant universe to compute quantiles is not limited to the articles that constitute the subfields in our data. Rather, the relevant universe includes the entire set of 17,312,059 articles that can be cross-linked between *PubMed* and the Web of Science. As a result, there is no reason to suspect that individual stars, or even our entire set of stars, could ever alter the shape of these distributions. For example, the article by Sopko et al. highlighted on Figure C5 (in Appendix C) received 40 citations from other articles in *PubMed* by 2015. This puts this article above the 79^{th} percentile of the citation distribution for articles published in 2002.

[22] Table E3 and Figure E3 in Appendix E break down these results further by examining separately the growth of subfields by cause of death (anticipated vs. sudden). As mentioned earlier, the case for exogeneity is stronger for sudden death, since when the death is anticipated, it would be theoretically possible for the star to engage in "intellectual estate planning," whereby particular scientists (presumably close collaborators) are anointed as representing the next generation of leaders in the subfield. Our core results continue to hold when analyzed separately by cause of death. However, we gain statistical power from pooling these observations, and some empirical patterns would be estimated less precisely if we chose to focus solely on observations corresponding to subfields for which the star died suddenly and

Table 5 parses the related articles in each subfield to ascertain whether contributions by non-collaborators constitute a genuine change in intellectual direction. Panel A distinguishes between contributions that are proximate in intellectual space to the source article from those that are more distant (though still part of the subfield as construed by PMRA). Because we have at our disposal both a cardinal and an ordinal measure of intellectual proximity, we present two sets of estimates. In both cases, the magnitude of the treatment effect pertaining to PMRA-proximate publication activity is larger, and more precisely estimated than the magnitude corresponding to PMRA-distant publication activity (relative to the same patterns for the control group of subfields). We can certainly rule out the conjecture that non-collaborators enter the field from the periphery. Rather, their contributions appear to tackle mainstream topics within the subfield.

Panel B sheds light on the intellectual direction of the field, by examining the cited references contained in each related article. The first two columns separate related articles in two groups: publications that cite at least some work which belongs to the subfield identified by PMRA for the corresponding source and publications that cite exclusively out of the PMRA subfield. Only articles in the second group appear to experience growth in the post-death era. The next two columns proceed similarly, except that the list of references is now parsed to highlight the presence of articles authored by the star (Column 3), as opposed to all other authors (Column 4). We find that subfield growth can be mostly accounted for by articles from non-collaborators who do not build on the work of the star.

Whereas Panel B highlighted the extent to which contributors were bringing new sources of inspiration into the subfield, Panel C focuses on the extent to which the treated subfields move closer to the scientific frontier in the wake of the superstar's passing. The first two columns do so by distinguishing between contributions that draw on recent versus more dated references. This exercise is repeated in Columns 3 and 4, with a focus on the vintage of the MeSH term combinations for each article in the subfield.[23] Both sets of results indicate that these new contributions are more likely to build on science of a more recent vintage.

Taken together, the results presented in Table 5 paint a nuanced picture of directional change in the wake of superstar passing. The new contributions do not represent a radical departure from the subfield's traditional concerns (Panel A). At the same time, the citation and MeSH evidence (Panels B and C) make it clear that these additional contributions are more likely to draw on new-to-the-subfield as well as new-to-the-world ideas. In short, they both rejuvenate the subfield, and alter its angular velocity by shifting its intellectual center of gravity away from its pre-death position.

unexpectedly.

[23] A two-way MeSH term combination is born in the year where an article is annotated by the keyword pair for the first time.

It is important to note, however, that the findings above do not imply that the published results of entrants necessarily contradict or overturn the prevailing scientific understanding and assumptions within a subfield. We provide indirect evidence regarding these contributions' disruptive impact by leveraging a measure recently proposed by Funk and Owen-Smith (2017). Their index captures the degree to which an idea consolidates or destabilizes the status quo, by measuring whether the future ideas that build on the focal idea also rely on its acknowledged predecessors. The results in Table E4 of Appendix E suggest that these contributions do not radically disrupt the subfield. Rather, they appear to reflect the impact of a myriad "small r," permanent revolutions whereby new ideas come to the fore without necessarily eclipsing prior approaches.

Outsiders vs. competitors. The next step of the analysis is to investigate the type of scientists who publish the articles that account for subfield growth in the wake of a star's death. We examine the proximity in intellectual space between non-collaborators in the subfield and the deceased superstar. One possibility is that non-collaborators are competitors of the star, with much of their publication activity falling into the subfield when the star was alive. Another possibility is that they are recent entrants into the subfield—intellectual outsiders. To distinguish these different types of authors empirically, we create a metric of intellectual proximity for each related author we can match to the AAMC Faculty Roster, by computing the fraction of their publications that belongs to the star's subfields up to the publication year for each related article.[24] The distribution of this field overlap measure is displayed on Panel A of Figure 3. The distribution is skewed, with a pronounced mass point at the origin: approximately 50% of the related articles turn out to have authors with exactly zero intellectual overlap with the star's subfield, and another 1.24% are authored by new scientists for whom this publication within the subfield is also their first publication overall.

We now use this metric to gauge the extent to which the post-death publication activity by non-collaborators (relative to the control group) can be attributed to related authors whose outsider status falls into one of twelve separate bins. This includes one bin for new scientists, one bin for the bottom half of the overlap distribution, one bin for every five percentiles above the median (50^{th} to 55^{th} percentile, 55^{th} to 60^{th} percentile,..., 95^{th} to 99^{th} percentile), as well as a top percentile bin. We then compute the corresponding measures of subfield activity by aggregating the data up to the subfield/year level. These results are presented graphically in Panel B of Figure 3. Each dot corresponds to the magnitude of the treatment effect in a separate regression with the outcome variable being the number of articles in each subfield that belong to the corresponding bin.

A striking pattern emerges. The authors driving the growth in relative publi-

[24]Whenever we match more than one author on a related article, we assign to that article the highest proximity score for any of the matched authors. Appendix E, Table E9 defines overlap with respect to all the subfields associated with a given star, rather than simply the focal subfield. This does not alter our conclusions.

cation activity following a star's death are largely outsiders. They do not appear to have been substantially active in the subfield when the star was alive. In other words, they are predominantly new entrants into these subfields, though not necessarily novice scientists.

D. The Nature of Entry Barriers

The evidence so far points to fields of deceased stars enjoying bursts of activity after the death event. The influx of outsiders documented above suggests that stars may be able to regulate entry into their field while alive. In this section, we attempt to uncover the precise nature of barriers to entry into the subfields where the stars were prominent prior to their untimely demise. Methodologically, we do so by splitting the sample of fields across the median for a series of relevant covariates. Because there is no presumption that death events are exogenous with respect to subfield growth and decline within the strata delineated by these covariates, it should be clear that we will only be able to document conditional correlations, and not causal effects in what follows.[25]

While it is tempting to envisage conscious effort by the stars to block entry through the explicit control of key resources, such as funding and/or editorial goodwill (Brogaard, Engelberg, and Parsons 2014; Li 2017), this explanation appears inconsistent with the facts on the ground. In the five-year window before death, only three of our stars (out of 452) were sitting on study sections, the funding panels that evaluate the scientific merits of NIH grant applications. Another three were journal editors in the same time window. This handful of individuals could not possibly drive the robust effects we have uncovered.[26] If barriers to entry are not the result of explicit control by stars, what is discouraging entry?

Goliath's shadow. One possibility is that outsiders are simply deterred by the prospect of challenging a luminary in the field. The existence of a towering figure may skew the cost-benefit calculations from entry by outside scholars toward delay or alternative activities. Table 6 examines this role of implicit barriers to entry by focusing on the eminence of the star. Eminence is measured through the star's publication count, the star's cumulative number of citations garnered up to the year of death, and the star's cumulative amount of NIH funding. We also have a "local" measure of eminence: the star's importance to the field, which is defined as the fraction of papers in the subfield that have the star as an author. Splitting the sample at the median of these measures reveals a consistent pattern of results. Stars that were especially accomplished appear to be an important deterrent to entry, with their passing creating a larger void for non-collaborators

[25]Instead of interacting the treatment effect with covariates, we prefer to estimate our benchmark specifications on subsamples corresponding to below and above the median of these covariates. For these two approaches to yield comparable results, one would need to also saturate the specification with interaction terms between the covariates and year/field age effects. In practice, we have found that the fixed effects Poisson models fail to converge with this full set of interactions.

[26]We verified that omitting these scientists from the sample hardly change the core results.

to fill. Rather than directly thwarting the efforts of potential entrants, it appears that the mere presence of a preeminent scholar is sufficient to dissuade intellectual outsiders from engaging with the field.

Of course, the accomplishment of the star alone may not be the only factor influencing entry. We next turn our attention to how the characteristics of the field and the star's coauthors may also modulate this relationship. Since entry is largely confined to those fields that have lost an eminent star, the analysis that follows limits attention to those subfields in which the most eminent among the stars were active, as measured by our citation metric in Table 6.[27]

Subfield coherence. Entry into a field, even after it has lost its star, may be deterred if the subfield appears unusually coherent to outsiders. A subfield is likely to be perceived as *intellectually* coherent, when the researchers active in it agree on the set of questions, approaches, and methodologies that propel the field forward. Alternatively, a field might be perceived as *socially* coherent, when the researchers active in it form a tightly-knit clique, often collaborating with each other, and perhaps also reviewing each other's manuscripts. To explore these purported barriers to subfield entry, we develop two alternative measures of intellectual coherence, and one measure of social coherence.

Our first index of intellectual coherence leverages PMRA to capture the extent to which articles in the subfield pack themselves into a crowded scientific neighborhood. Recall that for each article in a subfield, we have at our disposal both a cardinal and an ordinal measure of intellectual proximity with the source article from which all other articles in the subfield radiate. Focusing only on the set of articles published in the subfield before the year of death, we measure intellectual coherence as the cardinal ranking (expressed as a real number between zero and one) for the 25^{th} most related article in the subfield.[28] According to this metric, subfields exhibit wide variation in their degree of intellectual coherence, with a mean and median equal to 0.60 ($sd = 0.13$). The second index of intellectual coherence exploits the list of references cited in each article in the subfield before the star's death. In the spirit of Funk and Owen-Smith (2017), for all related articles published in the five years prior to the star's death, we compute the fraction of references that fall within the subfield. Our contention is that subfields that are more self-referential will tend to dissuade outsiders from entering. Once again, we observe meaningful variation across subfields using this second index ($mean = 0.05; sd = 0.04$).

Our measure of social coherence summarizes the degree of "cliquishness" within

[27]More precisely, Table 7 below drops from the sample subfields associated with stars who fall below the median of cumulative citations garnered by the year of death. Results are qualitatively similar when focusing on the most eminent stars as defined by publications or NIH funding. Table F6 in Appendix F presents the results corresponding to the subsample of less-eminent stars.

[28]The choice of the twenty fifth-ranked article is arbitrary, and also convenient. After purging from each subfield reviews, editorials, and articles appearing in journals not indexed by *WoS*, 95% of the subfields contain 25 articles or more in the period that precedes the star's death. In those rare cases where the number of articles is less than twenty-five, we choose as our measure of coherence the cardinal measure for the least-proximate article in the subfield.

a subfield by computing the clustering coefficient in its coauthorship network. The clustering coefficient is simply the proportion of closed triplets within the network, an intuitive way to measure the propensity of scientists in the field to choose insiders as collaborators.[29]

Panel A of Table 7 investigates the role of these intellectual and social barriers in modulating the post-death expansion of fields. We find tentative evidence of a role for both types of barriers, in that the magnitude of the treatment effect for coherent fields is always smaller than the magnitude for less coherent fields, regardless of how coherence is measured. The *difference* between the estimates for more or less coherent subfields does not reach statistical significance at conventional levels. What seems notable, however, is that the magnitudes are consistently ordered across the three measures.

Incumbent resource control. While we noted earlier that stars do not appear especially well positioned to directly block entry through the control of key resources, it is possible that those resources can be controlled indirectly through the influence of collaborators. If incumbent scholars within a field serve as gatekeepers of funding and journal access, they may be able to effectively stave off threats of entry from outsiders. The same may be implicitly true if collaborators are the recipients of the lion's share of funding within the field. To assess financial gatekeeping, we use information regarding the composition of NIH funding panels, to tabulate, for each star, the number of collaborators who were members of at least one of these committees in the five years preceding the death of the star. We would like to proceed in a similar fashion using the composition of editorial boards, but these data are not easily available for the set of *PubMed*-indexed journals and the thirty-year time period covered by our sample. As an alternative, we develop a proxy for editorial position based on the number of editorials or comments written by every collaborator of the star.[30] We then sum the number of editorials written by coauthors in the five years before the death. Together, the editorial and study section information allow us to distinguish between the stars whose coauthors were in a position to channel resources towards preferred individuals or intellectual approaches from those stars whose important coauthors had no such power.

Panel B of Table 7 presents the evidence on the role of indirect control. The results paint a consistent, if not always statistically significant, picture. While subfield expansion is the rule, it appears more pronounced when stars have rela-

[29]The clustering coefficient is based on triplets of nodes (authors). A triplet consists of three authors that are connected by either two (open triplet) or three (closed triplet) undirected ties. The clustering coefficient is the number of closed triplets over the total number of triplets (both open and closed, cf. Luce and Perry [1949]).

[30]We investigated the validity of this proxy as follows. In the sample of deceased superstars, every individual with five editorials or more was an editor. In a random sample of 50 superstars with no editorials published, only one was an editor (for a field journal). Finally, among the sixteen superstars who wrote between one and four editorials over their career, we found two whose CV indicate they were in fact editors for a key journal in their field. We conclude that there appears to be a meaningful correlation between the number of editorials written and the propensity to be an editor.

tively few collaborators in influential positions, or collectively capture a smaller portion of the funding that supported research in the subfield. Indirect control therefore appears to be a potential mechanism through which superstars can exert influence on the evolution of their fields, even from beyond the grave. Coauthors, either through their direct effort to keep the star's intellectual flame alive or simply by their sheer (financial) dominance in the field, appear to erect barriers to entry into those fields that prevent its rejuvenation by outsiders.

Taken together, these results suggest that outsiders are reluctant to challenge hegemonic leadership within a field when the star is alive. They also highlight a number of factors that may constrain entry even after she is gone. Intellectual, social, and resource barriers all seem to play a role in impeding entry, with outsiders only entering subfields whose topology offers a less hostile landscape for the support and acceptance of "foreign" ideas.

E. Reallocation and Welfare

What are the implications of our results for welfare? We approach this question with a great deal of caution, since much of the evidence presented thus far pertains to changes in the *direction*, rather than the *rate*, of scientific progress. Making welfare statements in this context is tantamount to valuing the importance of the new directions in which related authors take their fields (compared to the prior agenda inherited from the superstar), as well as ascertaining the fate of fields that the new entrants departed, and the agenda they otherwise might have pursued had the star remained alive. Such an exercise is fraught with peril. Below we synthesize the results that already speak to these questions, and provide a few additional suggestive pieces of evidence.

Our earlier evidence suggests that entrants bring different and more recent ideas into the subfields they enter to create highly impactful output (Tables 4 and 5). In Appendix E we further show that the subfields that experience the largest post-death boost in activity are those in which the star was presiding over an empire that was losing momentum in the years immediately preceding the star's death (Tables E5 and E8). These subfields are also those in which the star's close collaborators were less able to regulate entry (Table 7B).

It is important to note, however, that the additional output by entrants in treated subfields is largely offset by commensurate declines in output by the star's collaborators (Table 3). Moreover, these new contributions appear to come at the expense of the entrants' prior agenda. In Appendix G, we examine changes in total output at the related author level, using a difference-in-differences set-up that parallels our analyses at the subfield level. The results in Table G1 show that non-collaborators do not increase their overall output, measured in terms of publications and NIH grants awarded. Since we know from our main analysis that related authors are contributing more within the subfields of dead superstars, the absence of changes in total output imply that this additional work is displacing work they were doing in other subfields. Their new output replaces, at least in

part, articles that these authors would have written in other intellectual domains had the star remained alive.[31]

As a whole, these results imply that entrants are moving subfields in productive directions relative to the period immediately preceding the passing of the star, but without increasing scientific output in the aggregate. However, the impacts in the final years of a star's life are not necessarily indicative of their contributions writ large. Indeed, the lofty accomplishments which earned them superstar status suggest that their net contribution to society is likely positive. A longer view would also recognize that the scientific journeymen of today may well become the stars of tomorrow (as shown in Table E10 of Appendix E) with a career that slowly builds to an apex of socially valuable accomplishments, that will someday experience a similar decline (see Figure E4 in Appendix E).

F. Extensions and robustness

Appendix E presents results pertaining to extensions of the main analyses. Appendix F provides a number of robustness checks. In the interest of space, we only call out a subset of the analyses presented therein, but we have written these appendices as stand-alone documents, such that the interested reader can consult them for additional details.

Impact of research infrastructure needs. Our analysis is limited to the life sciences. Though this area accounts for a large fraction of publicly funded, civilian research funding in the United States, it is not necessarily representative of all fields of science. In particular, some domains of research require access to expensive and specialized capital equipment. When capital needs are large and lumpy, the evolution of subfields in the wake of an eminent scientist's death will likely depend on the institutions that govern access to the scarce capital equipment.

Within biomedical research, large-scale clinical trials most closely—albeit imperfectly—resemble the characteristics of capital-intensive scientific fields. These require a large infrastructure of data collection, monitoring, and management, which is why these activities are often consolidated in large cooperative groups such as the AIDS Clinical Trials Group, the Children's Oncology Group, or the Framingham Heart Study. *PubMed* has a "publication type" field which allows us to identify the subfields that are clinical-trial intensive (10% of the subfields) versus those that are not (the remaining 90%). Table E6 replicates the results of Table 3 separately for these two subsamples. Although our ability to estimate statistically significant effects is limited by sample size, the magnitudes are very similar.

Impact of star age and experience. As explained earlier, we do not impose

[31] We also estimate a dynamic version of these specifications and display the corresponding event study-style graphs in Figure G1 (publication output) and Figure G2 (grant output). In general, it appears from these figures that the total output of related authors neither expands nor contracts in the wake of a star's passing.

a strict age cutoff for the deceased star, we merely insist that they exhibit tangible signs of research activity, such as publishing original articles, obtaining NIH grants, and training students. Among our 452 departed superstars, the median age at death is 61, the seventy-fifth percentile 67, and the top decile 73. How do the core results change when the scientists who passed away at an advanced age are excluded from the sample? As can be observed in Table E7, the subfields of stars who passed away more prematurely are responsible for most of the effect. The effect for the fields associated with older stars is small in magnitude and imprecisely estimated. We chose to keep these older stars in the sample because a larger sample affords us opportunities to explore mechanisms without losing power to detect nuanced effects statistically.

Star level analyses. In Table F1, we probe the robustness of the core results presented in Table 3 after rolling up the data to the level of the star scientist (deceased or control). Recall that the treatment variable exhibits variation at the level of the star scientist, and not at the level of a single subfield. In this robustness check, we lump all related articles for each star together as if they belonged to a single subfield. The results in Table F1 are quite similar to those in Table 3, both in terms of magnitude and statistical significance. One exception is the coefficient on the effect of entry by collaborators, which is negative as expected, but smaller in magnitude, relative to the corresponding coefficient in Table 3. The corresponding event-study graphs, displayed in Figure F3, also display patterns fully consistent with those observed for our benchmark set of results. As explained in Section II.B, we strongly prefer performing the analyses at the the subfield level, for two reasons. First, the subfields delineated by PMRA exhibit limited overlap (see Figure C6 in Appendix C), and as a result the within-star, between subfield variation in publication activity can be exploited meaningfully. Second, we can track the differential position of the star across the subfields in which she was active. The covariates that leverage these differences help us shed light on mechanisms, as in Tables 7, E5, and E8.

Alternate functional forms. In Table F2, we examine the sensitivity of our benchmark set of results to the choice of alternative functional forms. In the three columns to the left, we simply use the "raw" number of articles in the subfield as the outcome, and perform estimation by OLS. Of course, the estimates are not directly interpretable in terms of elasticities. At the mean of the data, however, the treatment effect in the third column implies that subfield entry by non-collaborating authors expands by $0.409/3.335 = 12.26\%$, which is not all that different from the 8.2% reported in Table 3. In the three columns to the right, we report results corresponding to OLS estimation, but this time with the outcome variables transformed using the inverse hyperbolic sine function (Burbidge, Magee, and Robb 1988). In this case, coefficient estimates can be interpreted as elasticities, as an approximation. They are quite similar once again to those reported in Table 3, except for the effect on entry by collaborators, which is smaller

in magnitude.

IV. Conclusion

In this paper, we leverage the applied economist's toolkit, together with a novel approach to delineate the boundaries of scientific fields, to explore the effect that the passing of an eminent life scientist exerts on the dynamics of growth—or decline—for the fields in which she was active while alive. We find that publications and grants by scientists that never collaborated with the star surge within the subfield, absent the star. Interestingly, this surge is not driven by a reshuffling of leadership within the field, but rather by new entrants that are drawn from outside of it. Our rich data on individual researchers and the nature of their scholarship allows us provide a deeper understanding of this dynamic.

In particular, this increase in contributions by outsiders appears to tackle the mainstream questions within the field but by leveraging newer ideas that arise in other domains. This intellectual arbitrage is quite successful—the new articles represent substantial contributions, at least as measured by long-run citation impact. Together, these results paint a picture of scientific fields as scholarly guilds to which elite scientists can regulate access, providing them with outsized opportunities to shape the direction of scientific advance in that space.

We also provide evidence regarding the mechanisms that may enable the regulation of entry. While stars are alive, entry appears to be effectively deterred where the shadow they cast over the fields in which they were active looms particularly large. After their passing, we find evidence for influence from beyond the grave, exercised through a tightly-knit "invisible college" of collaborators (de Solla Price and Beaver 1966; Crane 1972). The loss of an elite scientist central to the field appears to signal to those on the outside that the cost/benefit calculations on the avant-garde ideas they might bring to the table has changed, thus encouraging them to engage. But this appears to occur only when the topology of the field offers a less hostile landscape for the support and acceptance of "foreign" ideas, for instance when the star's network of close collaborators is insufficiently robust to stave off threats from intellectual outsiders.

In the end, our results lend credence to Planck's infamous quip that provides the title for this manuscript. Yet its implications for social welfare are ambiguous. While we can document that eminent scientists restrict the entry of new ideas and scholars into a field, gatekeeping activities could have beneficial properties when the field is in its inception; it might allow cumulative progress through shared assumptions and methodologies, and the ability to control the intellectual evolution of a scientific domain might, in itself, be a prize that spurs much *ex ante* risk taking. Because our empirical exercise cannot shed light on these countervailing tendencies, we must refrain from drawing concrete policy conclusions from our results.

All of the evidence we have presented pertains to the academic life sciences. It is unclear how the lessons from that setting might apply to other fields inside the

academy. In particular, when frontier research requires access to expensive and highly-specialized capital equipment—as is sometimes the case in the physical sciences—the rules governing access to that capital are likely to favor succession by insiders. At the other end of the spectrum, more atomistic fields where scientists generally work alone or in very small groups may evolve in a more frictionless manner. Whether our findings apply to industrial research and development is also an open question. In that setting, the choice of problem-solving approaches is guided by market signals (however imperfectly, cf. Acemoglu [2012]), and thus likely to differ from those selected under the more nuanced system of pecuniary and non-pecuniary incentives that characterizes academic research (Feynman 1999; Aghion, Dewatripont, and Stein 2008). Assessing the degree to which our results extend to other settings, and the reasons they might differ, represents a fruitful area for future research.

References

Aad, Georges et al. 2015. "Combined Measurement of the Higgs Boson Mass in Pp Collisions at S=7 and 8 Tev with the Atlas and Cms Experiments." *Physical Review Letters*, 114(191803), 1-33.

Acemoglu, Daron. 2012. "Diversity and Technological Progress," J. Lerner and S. Stern, *The Rate & Direction of Inventive Activity Revisited.* Chicago, IL: University of Chicago Press, 319-56.

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), 617-35.

Aghion, Philippe and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica*, 60(2), 323-51.

Akerlof, George and Pascal Michaillat. 2017. "Beetles: Biased Promotion and Persistence of False Belief," *NBER Working Paper #23523.*

Azoulay, Pierre; Jeffrey L. Furman; Joshua L. Krieger and Fiona Murray. 2015. "Retractions." *Review of Economics and Statistics*, 97(5), 1118-36.

Azoulay, Pierre; Joshua Graff Zivin and Jialan Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics*, 125(2), 549-89.

Azoulay, Pierre; Danielle Li; Joshua S. Graff Zivin and Bhaven N. Sampat. 2019. "Public R&D Investment and Private Sector Patenting: Evidence from NIH Funding Rules." *Review of Economic Studies*, 86(1), 117-52.

Bertanha, Marinho and Petra Moser. 2016. "Spatial Errors in Count Data Regressions." *Journal of Econometric Methods*, 5(1), 49-69.

Bertrand, Marianne; Esther Duflo and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1), 249-75.

Borjas, George J. and Kirk B. Doran. 2015b. "Cognitive Mobility: Labor Market Responses to Supply Shocks in the Space of Ideas." *Journal of Labor Economics*, 33(S1), S109-S45.

———————————————————. 2015a. "Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors." *Review of Economics and Statistics*, 97(5), 1104-17.

Börner, Katy; Chen, Chaomei and Kevin W. Boyack. 2003. "Visualizing Knowledge Domains." *Annual Review of Information Science and Technology*, 37(1), 179-255.

Bourdieu, Pierre. 1975. "La Spécificité du Champ Scientifique et les Conditions Sociales du Progrès de laa Raison." *Sociologie et Sociétés*, 7(1), 91-118.

Bramoullé, Yann and Gilles Saint-Paul. 2010. "Research Cycles." *Journal of Economic Theory*, 145(5), 1890-920.

Brogaard, Jonathan; Joseph Engelberg and Christopher Parsons. 2014. "Network Position and Productivity: Evidence from Journal Editor Rotations." *Journal of Financial Economics*, 111(1), 251-70.

Burbidge, John B.; Lonnie Magee and A. Leslie Robb. 1988. "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association*, 83(401), 123-27.

Chavalarias, David and Jean-Philippe Cointet. 2013. "Phylomemetic Patterns in Science Evolution—the Rise and Fall of Scientific Fields." *PLoS one*, 8(2), e54847.

Cole, Jonathan R. and Stephen Cole. 1972. "The Ortega Hypothesis." *Science*, 178(4059), 368-75.

Crane, Diana. 1969. "Social Structure in a Group of Scientists: A Test of the 'Invisible College' Hypothesis." *American Sociological Review*, 34(3), 335-52.

Crick, Francis. 1970. "Central Dogma of Molecular Biology." *Nature*, 227(5258), 561.

de Solla Price, Derek J. 1963. *Little Science, Big Science*. New York: Columbia University Press.

de Solla Price, Derek J. and Donald D. Beaver. 1966. "Collaboration in an Invisible College." *American Psychologist*, 21(11), 1011-8.

Feynman, Richard P. 1999. *The Pleasure of Finding Things Out*. New York: Basic Books.

Fortunato, Santo and Darko Hric. 2016. "Community Detection in Networks: A User Guide." *Physics Reports*, 659, 1-44.

Foster, Jacob G.; Andrey Rzhetsky and James A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review*, 80(5), 875-908.

Funk, Russell J. and Jason Owen-Smith. 2017. "A Dynamic Network Measure of Technological Change." *Management Science*, 63(3), 791-817.

Gorham, Geoffrey. 1991. "Planck's Principle and Jeans's Conversion." *Studies in History & Philosophy of Science*, 22(3), 471-97.

Gouriéroux, Christian; Alain Montfort and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica*, 53(3), 701-20.

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), 159-77.

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), 909-38.

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

Hull, David L.; Peter D. Tessner and Arthur M. Diamond. 1978. "Planck's Principle." *Science*, 202(4369), 717-23.

Jaravel, Xavier; Neviana Petkova and Alex Bell. 2018. "Team-Specific Capital and Innovation." *American Economic Review*, 108(4-5), 1034-73.

Jones, Benjamin F. 2009. "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies*, 76(1), 283-317.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago 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.

Levin, Sharon G.; Paula E. Stephan and Mary Beth Walker. 1995. "Planck's Principle Revisited: A Note." *Social Studies of Science*, 25(2), 275-83.

Li, Danielle. 2017. "Expertise vs. Bias in Evaluation: Evidence from the NIH." *American Economic Journal: Applied Economics*, 9(2), 60-92.

Li, Danielle; Pierre Azoulay and Bhaven N. Sampat. 2017. "The Applied Value of Public Investments in Biomedical Research." *Science*, 356(6333), 78-81.

Lin, Jimmy and W. John Wilbur. 2007. "Pubmed Related Articles: A Probabilistic Topic-Based Model for Content Similarity." *BMC Bioinformatics*, 8(423), 1-14.

Lotka, Alfred J. 1926. "The Frequency Distribution of Scientific Productivity." *Journal of the Washington Academy of Sciences*, 16(12), 317-23.

Luce, R. Duncan and Albert D. Perry. 1949. "A Method of Matrix Analysis of Group Structure." *Psychometrika*, 14(2), 95-116.

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

Mohnen, Myra. 2018. "Stars and Brokers: Knowledge Spillovers among Medical Scientists," *Working Paper.* Working Paper, University of Essex.

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

_____. 1992. *The Lever of Riches: Technological Creativity and Economic Progress.* New York: Oxford University Press.

Morange, Michel. 1998. *A History of Molecular Biology.* Cambridge, MA: Harvard University Press.

Myers, Kyle. 2018. "The Elasticity of Science." Working Paper, Harvard University. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3176991.

Nagaoka, Sadao and Hideo Owan. 2014. "Author Ordering in Scientific Research: Evidence from Scientists Surveyed in the US and Japan," *IIR Working Paper #13-23.* Hitotsubashi University, Institute of Innovation Research,

Oettl, Alexander. 2012. "Reconceptualizing Stars: Scientist Helpfulness and Peer Performance." *Management Science*, 58(6), 1122-40.

Reese, Thomas S. 2004. "My Collaboration with John Heuser." *European Journal of Cell Biology*, 83(6), 243-44.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5), S71-S102.

Rosvall, Martin and Carl T. Bergstrom. 2008. "Maps of Random Walks on Complex Networks Reveal Community Structure." *Proceedings of the National Academy of Sciences*, 105(4), 1118-23.

Sampat, Bhaven N. and Frank R. Lichtenberg. 2011. "What Are the Respective Roles of the Public and Private Sectors in Pharmaceutical Innovation?" *Health Affairs*, 30(2), 332-39.

Santos Silva, João M.C. and Silvanna Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4), 641-58.

Shapin, Steven. 1996. *The Scientific Revolution.* Chicago, IL: University of Chicago Press.

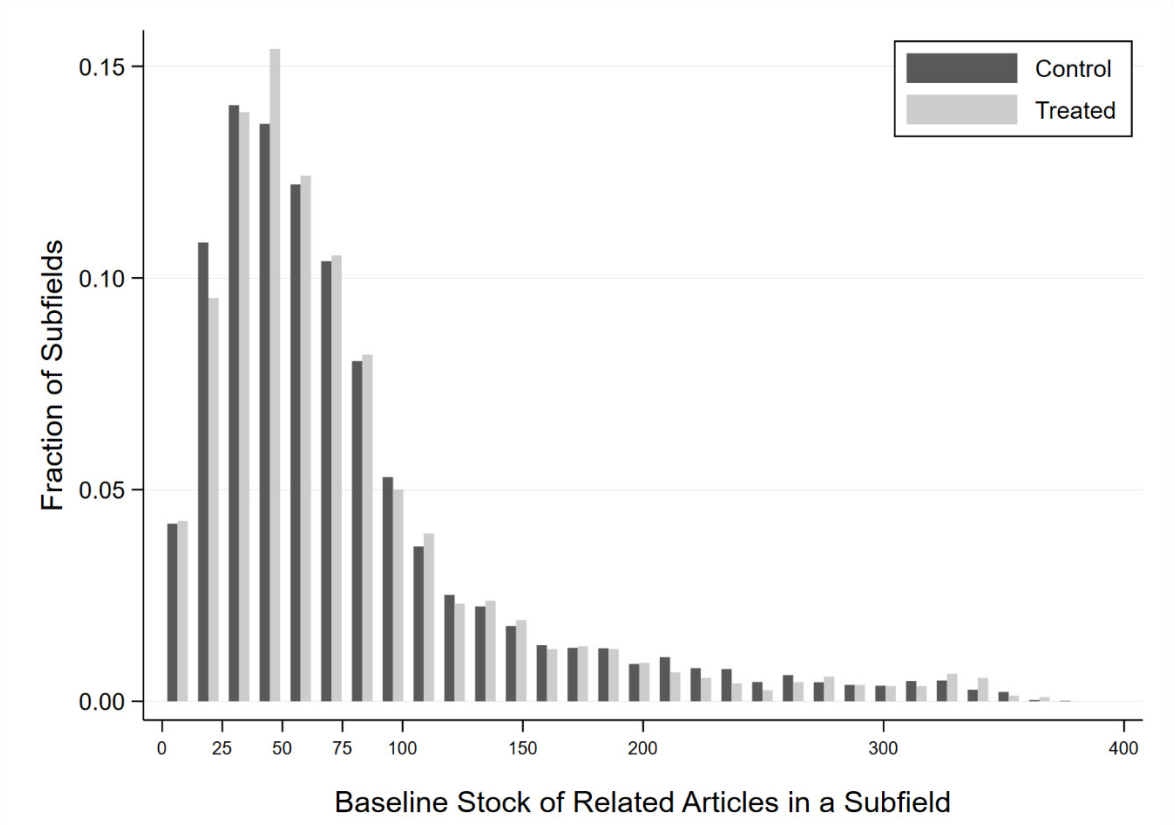
Shwed, Uri and Peter S. Bearman. 2010. "The Temporal Structure of Scientific Consensus Formation." *American Sociological Review*, 75(6), 817-40.

Solow, Robert M. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, 39(3), 312-20.

Wooldridge, Jeffrey M. 1997. "Quasi-Likelihood Methods for Count Data," M. H. Pesaran and P. Schmidt, *Handbook of Applied Econometrics.* Oxford: Blackwell, 352-406.

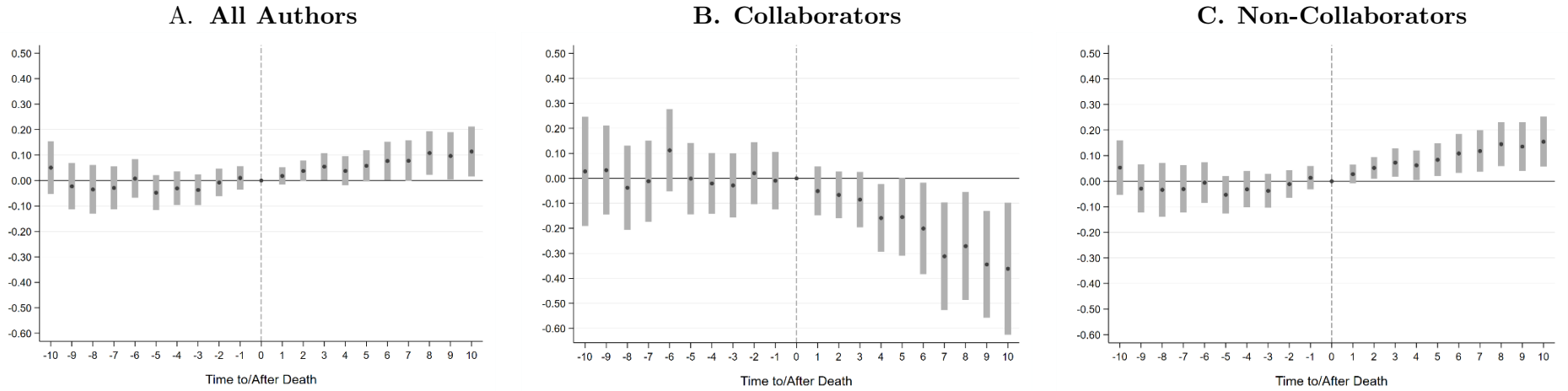
Zuckerman, Harriet A. 1968. "Patterns of Name Ordering among Authors of Scientific Papers: A Study of Social Symbolism and Its Ambiguity." *American Journal of Sociology*, 74(3), 276-91.

Figure 1: Cumulative Stock of Publications at Time of Death



Note: We compute the cumulative number of publications, up to the year that immediately precedes the year of death (or counterfactual year of death), between 3,076 treated subfields and 31,142 control subfields.

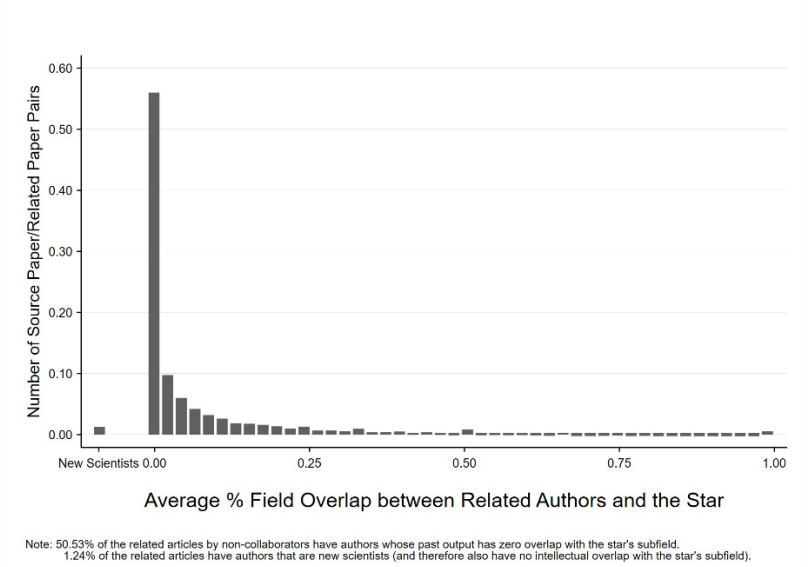
Figure 2
Effect of Star Scientist Death on Subfield Growth and Decline



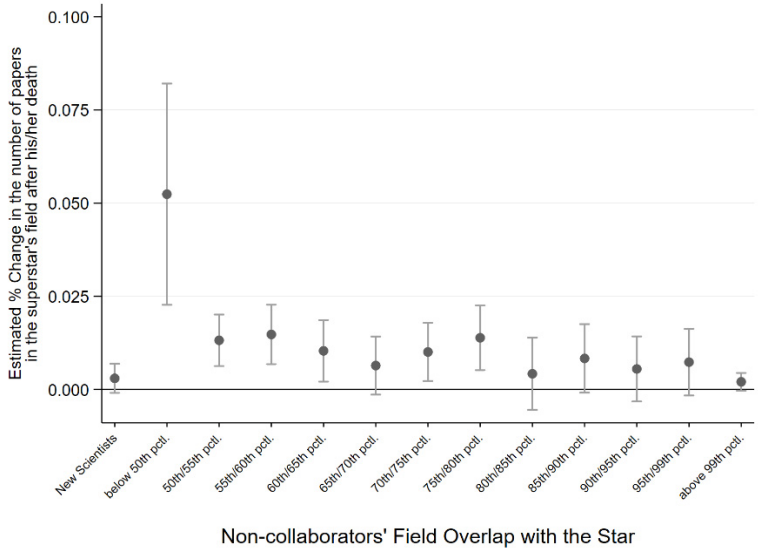
Note: The dark dots in the above plots correspond to coefficient estimates stemming from conditional (subfield) fixed effects Poisson specifications in which publication flows in subfields are regressed onto year effects, subfield age effects, as well as 20 interaction terms between treatment status and the number of years before/after the death event (the indicator variable for treatment status interacted with the year of death is omitted). The specifications also include a full set of lead and lag terms common to both the treated and control subfields to fully account for transitory trends in subfield activity around the time of the death. The 95% confidence interval (corresponding to robust standard errors, clustered at the level of the star scientist) around these estimates is plotted with vertical light grey lines; Panel A corresponds to a dynamic version of the specification in column (1) of Table 3; Panel B corresponds to a dynamic version of the specification in column (2) of Table 3; Panel C corresponds to a dynamic version of the specification in column (3) of Table 3.

Figure 3
Characteristics of Related Authors: Competitors or Outsiders?

A. Distribution of Intellectual Proximity



B. Entering Authors & Field Overlap



Note: Panel A displays the distribution of overlap between the past output of related authors and each star's subfield. For each author on a related article matched to the AAMC Faculty Roster, we create a metric of intellectual proximity by computing the fraction of their publications that belongs to the star's subfield. Slightly more than half of related articles have authors with zero overlap, i.e., this related article is their first contribution to the star's subfield. 1.24% of related articles are authored by new scientists for whom this publication within the subfield is also their first publication overall. Using this information, we aggregate the number of related articles in a particular subfield and in a particular year, e.g., *"the number of articles in the subfield in year t that have authors above the 95th percentile in our measure of field overlap."* In Panel B, each dot corresponds to the magnitude of the treatment effect in a separate regression where the dependent variable is the number of articles in each subfield authored by scientists who belong to a particular bin of intellectual proximity, as measured by field overlap above.

Table 1: Summary Statistics — Deceased Superstar Scientists (N=452)

	Mean	Median	Std. Dev.	Min.	Max.
Year of Birth	1930.157	1930	11.011	1899	1959
Degree Year	1957.633	1957	11.426	1928	1986
Year of Death	1991.128	1992	8.055	1975	2003
Age at Death	60.971	61	9.778	34	91
Female	0.102	0	0.303	0	1
MD Degree	0.403	0	0.491	0	1
PhD Degree	0.489	0	0.500	0	1
MD/PhD Degree	0.108	0	0.311	0	1
Sudden Death	0.409	0	0.492	0	1
Nb. of Subfields	6.805	4	7.308	1	57
Career Nb. of Pubs.	138.221	112	115.704	12	1,380
Career Nb. of Citations	8,341	5,907	8,562	120	72,122
Career NIH Funding	$16,637,919	$10,899,139	$25,441,933	0	$329,968,960
Sits on NIH Study Section	0.007	0	0.081	0	1
Career Nb. of Editorials	0.131	0	0.996	0	17

Note: Sample consists of 452 superstar life scientists who died while still actively engaged in research. See Appendix A for more details on sample construction.

Table 2: Summary Statistics — Control & Treated Subfields at Baseline

	Mean	Median	Std. Dev.	Min.	Max.
Control Subfields (N=31,142)					
Baseline Stock of Related Articles in the Field	76.995	59	64.714	0	384
Baseline Stock of Related Articles in the Field, Non-Collaborators	68.390	51	60.222	0	381
Baseline Stock of Related Articles in the Field, Collaborators	8.604	5	10.358	0	125
Source Article Nb. of Authors	3.970	4	1.901	1	15
Source Article Citations at Baseline	16.331	8	30.305	0	770
Source Article Long-run Citations	70.427	38	116.108	1	4495
Investigator Gender	0.067	0	0.249	0	1
Investigator Year of Degree	1960.546	1962	10.998	1926	1991
Death Year	1991.125	1991	7.968	1975	2003
Age at Death	58.100	58	8.795	34	91
Investigator Cumulative Nb. of Publications	164	131	123	1	1,109
Investigator Cumulative NIH Funding at Baseline	$18,784,517	$11,904,846	$25,160,518	0	$387,558,656
Investigator Cumulative Nb. of Citations	12,141	8,010	12,938	9	157,581
Treated Subfields (N=3,076)					
Baseline Stock of Related Articles in the Field	76.284	58	64.046	0	368
Baseline Stock of Related Articles in the Field, Non-Collaborators	67.752	51	59.725	0	357
Baseline Stock of Related Articles in the Field, Collaborators	8.532	5	9.841	0	86
Source Article Nb. of Authors	3.987	4	1.907	1	14
Source Article Citations at Baseline	16.694	8	36.334	0	920
Source Article Long-run Citations	70.432	35	180.528	1	6598
Investigator Gender	0.099	0	0.299	0	1
Investigator Year of Degree	1960.141	1961	10.898	1928	1986
Death Year	1991.125	1991	7.970	1975	2003
Age at Death	58.100	58	8.796	34	91
Investigator Cumulative Nb. of Publications	170	143	118	12	1,380
Investigator Cumulative NIH Funding at Baseline	$17,637,726	$12,049,690	$24,873,018	0	$329,968,960
Investigator Cumulative Nb. of Citations	11,580	8,726	10,212	120	72,122

Note: The sample consists of subfields for 452 deceased superstar life scientists and their matched control subfields. See Appendix D for details on the matching procedure. All time-varying covariates are measured in the year of superstar death.

Table 3: Effect of Superstar Death on Subfield Entry Rates

	Publication Flows			NIH Funding Flows (Nb. of Awards)		
	All Authors	Collaborators Only	Non-Collaborators Only	All Authors	Collaborators Only	Non-Collaborators Only
	(1)	(2)	(3)	(4)	(5)	(6)
After Death	0.051†	-0.232**	0.082**	0.046	-0.265**	0.110**
	(0.029)	(0.057)	(0.029)	(0.035)	(0.076)	(0.033)
Nb. of Investigators	6,260	6,124	6,260	6,215	5,678	6,202
Nb. of Fields	34,218	33,096	34,218	33,912	29,163	33,806
Nb. of Field-Year Obs.	1,259,176	1,217,905	1,259,176	1,049,942	902,873	1,046,678
Log Likelihood	-2,891,110	-729,521	-2,768,252	-1,350,204	-472,329	-1,223,913

Note: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications in a subfield in a particular year (columns 1, 2, and 3), or the total number of NIH grants that acknowledge a publication in a subfield (columns 4, 5, and 6). All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from zero to one after the death of the star, to address the concern that age, year and individual fixed effects may not fully account for trends in subfield entry around the time of death. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in column (3) imply that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away—a statistically significant $100\times(\exp[0.082]-1)=8.55\%$. The number of observations varies slightly across columns because the conditional fixed effects specification drops observations corresponding to subfields for which there is no variation in activity over the entire observation period.

Robust standard errors in parentheses, clustered at the level of the star scientist. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 4: Scientific Impact of Entry

	Vintage-specific long-run citation quantile						
	All Pubs	Bttm. Quartile	2nd Quartile	3rd Quartile	Btw. 75th and 95th pctl.	Btw. 95th and 99th pctl.	Above 99th pctl.
After Death	0.082**	-0.028	0.008	0.031	0.125**	0.232**	0.320**
	(0.029)	(0.036)	(0.033)	(0.032)	(0.035)	(0.049)	(0.081)
Nb. of Investigators	6,260	6,222	6,260	6,257	6,255	6,161	5,283
Nb. of Fields	34,218	33,714	34,206	34,212	34,210	33,207	21,852
Nb. of Field-Year Obs.	1,259,176	1,240,802	1,258,738	1,258,954	1,258,880	1,221,952	804,122
Log Likelihood	-2,768,252	-689,465	-1,125,555	-1,432,223	-1,469,096	-542,735	-156,519

Note: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications by non-collaborators in a subfield in a particular year, where these publications fall in a particular quantile bin of the long-run, vintage-adjusted citation distribution for the universe of journal articles in *PubMed*. All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from zero to one after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in column (1), Panel A, imply that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away—a statistically significant $100\times(\exp[0.082]-1)=8.55\%$.

Robust standard errors in parentheses, clustered at the level of the star scientist. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 5: Entry and Research Direction

Panel A	**Cardinal Measure**		**Ordinal Measure**	
	Intllct. Proximate Articles	Intllct. Distant Articles	Intllct. Proximate Articles	Intllct. Distant Articles
After Death	0.091**	0.028	0.117**	-0.024
	(0.030)	(0.035)	(0.028)	(0.037)
Nb. of Investigators	6,228	6,099	6,260	6,017
Nb. of Fields	33,375	32,232	34,218	31,712
Nb. of Field-Year Obs.	1,228,157	1,186,589	1,259,176	1,167,423
Log Likelihood	-1,628,374	-1,816,449	-1,893,982	-1,628,170
Panel B	**In-field vs. Out-of-field References**		**Backward Citations to the Star's Bibliome**	
	w/ in-field references	w/o in-field references	w/ references to the star	w/o references to the star
After Death	-0.023	0.128**	0.078*	0.152**
	(0.041)	(0.031)	(0.036)	(0.034)
Nb. of Investigators	6,195	6,260	6,247	6,259
Nb. of Fields	32,721	34,218	34,179	34,147
Nb. of Field-Year Obs.	1,204,315	1,259,176	1,257,747	1,256,576
Log Likelihood	-792,795	-2,510,344	-1,914,448	-1,767,571
Panel C	**Vintage of Cited References**		**Vintage of 2-way MeSH Term Combinations**	
	Young	Old	Young	Old
After Death	0.071*	-0.010	0.090**	0.029
	(0.035)	(0.034)	(0.033)	(0.036)
Nb. of Investigators	6,260	6,260	6,258	6,260
Nb. of Fields	34,218	34,214	34,206	34,210
Nb. of Field-Year Obs.	1,259,176	1,259,044	1,258,732	1,258,906
Log Likelihood	-2,124,598	-1,613,457	-1,853,062	-1,784,275

Note: Estimates stem from conditional (subfield) fixed effects Poisson specifications. In Panel A, the dependent variable is the total number of publications by non-collaborators in a subfield in a particular year, where these publications can either be proximate in intellectual space to the star's source publication, or more distant (in the PMRA sense). Since PMRA generates both a cardinal and an ordinal measure of intellectual proximity, we parse the related articles using both measures, yielding a total of four different specifications. For the cardinal measure, a related article is deemed proximate if its similarity score is above .58, which corresponds to the median of relatedness in the sample. For the ordinal measure, a related article is deemed proximate if its similarity rank is below 90, which also corresponds to the median of similarity in the sample. In Panel B, we focus on whether the content of entrants' contributions in the subfield change after the superstar passes away. Each cited reference in a related article can either belong to the subfield, or fall outside of it; it can cite a publication of the star scientist associated with the subfield, or fail to cite any of the star's past contributions. In Panel C, the dependent variable is the total number of publications by non-collaborators in a subfield in a particular year, where these publications can either be "fresh" (citing young references, or being annotated by MeSH terms of recent vintage) or stale (citing old references, or being annotated by MeSH terms of distant vintage). All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from zero to one after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in the first column of Panel A imply that treated subfields see an increase in the number of PMRA-proximate contributions by non-collaborators after the superstar passes away—a statistically significant $100\times(\exp[0.091]-1)=9.53\%$. Robust standard errors in parentheses, clustered at the level of the star scientist. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 6: Breakdown by Star Scientist Characteristics

	Publications		Citations		Funding		Importance to the Field	
	Below Median	Above Median	Below Median	Above Median	Below Median	Above Median	Below Median	Above Median
After Death	0.059	0.116*	0.036	0.125**	0.014	0.162**	0.063*	0.123**
	(0.037)	(0.050)	(0.042)	(0.040)	(0.040)	(0.052)	(0.031)	(0.045)
Nb. of Investigators	2,901	4,836	2,792	4,619	3,048	4,287	5,019	4,493
Nb. of Fields	17,210	17,008	17,328	16,890	15,731	15,487	16,985	17,233
Nb. of Field-Year Obs.	632,089	627,087	636,750	622,426	578,277	570,665	625,140	634,036
Log Likelihood	-1,377,727	-1,387,650	-1,367,335	-1,396,652	-1,268,559	-1,252,952	-1,462,538	-1,257,973

Note: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications by non-collaborators in a subfield in a particular year. Each pair of columns splits the sample across the median of a particular covariate for the sample of fields (treated and control) in the baseline year. The table examines differences in the extent to which the eminence of the star at death (respectively counterfactual year of death for controls) influences the rate at which non-collaborators enter the field after the star passes away. Eminence is measured through the star's cumulative number of publications, the star's cumulative number of citations garnered up to the year of death, and the star's cumulative amount of NIH funding. We also have a "local" measure of eminence: the star's importance to the field, which is defined as the proportion of articles in the subfield up to the year of death for which the star is an author. All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from zero to one after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimate in the second column implies that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away—a statistically significant $100\times(\exp[0.116]-1)=12.30\%$.

Robust standard errors in parentheses, clustered at the level of the star scientist. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 7: The Nature of Entry Barriers

Panel A	Subfield Coherence					
	PMRA-based definition		Citation-based definition		Cliquishness	
	Below Median	Above Median	Below Median	Above Median	Below Median	Above Median
After Death	0.202**	0.067	0.161**	0.096*	0.129**	0.064
	(0.038)	(0.048)	(0.053)	(0.041)	(0.049)	(0.052)
Nb. of Investigators	3,353	3,203	3,422	3,157	2,865	3,561
Nb. of Fields	9,062	7,828	8,731	8,159	8,044	8,846
Nb. of Field-Year Obs.	334,142	288,284	321,826	300,600	296,704	325,722
Log Likelihood	-711,335	-664,170	-760,842	-631,287	-692,330	-685,682

Panel B	Indirect Control through Collaborators					
	Editorial Channel		NIH Study Section Channel		Fraction of Subfield NIH Funding	
	Below Median	Above Median	Below Median	Above Median	Below Median	Above Median
After Death	0.147**	0.086†	0.134**	-0.078	0.174**	0.084
	(0.056)	(0.048)	(0.043)	(0.095)	(0.051)	(0.051)
Nb. of Investigators	3,452	2,068	4,385	664	3,559	2,525
Nb. of Fields	11,110	5,780	15,338	1,552	9,863	7,027
Nb. of Field-Year Obs.	410,025	212,401	565,219	57,207	363,690	258,736
Log Likelihood	-951,705	-461,769	-1,293,997	-125,950	-840,777	-545,782

Note: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications by non-collaborators in a subfield in a particular year. The sample is limited to the subfields in which the most eminent among the stars were active (specifically, above the median of the "cumulative citations up to the year of death" metric). Each pair of columns splits the sample across the median of a particular covariate for the sample of subfields (treated and control) in the baseline year. For example, the first two columns of Panel B compare the magnitude of the treatment effect for stars whose collaborators have written an above-median number of editorials in the five years preceding the superstar's death, vs. a below-median number of editorials. All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from zero to one after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in the first column of Panel B imply that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away—a statistically significant $100\times(\exp[0.147]\text{-}1)=15.84\%$.

Robust standard errors in parentheses, clustered at the level of the star scientist. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.