Matthew: Effect or Fable?

Pierre Azoulay
MIT & NBER
Sloan School of Management
100 Main Street, E62-482
Cambridge, MA 02142

Toby Stuart
University of California – Berkeley
Haas School of Business
2220 Piedmont Avenue
Berkeley, CA 94720

Yanbo Wang
Boston University
School of Management
595 Commonwealth Avenue
Boston, MA 02215

December 10, 2012

Abstract

In a market context, a status effect occurs when actors are accorded differential recognition for their efforts depending on their location in a status ordering, holding constant the quality of these efforts. In practice, because it is very difficult to measure quality, this ceteris paribus proviso often precludes convincing empirical assessments of the magnitude of status effects. We address this problem by examining the impact of a major status-conferring prize that shifts actors positions in a prestige ordering. Specifically, using a precisely constructed matched sample, we estimate the effect of a scientist becoming a Howard Hughes Medical Investigator (HHMI) on citations to articles the scientist published before the prize was awarded. We do find evidence of a post-appointment citation boost, but the effect is small and limited to a short window of time. Consistent with theories of status, however, the effect of the prize is significantly larger when there is uncertainty about article quality, and when prize-winners are of (relatively) low status at the time of appointment to HHMI.

Keywords: sociology of science, status, stratification, Matthew effect.

* We gratefully acknowledge the financial support of the National Science Foundation through its SciSIP Program (Award SBE-1063907). We thank Matthew Bothner, Stanislav Dobrev, Benjamin Jones, Michael Hannan, Casey Ichniowski, and Ezra Zuckerman, as well as seminar participants at Columbia, Berkeley, Stanford, MIT, University of Utah, UC Irvine, and Dartmouth, for many helpful comments. The usual disclaimer applies.

1 Introduction

Status orderings are ubiquitous in social and economic life. These hierarchies are of great interest because of their role in generating and reproducing inequality in social and economic outcomes, which occurs because an actors status often is a lens through which critical audiences form judgments about the actors quality. In consequence, status orderings can become self-perpetuating; because status influences perceptions of quality, those of high status often garner favorable assessments, which then reifies their positions atop the very status ordering that first served as an allocation mechanism for quality appraisals. Of course, the converse is true for those on the bottom rungs of the status ladder.

This idea has animated much research. Merton (1968) famously developed this argument in the sociology of science. He posited that small differences in initial status amplify over time to generate cumulative advantages. In Mertons classic account, not only does status itself influence perceptions of quality, but high status scientists are more likely to attract tangible resources, such as research funding and outstanding graduate students, which can then be parlayed into scientific outputs of higher quality. Of course, although it has proved to be a fertile research site, work on status extends well beyond the study of science. For example, Podolny (1993; 2005) and colleagues (e.g., Podolny, Stuart, and Hannan 1996) have examined these ideas in investment banking and semiconductors. Among recent studies with convincing empirical designs, Waguespack and Sorenson (2011) study the film industry, Hsu (2004) examines venture capital firms, Simcoe and Waguespack (2011) study a standards-setting body, and Roberts, Khaire, and Rider (2011) investigate wine producers.

Despite the general consensus about status dynamics, however, much of the nonexperimental empirical evidence on the consequences of social status remains fragile. This is because of the intricate coupling between an actor's quality and status, which engenders the question: Does status itself affect outcomes, or is status simply a byproduct of quality? In much of the work on the subject, there is an assumed feedback loop between these two constructs; status rankings may first emerge from quality distinctions among actors or differences in their social or ascriptive attributes, but these characteristics then interact (cf. Lynn,

Podolny, and Tao 2009). In many settings, causality seems as likely to flow from status to quality as it is to travel in the reverse direction. Therefore, few of the archival studies of the performance effects of status present evidence that would persuade a skeptic.

Our contribution in this paper is a research design that offers a more definitive test of the effect of a change in status on performance. To foreshadow our argument, assume a context with a group of producers, each of whom generates a set of products. Producers might be investment banks and their products underwritten securities (Podolny 1993); producers might be scientists and their products journal articles (Merton 1968); they could be semiconductor firms that produce innovations (Podolny and Stuart 1995); they could be engineers who draft standards documents (Simcoe and Waguespack 2011); or they could be law firms that employ attorneys from different calibers of law schools (Phillips and Zuckerman 2001). In the typical archival study of social status, there is a measure of producer-level status that is either derived from a social network among producers (e.g., Podolny 1993; Roberts and Sterling 2012) or that is an aggregation over measures of product-level status (e.g., Podolny and Stuart 1995).

Here we examine the effect of scientists winning a highly desired recognition, which results in a shift in their status. We analyze the effect of this prize on the subsequent-to-award citation rates to journal articles that were published before the award was granted. This research design has two advantages. First, because the producers (scientists) in the data have created thousands of products (articles), we can generate a control group of papers that contains nearly exact matches to articles written by prize-winning scientists. This enables us to net out the effect of product-level (article) quality when we estimate the effect of a change in producer-level status. Second, because the prizes we study are awarded after scientists already are well established in their careers, we can further restrict the analysis to the effect of a status shock on articles that were written before the award was bestowed. As we discuss below, the benefit of limiting the analysis to the effect of the award on pre-existing products is that we can restrict the operative mechanism to changes in perceptions of quality, versus the potential effect of an elevation in status on enhancing the actual quality of products that are created after the status change. Because the research setting offers a measure of a status

change and relatively precise measures of product quality, we can isolate what we believe to be an almost-pure status effect. The payoff of this research design is that we can present a narrow test of Merton's (1968: 58) often-quoted description:

> *"...the Matthew effect consists in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark."*

In our research design, we measure "increments in recognition" (changes in citations to papers) as the dependent variable, and a change in status (winning a prize) as the central explanatory variable.

We feature four findings. First, results can be quite misleading when, as is typical in the literature, the effect of status is measured at the producer level without adjustments for product-level quality. When we analyze citation rates to articles while controlling only for scientist-level quality, we observe a relatively large effect of the status shock. This effect falls significantly when we isolate the result to be net of scientist- and article-level controls for quality.[1] Therefore, in many settings, datasets that disaggregate producers into component product offerings may be necessary to estimate a reliable status effect. Second, the effect of the status shock we observe on the deference accorded to a producer's products is smaller than we had anticipated, and it is largest for products that were produced near to the time that the status shock occurred. Third, as theory suggests, we find that the magnitude of the status effect is contoured by the level of uncertainty around product quality. The status shock has a larger effect on citations to papers that were published in low impact journals; it is larger for articles that were written in novel areas of science; and it is larger for articles that combined ideas from multiple fields of scientific endeavor, rather than those that drew from a single field. Therefore, the findings support the idea that status is a social cue that conveys the greatest information to audiences when there is uncertainty about product quality (Stuart et al. 1999). Finally, we show that there are ceiling effects in the data (cf. Bothner, Smith and White 2010; Bothner, Podolny and Smith 2011). Specifically, the effect

[1] In a compelling article we describe in greater detail below, Simcoe and Waguespack (2011) present a similar result. These authors show that the standard approach can significantly over-estimate the true effect of status.

of a status shock is greater for actors whose pre-award positions in the status ordering are less-well established at the time they garner recognition.

2 Theoretical Background

Sociologists and social psychologists have demonstrated that status orderings are an always-present feature of social life (Gould 2002; Bothner, Godart, and Lee 2009). For instance, those who study small group dynamics find that even in a collection of strangers, a status hierarchy emerges almost immediately upon the designation of the individuals as members of a common group (e.g., Bales et al. 1951). The literature describes status hierarchies across all types of social settings, from the schoolyard to the marketplace, from ephemeral groups to occupational communities, from hospitals to street gangs. The differentiation of actors into positions in status orderings truly does permeate social life.

A number of theories explain the emergence of these hierarchies. In Merton (1968), Podolny (1993), Gould (2002) and other formulations of status theory, the prestige hierarchy takes shape from audience members' perceptions of quality, but it can quickly diverge from the distribution of actual quality. In theories in which perceived quality determines position in the status ordering, there is one essential—albeit not particularly limiting—boundary condition: there must be some degree of uncertainty among the members of the social system regarding their ability to assess actual quality.[2] Actors then resort to social cues to resolve their uncertainty about a producer's underlying quality. Through a variety of mechanisms, the reliance on social cues to form perceptions of quality can drive a wedge between the distributions of status (as perception) and quality (as reality).

One approach to resolve uncertainty about another's quality is to infer it from the identities of a producer's affiliates (Blau 1955). When this transpires, the perception of a focal producer becomes an attribution in part made from the prominence of its affiliates. The

[2]The term "uncertainty" often is not precisely defined in research on social status, but its use generally is consistent with the idea that the distribution from which the expectation of product quality is drawn is unknown, rather than that there is greater dispersion around a known mean for quality. In other words, the conception of uncertainty in the status literature is more akin to Knightian uncertainty than to the classic notion of risk.

implicit endorsement conveyed by the fact that an actor of high regard enters an association with a focal producer inevitably shapes how others perceive that actor. In fact, this dynamic has been central in accounts of scientific careers. For instance, given their short and often non-diagnostic tenure in the profession, recently minted PhDs frequently are assessed according to the prestige of their mentors or the status of the university in which they gain employment.

Another reason that a producer's status and quality may decouple—and one that is central to the research design in this paper—is that status often is amplified by designations, such as prizes, which create break points in smooth quality distributions. This is an important element of Merton's (1968) discussion of the Matthew effect. He writes that highly prestigious accolades such as the Nobel Prize engender almost-capricious status distinctions. To illustrate this point, Merton (1968:56) identified what he labeled, "the phenomenon of the 41st chair." He wrote:

> *"The French Academy, it will be remembered, decided early that only a cohort of 40 could qualify as members and so emerge as immortals. This limitation of numbers made inevitable, of course, the exclusion through the centuries of many talented individuals."*

He posits that the fixed number of members of the French Academy (or more generally, the tight supply of places in all recognitions of distinction) causes a disjuncture between status and quality, in which the status of those who are awarded membership jumps significantly, possibly far above any actual quality difference that separates them from others who were on the cusp of recognition. Merton's discussion of the 41st chair suggests a fascinating experiment: How does the post-award perception of status-enhancing prize winners compare to that of the producers who were in the consideration set, but failed to achieve consecration (cf. Simcoe and Waguespack 2011)? Or, conversely, if we contemplate the counterfactual, how would the career outcomes of the 41^{st} chairs have differed if, contrary to the fact, they had been elected to the French Academy?

In attempting to answer this question, it is important to recognize that there are multiple routes through which a change in a producer's status, such as election to the French Academy, can affect outcomes. Here, we describe two such paths. First, for reasons already discussed,

changes to a producer's status may directionally influence perceptions of the quality of the producer's products, and thus set in motion socially endogenous (to the status distribution) assessments of producer quality. In effect, the prize consecrates the producer. Second, a change in status often affects actual (versus just perceptions of) quality. Most importantly, this occurs because one of the benefits of status is that those who possess it attract resources, and thereby experience an enhanced ability to create goods of high quality.

Once again, Merton (1968) describes both pathways in the context of scientific careers. In addition to the fact that high status scientists garner greater acknowledgment for an article of a given level of quality than would a lower-status scientist, Merton also argues that the higher status scientist is more likely to gain access to the resources that are consumed in scientific production. For instance, prestigious scientists are more likely to find positions at elite universities, which attract the best students and possess state-of-the-art facilities. Through these and other resources that disproportionately flow to high status producers, those who occupy positions at the top of the prestige hierarchy often have the means to produce higher quality goods. In the empirical work to follow, we present an analysis that closely conforms to Merton's 41^{st} chair thought experiment. Formulating the inquiry in broader terms, we seek answers to three questions:

1. Does a shock to a producer's status truly cause others to change their perceptions of the quality of a producer's products?

2. Does the extent to which audience members update their perceptions of quality after a status shock depend on the uncertainty surrounding producer and/or product quality?

3. Does the extent to which audience members update their perceptions of quality after a status shock depend on a producer's initial position in the status hierarchy, such that there are greater returns to lower-status (at the time of recognition) prize winners?

In pursuing answers to these three questions, we are able to narrow our focus to the question of how a shock to status influences perceptions of quality. That is, in the aforementioned pathways through which a status shock affects outcomes—by altering other-party percep-

tions of a given level of quality or by attracting resources that are invested to produce higher quality goods—we strive to narrowly focus the empirical test on the former mechanism.

3 Empirical Approach

Our research design differs from most of the archival research on the effects of status. Therefore, before providing the full details of the methodology, we present a high-level roadmap of the empirical approach.

First, we have identified a set of producers/scientists who are recipients of a prize that boosts their status. Second, we create a control group of scientists who were close contenders for the prize, but were not selected. Third, we collect data on all products/journal articles written by prize winners and control group members. For prize winners, we limit the data analysis to articles that were written before the prize was awarded, which guarantees that the quality of the articles in the dataset could not have been influenced by resources that are connected to the prize. This is a pivotal aspect of the research design because it is the means by which we exclude the resource-based pathway on the measured effect of status.

Fourth, we create a sample of matched pairs at the product level, in which we match prize winners' articles to those of control group members on a set of criteria that, we assert, makes it likely that the actual quality of the two papers in a pair are very similar. This results in a matched sample comprising pairs of very similar articles, with one belonging to a prize winner and the second authored by a prize contender who did not receive an award. The final step of the analysis is to assess whether the prize—a shock to a producer's status—affects third-party perceptions of quality, relative to the control group of equal-quality products.

3.1 Identifying Status Effects

Returning to the existing literature, the most widespread approach to identifying the magnitude of status effects is to focus on a set of organizations, people, or products that vary in status. When the researcher observes this variation, the analyst can then regress a measure

of performance on status. With detailed controls for characteristics that might correlate with both status and performance, it is in principle possible to quantify the benefits of status. However, because variation in status often occurs only in the cross-sectional dimension of the data, these estimates may incorporate the confounding influence of omitted variables, such as producers' true quality or resource endowments.[3]

One recent article provides a compelling approach to address this causality problem. Simcoe and Waguespack (2011) examine the diffusion of internet standards through proposals brought before the Internet Engineering Task Force (IETF). These authors exploit a natural experiment to estimate the effect of author status. Specifically, they identified a context in which the use of *et al.* occasionally and randomly obscures the identity of authors who submit proposals to the IETF. Simcoe and Waguespack find that when the name of a high status authors is excluded, the likelihood that a proposal is downloaded drops relative to cases in which the identity of a low status author is concealed. In effect, the use of *et al.* in this context creates a manipulation in the status signal that audience members are able to observe.

We are not aware of any natural experiment in the context we study. Instead, we exploit a setting in which we can unambiguously isolate the timing of a one-time jump in a producer's status, and then examine the benefits that accrue to the same producer and his/her products, before and after the shock. This longitudinal (before-to-after) contrast purges our estimates of most sources of omitted variable bias that plague cross-sectional comparisons, but it remains possible that there is a second estimation problem: the timing of the shock itself may be endogenous. Specifically, when status shocks incorporate forecasts about actors' expected performance improvements, the standard difference-in-difference estimate can be

[3] A number of studies estimate the effect of status on a performance outcome while including actor fixed effects (e.g., Podolny, Stuart and Hannan 1996). Under a strict set of assumptions, these studies address the problem of unobserved quality differences among producers. For the fixed effects estimator to yield informative coefficient estimates, producer status positions must change meaningfully during the time period of the study so that there is adequate within-producer variation to estimate the effect of status. In addition, quality differences must be non-time-varying, or the fixed effects will not resolve the problem of unobserved differences between producers. In general, status theory suggests that status positions are far stickier than producer quality (Podolny 2005), which calls into question the ability of fixed effects estimators to solve the measurement problems in the status literature.

unreliable. To understand why, consider our case—the bestowing of a prestigious accolade. It is possible, perhaps even likely, that the evaluators who award such a prize are aware of the performance trends of the members of the application pool. If the individuals in the pool on the best trends are selected to win the prize, any estimated change seemingly caused by the prize simply may reflect the selection of the best producers into the treatment condition, rather than a causal effect of the status change *per se.*

To remedy this problem, we pair each product with a control that both provides a very near match based on time-invariant characteristics and exhibits an almost-identical performance trend prior to the status shock. When we analyze the data at the matched-pair level of analysis, a difference-in-difference framework provides a flexible and non-parametric methodology to evaluate the effects of the status shock. In fact, conditional on the assumption that the matching algorithm we employ successfully pairs products of comparable quality, we are able to present the findings in a straightforward, graphical form.

3.2 Status Shocks and Their Associated Counterfactuals

The academic sciences provide an ideal laboratory for our study. First, the production of scientific knowledge is a classic context for investigations of status and its effects, beginning with Merton's original statement of the Matthew effect, and continuing with many of his intellectual disciples (e.g., Cole and Cole 1968; Allison, Long, and Krauze 1982). Second, this setting provides a clear distinction between individual producers (scientists) and their products (scientific articles). Third, scientists may garner shocks to their status at several career stages, in the form of prizes or election to prestigious societies. Finally, the flow of citations to scientific articles provides a metric to evaluate the effects of status, since by citing another piece of knowledge, producers inscribe into their own products esteem for their peers.

HHMI Appointment. We analyze the effect of a shock to the status of mid-career academic life scientists in the U.S.—appointment to be investigators of the Howard Hughes Medical Institute (HHMI). HHMI, a non-profit foundation, is a major participant in the

biomedical research funding ecosystem. Indeed, the Institute's annual budget is larger than the amount the NSF typically commits to the biological sciences. During periodic, open competitions, the Institute solicits applications from scientists in the U.S. The selection committee for HHMI almost exclusively comprises members of the National Academy of Sciences, so the profession's elite scientists choose prize winners. Once selected, awardees continue to be based at their home institutions, but they are entitled to append the prestigious "& HHMI" to their affiliation, so that other scientists are reminded of their status.[4]

Appointment to HHMI is a major honor. Consistent with its stature, HHMI appointment is a harbinger of greater accomplishment: the current group of HHMI investigators includes 16 Nobel laureates and 152 members of the National Academy of Sciences.

HHMI's award cycles last five years and typically are renewed at least once. Appointees also are recipients of large research budgets and may benefit from intangible assistance such as editorial goodwill and access to advice from an elite peer group (Azoulay et al. 2011). As such, HHMI appointment combines status with other forms of resources, and it is therefore likely to affect both the perceived and actual quality of a prize winner's work. To separate these two effects, the bulk of our analysis focuses on the consequence of the prize for the citation trajectories to articles that were written before the award was granted. We do this because future appointment to HHMI cannot influence the actual quality of pre-existing work; it only can affect the perception of and/or the attention directed to past work.

The Producer Control Group: Early Career Prize Winners. Given the high degree of accomplishment exhibited by HHMI investigators at the time of their appointment, a random sample of scientists of the same age and scientific fields would not be fitting as a control group. We therefore construct a control group comprising only scientists who received early career prizes that are awarded in the same subfields of the life sciences as HHMI. The Pew, Searle, Beckman, Packard, and Rita Allen Scholarships all are early career prizes that target scientists in the same life science subfields and similar research institutions as HHMI. These scholarships are accolades that young researchers can receive in the first years of their

[4]The subfields of the life sciences of interest to HHMI investigators have tended to concentrate on cell and molecular biology, neurobiology, immunology, and biochemistry.

independent career. We label members of this control group "Early Career Prize Winners," or ECPWs.[5]

3.3 Data Sources

Individual scientist data. We start from the set of all 443 HHMI investigators appointed between 1984 and 2003. We eliminated 16 scientists who left academe and three investigators who work in a small field (computational biology) because we are unable to find controls in that field. We track the careers of the remaining 424 scientists from their first positions as independent investigators until 2007. We do so through a combination of curriculum vitæ, NIH biosketches, *Who's Who* profiles, National Academy of Sciences biographical memoirs, and Google searches. For each one of these individuals, we record employment history, degree held, date of degree, gender, and up to three departmental affiliations. To construct the control sample we proceeded in parallel fashion to track the careers of ECPW scientists. The final sample contains 2,375 early career prize winners.

The timing of appointment for HHMI investigators is identified from the HHMI web site and scientists' CVs, rather than inferred from affiliation information in published articles. To be precise, we know the calendar year in which each investigator joined HHMI's payroll, but not the exact date. We adopt the following convention: we define the baseline year for each treated scientist as the year that precedes the appointment year listed on HHMI's web site. Although some of the publications that appear in the year of appointment in fact correspond to pre-appointment output, we wish to avoid the mistaken assignment of post-appointment output to the pre-award period. Because appointment carries access to resources that may enhance the actual quality of work, our claim about causal mechanisms will depend on strictly limiting the analysis to articles that were published prior to appointment. In this respect, we err on the side of caution: our assignment of the treatment date guarantees that

[5]In addition to the career stage at which they are bestowed, these prizes differ from HHMI investigatorships in one essential respect: they are structured as one-time grants. The corresponding amounts are relatively small, roughly corresponding to 35% of a typical NIH R01 grant.

all articles in the matched pair sample we analyze were written before the treated scientist was appointed to HHMI.

Article data. The second step in the construction of our dataset is to link scientists to journal articles. We collect articles from PubMed, a comprehensive bibliographic database covering the fields of the life sciences. In practice, the challenge in using these data is name uniqueness: common names make it difficult to distinguish between scientists, and scientists with relatively rare names sometimes are inconsistent in their use of publication names. We resolve this problem by designing a set of customized search queries for all 424 HHMIs in the treated group and all 2,375 EPCWs in the control groups, which enhances the accuracy of each scientist's bibliome. Details on the data and linking process are provided in Appendix I (cf. supplementary online material).

We begin by downloading all publications of the HHMI and EPCW scientists using the customized queries. We then eliminate from the consideration set letters, comments, reviews, and editorials. Next, we eliminate all articles published 11 or more years prior to the date of the earliest appointment to HHMI in the sample (1984); similarly, we eliminate all articles published after 2003 (the latest HHMI competition we record) so that we always observe a minimum number of three years of citation information for each article.

From Control Producers to Control Products: A Non-Parametric Matching Procedure. A key aspect of our empirical approach is to unbundle producers' status from their products. Empirically, we begin with all products (articles) of "treated" producers (HHMIs) and then we search for nearly exactly matching products (articles) written by control group producers (EPCWs). The goal of the construction of this matched sample is to select a set of articles that pin down the citation trajectories associated with HHMI investigators' papers had they, contrary to the fact, not been awarded this prize.

In practice, identifying close matches is difficult. Because we are interested in the fate of articles, but the status shock we observe occurs at the scientist-level of analysis, semi-parametric matching techniques such as the propensity score and its variants are of limited

use in this context.[6] Instead, we use a non-parametric matching approach, so-called "Coarsened Exact Matching" (CEM) (Iacus, King, and Porro 2011; Blackwell et al. 2009).

The selection of controls proceeds as follows. The first task is to choose a relatively small set of covariates on which we would like to guarantee balance between the treatment and control group. The second step is to create a large number of strata to cover the entire support of the joint distribution of the covariates selected in the previous step. Next, each observation is allocated to a unique stratum; any stratum that either has no articles written by an HHMI, or that has less than five potential control articles, is then dropped from the data. Finally, we select in each stratum a unique control article such that the sum of squared differences in citation flows between the treated and control article from the year of publication until the year preceding the time of appointment is minimized. We break ties at random when there are several candidate articles that minimize this distance metric.

The procedure is coarse because we do not precisely match on covariate values; rather, we coarsen the support of the joint distribution of the covariates into a finite number of strata, and we match a treated observation if and only if a control observation can be found in the same stratum. An important advantage of CEM is that the researcher can guarantee the degree of covariate balance ex ante. However, the more fine-grained the partition of the support for the joint distribution (i.e., the higher the number of strata incorporated into the analysis), the larger the number of unmatched, treated observations. In general, the researcher must trade off the quality of the matches with external validity: the longer the list of matching covariates, the more difficult it is to identify a "fraternal twin" for each article in the treatment group.

Control articles are selected if they share the following characteristics with treated articles: (1) year of publication; (2) specific journal (e.g., *Cell* or *Science*); (3) number of authors; (4) focal-scientist position on the authorship list; and (5) the cumulative number of

[6] A propensity score approach would entail estimating the probability of treatment—becoming an HHMI—and then using the inverse of this estimated probability to weight the data in a second stage analysis of the effect of the HHMI on subsequent citation rates. However, because citations occur at the article level, achieving covariate balance by weighting the data by the scientist-level likelihood of winning the prize, even if the determinants of winning were fully observable, would not resolve the problem of controlling for article-level quality.

citations the articles received between the time they were published and the year the treated scientist was appointed to HHMI. Implementation details can be found in the supplementary online material (Appendix II).

We start from a universe of 195,865 articles published by HHMI or ECPW scientists. Of these 195,865 papers, 7,469 are pre-appointment publications written by HHMI investigators prior to the time they receive the prize. We sought to match this group of treated articles to twins in the set of papers written by control group members, and we succeeded with 3,636 of these 7,469 publications (47.80%). This relatively low rate is to be expected because non-parametric matching procedures are prone to a "curse of dimensionality" whereby the proportion of matched observations decreases rapidly with the number of strata that are imposed.[7]

Citation data. We match PubMed with Thomson-Reuters' *Web of Science* database between the years 1965 and 2007 to generate a dataset with 190 million cited-to-citing paper pairs (details regarding the corresponding cross-walk file can be found in the supplementary online material, Appendix III). Before conducting the statistical analysis, we first eliminate all self-citations from any member of the authorship team. Next, we match the citing article to another database we have assembled that contains all publications of members of the NAS. Whenever a citing article is authored by at least one scientist who was a member of the NAS (or who was previously appointed as an HHMI investigator), we flag this citation as being "high-status." This enables us to decompose the total number of citations flowing to individual articles at a given point in time into an "ordinary" and a "high status" set. With additional processing, these data also enable us to distinguish repeat citers of a line of work from new ones, or citers in the same precisely defined scientific field vs. citers from outside the field. These data enable us to further explore the mechanisms that generate the core set of findings.

[7]For instance, matching on one additional indicator variable, scientist gender, drops the match rate to about 30%. Likewise, only 10.7% of the pairs would remain if we constrained the degree dates for the control and treated investigator to be at most one year apart from one another. Conversely, if we relaxed the constraint that articles for the treated and control groups are drawn from the same scientific journal, the match rate would jump to 70%. Loosening this constraint, however, would come at the expense of the internal validity of the findings.

3.4 Descriptive Statistics.

The final sample contains 3,636 treated articles and 3,636 control articles. The average article was written a number of years before HHMI appointment and we observe it for multiple years after, so the final dataset contains 144,890 article-year observations. We report descriptive statistics in Table 1. For the sake of computing descriptive statistics, we measure all time-varying covariates at baseline. Recall that baseline is defined to be the year preceding appointment for HHMI investigators. A control article inherits the appointment year associated with its treated article match, resulting in a counterfactual appointment year for the ECPW scientist who authors this control article.

Based on the descriptive statistics, four facts merit attention. First, article-level, time-invariant characteristics are very closely matched between treated and control groups. For some covariates (e.g., number of authors, focal author position, article age), this is a mechanical reflection of the CEM procedure, but for others (e.g., the article's novelty as indicated by MeSH keyword vintage, as described below), the close match occurs incidentally. Second, the distribution of citations received at baseline is also very similar between the treated and control papers. Third, as we would expect when we create a paper-level control group, covariate balance does not fully extend to scientist-level characteristics, such as PhD or MD graduation year, though the treatment and control groups appear well-balanced on most characteristics, including the number of "hit articles" they have previously published at baseline.[8] Fourth, one can only discern a small difference in the cumulative number of citations received ten years after appointment—the dependent variable of interest.

In short, the comparisons between control and treated observations illustrate that our matching procedure succeeds at the product-level, rather than producer-level. Imposing a match on a full suite of producer characteristics in addition to article-level covariates would result in a very low match rate. Conversely, one could modify the procedure to achieve a closer match on focal scientist characteristics, but the articles matched in this way would

[8]We classify a paper as a hit if its cumulative citation count in 2008 places it above the 95^{th} percentile of the article-level citation distribution for the focal article's birth year, where the citation count distribution includes all publications in the universe of PubMed for the given birth year.

differ in the pre-appointment flow of citations. By restricting the set of potential control producers to early career prize winners, and then imposing a very close match at the article level, we seek a balance between internal and external validity.

3.5 Statistical Approach

A natural starting point for an analysis of the effect of HHMI appointment on citation trajectories would be to run regressions using all article-year observations (treated and control) as the estimation sample, with article fixed effects. If we followed this approach, because we have several cohorts of HHMI investigators in the sample (appointment years are staggered from 1984 to 2003), the control group that pins down the counterfactual vintage and calendar time effects for the articles that were written by currently appointed HHMI investigators would contain three categories of articles: (i) articles written by EPCWs; (ii) articles by scientists who will become HHMI investigators in the future; and (iii) articles written by HHMI investigators who were appointed in earlier periods. The articles that are part of the last two categories are problematic controls, since they were treated in the past or will be treated in the future. If HHMI appointment events influence citation trends (rather than just levels), the fixed effects estimates will reflect in part this unwanted source of variation, occluding the interpretation of the results.

To produce an analysis in which the control group solely consists of articles written by ECPW scientists, we perform the statistical analysis at the article-pair level. Specifically, the outcome variable is the difference between the citations received in a given year by a treated article and its associated control identified in the matching procedure described above. Let i denote an article associated with an HHMI scientist and let $i^{'}$ index the corresponding control article. Then our estimating equation relates $\Delta CITES_{ii't} = CITES_{it} - CITES_{i't}$ with the timing of HHMI appointment in the following way:

$$E\left[\Delta CITES_{ii't} | X_{ijt}\right] = \beta_0 + \beta_1 AFTER\_HHMI_{jt} + f(AGE_{jt}) + \gamma_{ii'} \quad (1)$$

where $AFTER\_HHMI$ denotes an indicator variable that switches to one in the year scientist j becomes an HHMI, $f(AGE)$ is a flexible function of the scientist's age, and the $\gamma_{ii'}$

are article-pair fixed effects, consistent with our approach to analyze *changes* in the citation rates to articles in each pair following the appointment of investigator j to HHMI.[9] We also run slight variations of this specification in which the dependent variable has been parsed so that we can break down citation flows by citer status (i.e., citations from members of the National Academy of Sciences vs. non-members, from within vs. outside the field, or from *de novo* vs. *de alio* citers).

There is another benefit to conducting the analysis at the product-pair level: since treated and control products always originate in the same year, experimental and calendar time coincide, making it simple to display the results graphically. The graphical approach is advantageous in that it is simple and it enables to us go beyond a temporal averaging of status effects (i.e., a single point estimate of the treatment effect that averages its impact over time) to illustrate their dynamics. Conversely, a drawback of the graphs is that the standard errors are naïvely computed from the raw data. As a result, the confidence bands ignore the clustered structure of the data, specifically the presence of multiple articles per scientist. The supplementary online material (Appendix IV) presents a number of alternative graphs corresponding to slightly different statistical assumptions. Our results, however, are extremely robust to these modeling choices.

4 Results

Main Effect of HHMI Appointment on Citation Rates to Articles Published Post-Appointment. Before we present our analysis of the treatment effect of appointment to HHMI using the article-pair matched sample, we report an estimate of the effect of becoming an HHMI that approximately follows the "standard" methodology in the literature. This analysis differs from our subsequently reported findings in two, essential respects. First, to replicate the standard approach, we need to impose a match at the producer level, rather

[9]For two reasons, article-pair fixed effects may alter the estimations despite our use of the matching procedure. First, recall that the matching procedure is coarse, not exact. Second, the fixed effect transformation subtracts the mean of the outcome variable within the sample *over the entire sample period*, from the contemporaneous level. But there is no presumption that this mean should be zero, even if our matching procedure was more "exact" than it is.

than at the product level. To do this, we will control for producer-level but not product-level quality. Second, the articles in these data were written after the treated scientist won the prize, rather than before. In other words, the first set of results are akin to an estimate of the effect of the shock to producer status that, (i) accounts for producer-level quality but does not include careful controls for product quality, and (ii) depends on the output the producer creates after the status shock.

To implement the standard approach, we pair an HHMI winner with an ECPW scientist who is very similar on covariates that we know matter for selection into the HHMI program. These covariates are (i) year of highest degree; (ii) gender; and, most importantly, (iii) the number of independent "hits" recorded up to the appointment year. By limiting the control group to early career prize winners and then further matching on the number of hits, we effectively incorporate into the analysis excellent controls for scientist-level quality. In addition, we match on a few basic article characteristics, including the length of the authorship roster, the focal author's position on it, and publication year.

The results of this analysis, which estimates the effect of appointment to HHMI on scientists' future performance while controlling for scientist-level quality, are presented in Panel A of Figure 1. In this and subsequent figures, we display the difference in average citation trends for the article pairs in the sample (the solid line), along with a 95^{th} confidence interval (the dashed lines). Panel A shows that articles written by HHMIs are cited more frequently than articles written by EPCWs. The citation premium exists in the first year of an article's life, increases in the article's second year, and gradually declines over the next 10 years without ever vanishing. By 2007 (the conclusion of our observation period), the conventional approach to estimating the effect of HHMI on article-level citations suggests that HHMI-authored articles garner 10 more citations than those of early career prize winners. This boost corresponds to 12% of the cumulative number of citations that the average control articles will receive in the observation window.

In this analysis, however, interpreting appointment to HHMI as causing a change in the perceived quality of prize winners' work is problematic due to at least two alternative

possibilities. The first is that the premium reflects the presence of correlated resource endowments. For example, relative to ECPW scientists, HHMI appointees receive large research budgets. Second, the actual quality of their post-appointment publications might increase relative to a control group of non-prize-winners because HHMI's may benefit from access to higher quality graduate students, better-equipped laboratories, advice from an elite peer group, and so forth. In other words, resource flows tied to the award may result in actual improvements to the quality of prize winners' articles, rather than simply changes in others' perception of their quality.

In fact, we present evidence that this is indeed the case. Specifically, in Panel B we repeat the analysis presented in Panel A, but this time we incorporate a single, additional criterion in the matching algorithm: instead of matching on just a scientist-level measure of quality (the number of hits), we also require that the treated and the control papers in each article pair were published in the same journal. In other words, we incorporate a product-level quality control to the matching algorithm, so that we better account for potential quality differences between the articles written by treated and control group members.

The control for product-level quality brings us much nearer to a true test of the Matthew Effect. Merton (1968) argued that the Matthew Effect occurs when work of a fixed quality is accorded greater recognition when it is the product of a higher status producer. To empirically assess the Matthew Effect at the product level, it is therefore necessary to hold product quality constant. When we take a first pass at this by matching articles on scientific journal, a citation premium for HHMI-authored papers is still evident, but its magnitude is reduced (to about 0.6 citation per year). Immediately, we see the potential bias in estimating the effect of status while controlling only for producer-level quality, without accounting for quality differences at the product level. The inclusion of a product-level control erases more than one third of the estimated status effect.

There is, however, a second complication that raises further questions about the interpretation of the residual treatment effect: it remains possible that the citation increase could be an artifact of the selection process. Even if ECPW and HHMI scientists were perfectly

matched on past achievement, the HHMI appointment event may also incorporate information about scientists' future *potential*. If this were the case, one would expect to observe a citation premium for articles by HHMI investigators, even in the absence of any status effect. In this interpretation, the award does not cause changes in perceived article quality; it simply reflects actual differences in quality. By focusing on changes in citation rates following HHMI appointment for articles published before appointment while matching on the pre-appointment citation trajectory, we believe that our research design enables us to isolate the operation of status-based changes in perception from these competing effects.

Main Effect of HHMI Appointment on Citation Rates to Articles Published Pre-Appointment. The comparisons in Figure 1 contrast articles written after a scientist is appointed to HHMI with a matched pair written by an ECPW. Figures 2 and 3—the core of our graphical analysis—confine the analysis to articles written before the HHMI is appointed, which are each paired with a matching ECPW article. These figures report the difference in citation trajectories for the 10-year period following the treated scientists' appointment to HHMI. The zero point is the year the HHMI is appointed; negative years indicate the pre-appointment period and positive years correspond to the post-appointment period. Because we now limit the analysis to articles that were written before appointment, we can incorporate an additional, stringent control for article-level quality in the matching procedure; these graphs restrict the dataset to pairs of HHMI and ECPW articles that were: (i) published in the same year; (ii) in the same journal; (iii) with approximately the same number of authors; (iv) in which the HHMI and ECPW scientists occupy the same authorship position; and (v) that were following a nearly identical citation trajectory up to and including the year that precedes HHMI appointment. We then investigate whether there is a citation boost associated with HHMI appointment.

In Figure 2, the average citation premium hugs the horizontal axis of the graph until the year the HHMI is appointed. The similarity in the years between when a pair of articles was published and when one scientist in the pair becomes an HHMI confirms that the matching algorithm successfully selects control articles with citation trends that were nearly identical

to treated articles. The magnitude of the status effect for appointment to HHMI in the overall sample is captured as the difference in the curves in the post-appointment (after time zero) period. Inspection of the figure reveals that the effect is not large: there is a slight uptick in the citation rate in the first post-appointment year, and then a gradual decrease in subsequent years. Based on this evidence, the effect of the status shock on the perceived quality of scientists' existing work is small.

While this is an overall conclusion, it is altered by cutting the data into different subsets to examine contingencies in the effect of the status shock. First, we find that the results depend on the vintage of the papers being examined. Figure 3, Panel A performs an identical analysis, but limits the sample to articles published at least three years before appointment. For this sample, there is no evidence of an HHMI citation boost. Figure 3, Panel B limits the sample to articles published exactly two years before the year of HHMI appointment. Once again, there is no hint in this subsample of an HHMI citation premium. Finally, Panel C focuses on articles published in the year immediately preceding the year of appointment.[10] In this sample, there is a post-appointment citation increase. HHMI articles receive approximately three citations more on average than their ECPW counterparts in the year following appointment. While the quantity of the citation premium slowly attenuates after the first year, the effect remains throughout our data. On average, HHMI articles receive 24 extra citations over the ten-year period that follows appointment. To contextualize this effect, the average number of cumulative citations in the control sample is 68, and 68+24=92 citations map into a change from the 72^{nd} to the 81^{st} percentile of the distribution. This nine-percentile point rightward shift strikes us as being relatively small, though still meaningful.

We verify this conclusion in a regression that incorporates article-pair fixed effects, corresponding to the estimating equation above. Specifically, we regress the difference in the level of citations within a treated/control article pair onto year effects, career age indicator variables for the focal HHMI investigator, and interaction terms between the treatment effect and the vintage of each article at the time its author was appointed. Since related articles

[10]For the prior year's articles, each of the papers in a pair are matched on publication month. This is necessary because two articles with an equal number of citations in, say 1990, the first appearing in February, and the second appearing in October, might be on different citation trends.

in the sample are published between one and ten years before their associated appointment event, there are ten interaction terms. Figure 4 reports these interaction effect, with 95% confidence intervals denoted by vertical bars.

Although there is evidence of a treatment effect for articles published in past years, most notably five years prior to the time of HHMI appointment, the regression results broadly are consistent with the graphics in Figure 3: the largest effect of HHMI appointment occurs for articles published in the year prior to the status shock. Therefore, we conclude that the status jump associated with appointment to HHMI does influence how past work is perceived, but its effect is limited in time. Its impact is greatest on recent product introductions.

Variation in the HHMI Citation Premium. The foregoing results show the *average* citation premium that accrues to pre-appointment articles after individuals become HHMIs. In Table 2, we now examine whether the magnitude of the premium correlates with specific attributes of articles or scientists. Because the status effect is most discernible for articles that were written one year prior to the time of the treatment effect, we focus on this set of articles in the regression analysis. Robust standard errors, clustered at the level of the focal HHMI scientist, appear below the coefficient estimates in brackets.

We now return to status theory and proceed with the analyses in two steps. First, we examine whether the effect of status is amplified when there is greater uncertainty surrounding product-level quality; second, we investigate whether there is evidence of a ceiling-like effect, in which the returns to a status shock depend on the producer's pre-award level of status.[11]

At the product level, we introduce three contingencies to examine the impact of uncertainty in product quality on the magnitude of the HHMI treatment effect. First, we create an indicator for whether the journal in which an article pair is published exceeds the median Journal Impact Factor (JIF)—a measure of the frequency with which the average article in a journal has been cited in a particular year (column 2). We interact the JIF indicator with

[11] Although it is possible to present the results of all of these interaction effects in graphical form too, to economize on space, we only report regression results. A complete set of graphs is available from the authors upon request.

the treatment effect following the logic that journal quality signals article quality, and that the variance in article quality may be greater in lower impact journals. Therefore, the effect of author status on perceptions of article quality should be greater in low JIF journals. The evidence in Table 3, Column 2 is consistent with this reasoning: the HHMI citation premium is approximately twice as large for articles appearing in less prestigious journals.

Second, we explore whether the effect of status is larger when market participants face difficulty in evaluating the inherent worth of particular products because these products embody novel concepts or categories (Podolny, Stuart, and Hannan 1996; Zuckerman 1999). To measure novelty at the article level, we first calculate the vintage of all the MeSH keywords tagging the articles in our sample.[12] Concretely, we define the birth year of a keyword as the year in which it first tags a paper indexed by PubMed. For each article, we then difference the average MeSH vintage from the publication year to produce our index of novelty. We assume that more novel articles are ones that are tagged by MeSH keywords that were first used in the recent past.

Next, we create an indicator variable based on a median split of the measure of novelty. In Table 2, Column 3, we interact article-level novelty with the treatment effect. Once again, the finding confirms status theory: the post-HHMI citation premium is larger when the focal article in the pair is relatively more novel. In fact, the absence of a statistically significant main effect of HHMI appointment indicates that the citation benefit of the award is concentrated on relatively novel articles.

As a third proxy for product-level uncertainty, we construct a measure of recombinant scope, or the extent to which a focal article builds upon antecedents from areas of science that are unrelated to it. We posit that papers that cite a higher proportion of out-of-field prior work are of greater uncertainty. These papers typify what Fleming (2001) labels, "recombinant uncertainty." To quantify recombinant uncertainty, we use PubMed's "related articles" algorithm, which combines MeSH keywords, abstract words, and title words to

[12]MeSH is the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. There are 24,767 descriptors in the 2008 MeSH (new terms are added to the dictionary as scientific advances are made).

identify the set of most-related papers for all referenced articles. In effect, by identifying the nearest neighbors for each article, the algorithm allows one to delineate a scientific "field" around each source publication, where field denotes a specific niche in the space of ideas. We use this algorithm to identify related versus unrelated trailing citations. To construct the measure, we start with the set of all articles cited by the focal article. We then match these article-level sets of all cited papers to the set that PubMed identifies as being related to the focal article. From this, we calculate the proportion of a focal paper's citations that go to its scientific near neighbors. Finally, we conduct a median split of the covariate, and run an interaction between it and appointment to HHMI.

Table 2, Column 4 reports the result. Once again, we find that the effect of the status shock is much greater for articles that are high in recombinant scope. In fact, the results show that the HHMI treatment effect for articles that build on intellectually proximate antecedents does not statistically differ from zero. Therefore, for all three measures of product-level uncertainty, we find that the effect of the producer-level status shock is larger for products that are of less determinate quality.

Before concluding the discussion of interaction effects between the extent of product-level uncertainty and producer status, we return to one of the key patterns in the data: the effect of HHMI appointment on subsequent citations is strongest for articles that were published during the year preceding appointment. There are a few interpretations of the finding that recent articles exhibit the largest post-treatment citation boost. On one hand, it may correspond to a true, dynamic feature of status effects: the designation of the prize leads to reassessments of quality, but only for recent work. On the other hand, this finding too may be attributable to the sensitivity of the effect of the status signal conveyed by the HHMI award to the presence of uncertainty. Specifically, there is likely to be more uncertainty about the quality of recently published articles than of those that have been on the market for a longer interval of time. In other words, the temporal patterns in the data support the conclusion that producer status may have a larger effect on perceptions of quality for new products.

Journal Impact Factor, MeSH keyword novelty, recombinant scope, and article age are product-level measures of uncertainty. In addition, our data also contain some variation in the level of pre-Award producer-level status. Although all of the scientists in our data are distinguished, there are shades of gray. This is because there are differences in the scientific track records of HHMI awardees at the time of their appointment. Some are appointed only a few years after starting their independent careers, while others are much more senior in their fields.

This brings us to the third question we posed: does the effect of the status shock depend on where the producer is positioned in the status hierarchy at the time of the award? Once again, the literature offers an expectation: because of the possible presence of a ceiling effect, there may be diminishing returns in the benefits of a jump in status for producers who already are near the peak of the status hierarchy (Bothner, Smith and White 2010; Bothner, Podolny and Smith 2011). This is, in fact, exactly what the data show. In Table 2, column 5, we interact the treatment effect with the scientific eminence of the HHMI at the time of the award, as measured by the cumulative stock of citations to all articles he or she had published up to the year before appointment. We find that the articles published by (relatively) less eminent scientists benefit more from the status shock. Likewise, we create an indicator based on the median age of HHMI awardees. Column 6 in Table 2 interacts this variable with the treatment effect. As anticipated, the citation premium is larger in magnitude for younger investigators. Therefore, the evidence is consistent with the view that the effect of a status shock of this nature declines in the pre-accolade status of the awardee.

5 Robustness Checks

We present three robustness checks to further assess the integrity of the results.

Salience of HHMI status marker. First, we conduct a form of a falsification test by examining whether the citation premium accruing to HHMI awardees varies with authorship

credit for the scientists in the data. If HHMI appointment represents a genuine status shock and if future citations to past articles is a good measure of perceptions of quality, the strength of the results should depend on which position the HHMI holds on the authorship roster of an article. In particular, a strong norm in the life sciences assigns last authorship to the principal investigator, and therefore the consequence of the award should be most significant for articles in which the focal HHMI occupies the last author's position. To examine this, we created an indicator variable for all article-pairs in which the HHMI awardee is the last author, which we then interact with the treatment effect. In Table 2, column 7, the status effect appears to be twice the size for article pairs in which the HHMI scientist is the last author, although the interaction effect only is significant at $p < 0.10$.

Appointment panel effects. We observed a post-citation boost only for articles published in the year immediately preceding appointment. If these recent articles are precisely those that convinced the HHMI selection panel to choose these particular nominees for the Award, then we run the risk that the results could be influenced by the citation patterns of the panelists themselves, as they are active publishers who may become aware of applicants' work in the selection process. If this were the case, it may be stretching the meaning of the Matthew Effect to interpret our results through its lens.

Although we cannot identity the panelists by name, we do know they are recruited from the ranks of the National Academy of Science and long-established HHMI investigators. In Table 3, columns 1a and 1b, we split the pool of citing articles into two separate buckets. The first includes articles in which the authorship roster does not list any member of the NAS or a past HHMI investigator. Conversely, the second is limited to articles in which at least one author is a member of the academic "super-elite." The results show that existing HHMI investigators and NAS members do not cite papers of soon-to-be-appointed HHMI investigators more than would be expected given their relatively tiny share of the overall pool of citing articles.

Epiphenomenal? Finally, we extend this analysis to a broader exploration of the composition of post-HHMI citations. The objective of these analyses is to determine whether the sta-

tus shock appears to cause an epiphenomenal bounce in citations, in that it merely leads to ceremonial citations. This cannot be resolved definitively, but we provide suggestive evidence by analyzing two outcomes: the before-versus-after incidence of within- vs. across-field citations, and the before-versus-after incidence of *de novo* vs. *de alio* citations.

To determine whether the award leads to across-field or within-field citations, we again exploit PubMed's "related articles" algorithm, described above. For each HHMI- or ECPW-authored article, we collect all forward citations to it and match them with the set of related articles harvested from PubMed. Using this match, we can then parse the citation counts to create two separate buckets, the first with "within field" or related citations; the second with "outside-of-field" citations, i.e., articles that cite the source article but that PubMed does not recognize as being related to the source in intellectual space. We then investigate whether the HHMI treatment influences within-field citations, outside-of-field citations, or both. As can be seen in Table 3, columns 2a and 2b, the answer is "both." When we estimate (not-reported) regressions in logs, we find that the elasticities of out-of-field and within-field citations are, respectively, 0.10 and 0.20, ($p < .10$). Thus, there is a marginally significant shift toward within-field citers, post-treatment. If anything, the award drives a greater depth of related work.

Next, we investigate whether post-award citers are *de novo*—new to the article—or *de alio*—repeat citers. One hurdle in making this determination is that the treatment effect is greatest for publications that only slightly predate the HHMI appointment. For these articles, there is a limited pre-award citation history, which means that there is little basis to distinguish between new and old citers for the articles that experience the strongest treatment effect. We can, though, implement a variant of this idea; we can measure whether the citer has cited a different paper from the same HHMI (treated) or ECPW (untreated) scientist in the past five years. Using these measures of *de novo* vs. *de alio* citations, we conclude (Table 3, columns 3a and 3b) that there is no consequence of the award for these proclivities. When we estimate (not-reported) regressions in logs, we find that the elasticities of *de novo* and *de alio* citations are, respectively, 0.18 and 0.09 (difference not statistically significant

at conventional levels). The elasticities suggest a modest shift toward de novo citers, but the coefficients are not precisely estimated.

Based on these two sets of analyses (and those on elite vs. non-elite citers), we come to a tentative conclusion that the consequence of the HHMI for subsequent citing behavior is not epiphenomenal. To the best of our ability to gauge, the composition of post-award citers is relatively comparable to that in the pre-treatment interval. The prize may slightly broaden the recognition of an HHMI's work to new scholars, but it does so to scholars who are within field. Therefore, we do not believe that the status shock merely causes a jump in ceremonial citations; the findings suggest that it legitimately deepens the collective acknowledgment of Award winners' previous scientific contributions.

6 Conclusion

This paper presents a novel research design that enables a focused test of the Mertonian hypothesis that a producer's status is a lens through which audience members assess the quality of work. Specifically, the research design first zeroes in on the effect of a change in status caused by winning a major prize on other-party perceptions of the quality of a focal producer's goods. To identify the effect of the change in status that is independent of its potential influence on the actual quality of a producer's outputs, we limit the analysis to the effect of the change to evaluations of outputs that were produced prior to the time the prize was granted. To further insure that the results truly reflect changes in perceptions (versus forecasts that endogenously relate to the selection of specific producers as prize winners), we implement a product-based, matched sample design that pairs each treated product to a nearly identical twin that is closely matched on product quality.

Our findings suggest that the standard approach to estimating the effect of status on performance is likely to overstate its true, causal influence. This likely occurs for two reasons. First, controls for quality often are inadequate, particularly if quality is held constant at the producer level but performance is measured at the product level. Second, changes in a focal actor's status follow acts of deference from high status actors, whether through the awarding

of prizes or other forms of recognition, or through creation of an affiliation that conveys an endorsement. While these actions on the part of already high status actors do cause changes in alters' prestige, the intentionality of these status-conferring behaviors often is rooted in forecasted changes in performance. If high status actors (such as the members of the HHMI selection committee) bestow recognitions or affiliations because they anticipate that the recipients of these acts of deference are on positive performance trajectories, status changes may reflect—rather than cause—changes in performance.

For both reasons, much of the existing empirical literature on status may overestimate its true effect. It is for this reason that we introduced the paper with a bold title. We do not claim that the evidence presented here is definitive, and we acknowledge that we do not address the principle avenue through which the accumulative advantage process may unfold (the preferential access to resources enjoyed by producers who benefit from a status shock). However, we strongly stand by the point that much of the existing evidence for a large Matthew Effect in market contexts also is not definitive. Moreover, our findings and those of Simcoe and Waguespack (2011) suggest that, at least in the contexts we study, the direction of the bias in the existing literature is to over-estimate the true magnitude of status effects. Therefore, our hope is to re-open the debate about the magnitude and mechanisms of the Matthew Effect.

Despite the result that the conventional estimation approach overstates the true effect of status, we still find that appointment to HHMI causes an increase in citations to articles written before the award was granted. In a strict test of the Mertonian hypothesis that prestigious scientists garner greater recognition for outputs of a given level of quality, we find modest support for a main effect of a change in status. However, consistent with the subsequent literature and with Merton's own argument, we show that the effect of status is much larger when there is significant uncertainty surrounding product-level quality. We also find that there are diminishing returns at the apex of the status hierarchy—the benefits of the Award are smaller for already eminent scientists. Extending this finding beyond the information in our data, it is also possible that the relatively modest main effect of the status shock of HHMI appointment occurs because of the range in the status distribution on which

we focus the analysis. Specifically, HHMI prize winners are prominent even before they win the award, and the control group of early career prize winners is almost if not equally noteworthy. Appointment to HHMI undoubtedly elevates recipients' status, but the effect of a given change in status may be much greater if it is experienced by actors who begin on a lower rung of the prestige hierarchy. In other words, there may be non-linearities such that the causal effect of a change in status may greatly depend on a producer's origin point in the status hierarchy. To our knowledge, there is no empirical evidence of differential returns across the status distribution. This is a question that clearly merits further theoretical elaboration and empirical investigation.

In concluding, a few remarks about the methodology and the scope of this analysis are in order. First, natural experiments like the one in Simcoe and Waguespack (2011) are ideal for identifying causal effects in research on social status, but they are difficult to discover. Matching methods implemented at the product level may be feasible in a wider range of contexts. The clear advantage of our approach is broader applicability, but it does carry a few costs. First, relative to a natural experiment, implementing a matching procedure always requires significant judgment; the researcher must select the dimension on which the match is constructed. Second, when performing inference, we have ignored the influence of the prior matching step and focused only on the second-stage standard errors. At the moment this is the only alternative because the full statistical properties of these estimators have yet to be clarified.

Finally, a clear scope condition of our work concerns the primary mechanism through which a status jump translates into superior performance. By limiting the empirical analysis to the effect of the prize on the citation trajectories of previously published articles, we have attempted to precisely estimate the effect of a shock to an actor's status on changes in perceptions of the quality of that actor's products. The narrowness of the empirical test in the paper is both its core strength and weakness. On one hand, we believe that it is one of the cleanest tests yet of Merton's famous hypothesis. Moreover, we believe the research design is very much aligned with the spirit of Merton's (1968) thought experiment in which he compares the careers of the 40^{th} to the so-called 41^{st} chair, when the former is actually

elected to the French Academy and the latter barely misses the cut. On the other hand, the consequence of this tight comparison is that we neglect other pathways through which changes in status influence performance outcomes.

It may be, for instance, that through the implicit anointment into the academic super-elite that co-occurs with appointment to HHMI investigatorship, prize-winners gain preferential access to the most prominent journals in their fields. Or, they may benefit from privileged access to very tangible forms of resources, such as state-of-the-art laboratory equipment. Insofar as these forms of resource access can be causally related to changes in status, our analysis may significantly understate the full consequence of gains in status, even if it correctly spotlights its effect through changes in other-party perceptions of a focal actor's outputs. Our goal in this paper was to present a narrow test of this specific mechanism. In future work, similar research designs can be developed to illuminate the other routes through which status affects attainment.

References

Allison, Paul D., J. Scott Long, and Tad K. Krauze. 1982. "Cumulative Advantage and Inequality in Science." *American Sociological Review* **47**(5): 615-625.

Azoulay, Pierre, Andrew Stellman, and Joshua Graff Zivin. 2006. "PublicationHarvester: An Open-source Software Tool for Science Policy Research." *Research Policy* **35**(7): 970-974.

Azoulay, Pierre, Joshua Graff Zivin, and Gustavo Manso. 2011. "Incentives and Creativity: Evidence from the Academic Life Sciences." *RAND Journal of Economics* **42**(3): 527-554.

Bales, Robert F., Fred L. Strodtbeck, Theodore M. Mills, and Mary E. Roseborough. 1951. "Channels of Communication in Small Groups." *American Sociological Review* **16**(4): 461-488.

Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. "cem: Coarsened Exact Matching in Stata." *The Stata Journal* **9**(4): 524-546.

Blau, Peter M. 1955. *The Dynamics of Bureaucracy.* Chicago, IL: University of Chicago Press.

Bothner, Matthew S., Frederic Godart, and Wonjae Lee. 2009. "What is Social Status? Comparisons and Contrasts with Cognate Concepts." Working Paper, University of Chicago.

Bothner, Matthew S., Joel M. Podolny, and Edward B. Smith. 2011. Organizing Contests for Status: The Matthew Effect vs. the Mark Effect. *Management Science* **57**(3): 439-457.

Bothner, Matthew S., Edward B. Smith, and Harrison C. White. 2010. A Model of Robust Positions in Social Networks. *American Journal of Sociology.* **116**(3): 943-992

Cole, Stephen, and Jonathan R. Cole. 1968. "Visibility and the Structural Bases of Awareness of Scientific Research." *American Sociological Review* **33**(3): 397-413.

Gould, Roger V. 2002. "The Origins of Status Hierarchies: A Formal Theory and Empirical Test." *American Journal of Sociology* **107**(5): 1143-1178.

Hsu, David H. 2004. What Do Entrepreneurs Pay for Venture Capital Affiliations? *Journal of Finance* 59(4): 1805-1844.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2011. "Multivariate Matching Methods that Are Monotonic Imbalance Bounding." *Journal of the American Statistical Association* **106**(493): 345-361.

Lynn, Freda B., Joel M. Podolny, and Lin Tao. 2009. "A Sociological (De)Construction of the Relationship between Status and Quality." *American Journal of Sociology* **115**(3): 755-804.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* **159**(3810): 56-63.

Phillips, Damon J., and Ezra W. Zuckerman. 2001. "Middle Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets." *American Journal of Sociology* **107**(2): 379-429.

Podolny, Joel M. 1993. "A Status-based Model of Market Competition." *American Journal of Sociology* **98**(4): 829-872.

Podolny, Joel M. 2005. *Status Signals: A Sociological Study of Market Competition.* Princeton, NJ: Princeton University Press.

Podolny, Joel M., and Toby E. Stuart. 1995. "A Role-Based Ecology of Technological Change." *American Journal of Sociology* **100**(5): 1224-1260.

Podolny, Joel M., Toby E. Stuart, and Michael T. Hannan. 1996. "Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991." *American Journal of Sociology* **102**(3): 659-689.

Roberts, Peter W., Mukti Khaire, Christopher I. Rider. 2011. Isolating the Symbolic Effect of Employee Mobility: Price Increases after Hiring Winemakers from Prominent Wineries. *American Economic Review: Papers and Proceedings* **101**(3): 147-151.

Roberts, Peter W., Adina D. Sterling. 2012. Network Progeny? Prefounding Social Ties and the Success of New Entrants. *Management Science* **58**(7): 1292-1304.

Simcoe, Timothy S., and David Waguespack. 2010. "Status, Quality, and Attention: What's in a (Missing) Name?" *Management Science* **57**(2): 274-590.

Stuart, Toby E., Ha Hoang, and Ralph C. Hybels. 1999. "Interorganizational Endorsements and the Performance of Entrepreneurial Ventures." *Administrative Science Quarterly* **44**(2): 315-349.

Waguespack, David, and Olav Sorenson. 2011. The Ratings Game: Asymmetry in Classification. *Organization Science* **22**(3): 541-553.

Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* **104**(5): 1398-1438.

Table 1: Descriptive Statistics for Pre-Appointment Articles (n=2×3,636)

	ECPWs			HHMIs			Overall	
	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median	Min.	Max.
Number of Authors	3.981	2.099	4	3.982	2.084	4	1	35
Focal Author is Last	0.634	0.482	1	0.634	0.482	1	0	1
Journal Impact Factor	8.686	5.393	7	8.686	5.393	7	0	30
Baseline Stock of Citations	20.401	25.956	11	21.207	26.350	12	0	273
Baseline Stock of PubMed-unrelated Citations	18.117	24.052	9	18.290	23.889	10	0	257
Baseline Stock of PubMed-related Citations	2.284	4.413	0	2.919	4.820	1	0	53
Baseline Stock of *de novo* Citations	11.986	16.479	6	13.785	18.152	8	0	211
Baseline Stock of *de alio* Citations	8.141	12.100	4	7.168	10.599	3	0	119
Baseline Stock of Citations from Non-Elite Citers	19.160	24.353	11	19.435	23.999	11	0	250
Baseline Stock of Citations from Elite Citers	1.241	2.603	0	1.774	3.446	0	0	42
Publication Year	1988.06	5.735	1988	1988.076	5.735	1988	1974	2002
Appointment Year	1992.66	5.017	1993	1992.616	5.017	1993	1984	2003
Investigator Graduation Year	1970.77	7.938	1971	1976.661	7.677	1977	1956	1998
Investigator Gender	0.111	0.314	0	0.133	0.340	0	0	1
Investigator Career Cites at Baseline	5,756	5,911	4,042	6,311	5,880	4,855	0	77,787
Investigator Stock of Top 5% Pubs. at Baseline	14.087	15.417	9	15.150	11.097	12	0	187
MeSH Keywords Average Vintage	24.978	8.877	25	24.418	9.002	24	1	60
Proportion of PubMed-unrelated Bckwrd. Citations	0.873	0.163	1	0.821	0.164	1	0	1
Stock of Citations up to Year 10	56.100	72.675	36	60.798	77.214	39	0	1,551

Note: The match is "article-centric," i.e., the control article is always chosen from the same journal in the same publication year. The control article is coarsely matched on the number of authors (exact match for one, two, and three authors; four or five authors; between six and nine authors; and more than nine authors). We also match on focal scientist position in the authorship roster (first author; last author; middle author). For articles published one year before appointment, we also match on the month of publication. For articles published two years before appointment, we also match on the quarter of publication. In addition, control and treatment articles are matched on citation dynamics up to the year before the (possibly counterfactual) appointment year. The cost of a very close, non-parametric match on article characteristics is that author characteristics do not match closely. Imposing a close match on focal scientist age, gender, and overall eminence at baseline results in a match rate which is unacceptably low. A possible compromise is to not match on journal, but to match on author characteristics. This alternative does not change our overall message.

Table 2: Variation in the HHMI Post-Appointment Citation Boost

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After Appointment	1.921*	2.582**	0.785	0.915	2.903**	1.638**	1.219*
	(0.828)	(0.466)	(0.589)	(0.903)	(0.918)	(0.392)	(0.555)
After Appointment × Article in High-JIF Journal		-1.319*					
		(0.540)					
After Appointment × Novel Article			1.526*				
			(0.606)				
After Appointment × Recombinant Scope.				2.544*			
				(1.150)			
After Appointment × HHMI well-cited at Appnt.					-2.598*		
					(1.168)		
After Appointment × HHMI young at Appnt.						3.604**	
						(1.254)	
After Appointment × HHMI is PI							1.011†
							(0.582)
Nb. of Observations	6,292	6,292	6,292	6,292	6,292	6,292	6,292
Nb. of Article Pairs	549	549	549	549	549	549	549
Nb. of Scientists	247	247	247	247	247	247	247
Adjusted R^2	0.720	0.720	0.720	0.721	0.720	0.720	0.720

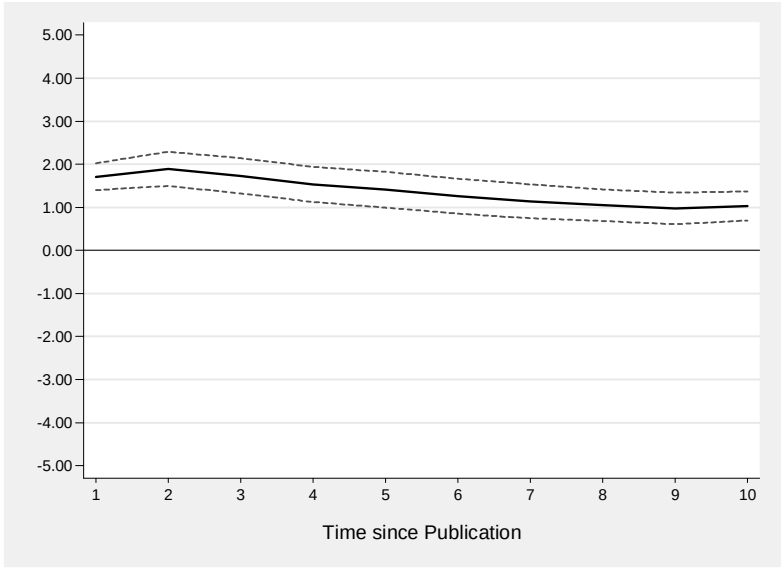
Table 3: Effects of HHMI Appointment on Citation Rates

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
	Citer Status		**Intellectual Overlap**		**Citer Familiarity**	
	Non-Elite	Elite	Unrelated	Related	*de alio*	*de novo*
After Appointment	1.693*	0.228*	1.396†	0.525**	0.549	1.363**
	(0.783)	(0.092)	(0.805)	(0.088)	(0.569)	(0.396)
Nb. of Observations	6,292	6,292	6,292	6,292	6,292	6,292
Nb. of Article Pairs	549	549	549	549	549	549
Nb. of Scientists	247	247	247	247	247	247
Adjusted R^2	0.722	0.413	0.723	0.304	0.699	0.660

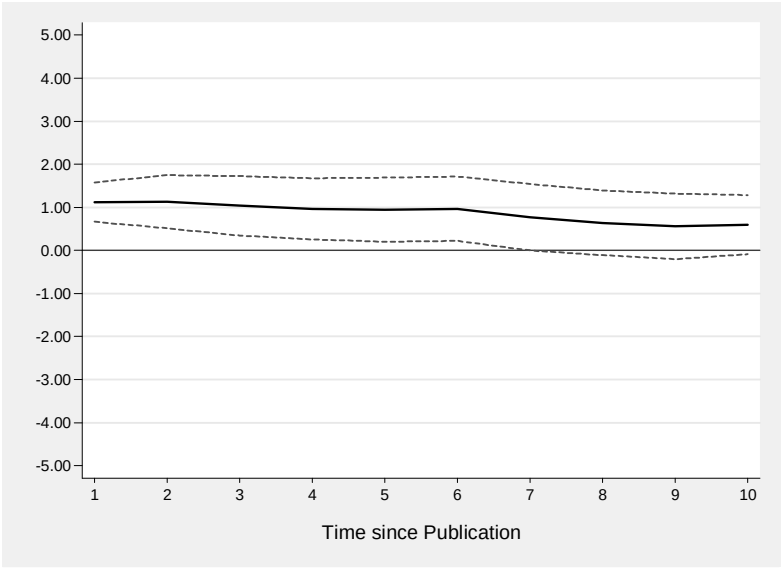
Note: The sample comprises 549 article pairs where both the treated and control articles were published one year before the year of HHMI appointment for the focal investigator. The dependent variable is the difference in citations between the treated and the control article in each pair, in a particular year. The observation window for each pair runs until ten years after appointment or 2007, whichever comes earlier. All specifications are estimated by OLS; the models include a full suite of year indicator variables, twenty five career age indicator variables, as well as article-pair fixed effects. Standard errors in parentheses, clustered by investigator. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure 1
Effect of HHMI Appointment on Citation Rates
Post-Appointment Articles

A. Without Journal Match
[N=12,842 article pairs]

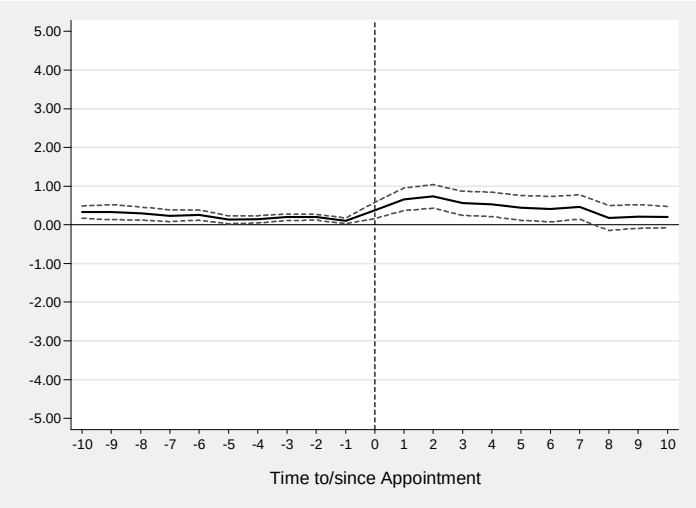


B. With Journal Match
[N=4,631 article pairs]



Notes: Dynamics for the difference in yearly citations between HHMI-ECPW matched articles written in the post-appointment period. The solid dark lines correpond to the sample mean difference in citations in each pair; the dashed grey lined correspond to 95% confidence intervals based on sample standard errors. Articles in each pair are published in the same year, and the focal scientists are matched on degree year, gender, and eminence as indicated by the number of articles they published up to the year of appointment that fall in the top ventile of the vintage-specific article-level distribution of citations. In Panel B, the match is further constrained so that the two articles in each pair appeared in the same scientific journal.

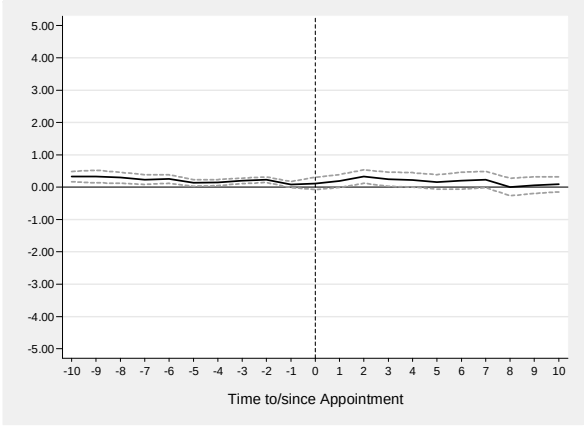
Figure 2
Effect of HHMI Appointment on Citation Rates
Pre-Appointment Articles



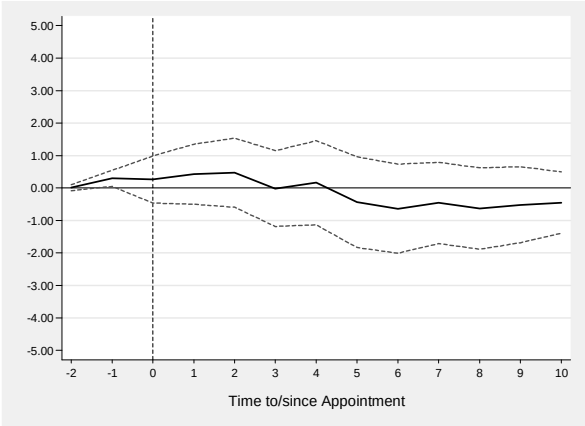
Note: Dynamics for the difference in yearly citations between HHMI-ECPW matched articles. The sample includes articles of vintage t_0-10 to t_0-1 — where t_0 is the year of (possibly counterfactual) appointment. The solid dark lines correpond to the sample mean difference in citations in each pair; the dashed grey lined correspond to 95% confidence intervals based on sample standard errors.

Figure 3
Effect of HHMI Appointment on Citation Rates [Pre-Appointment Articles]

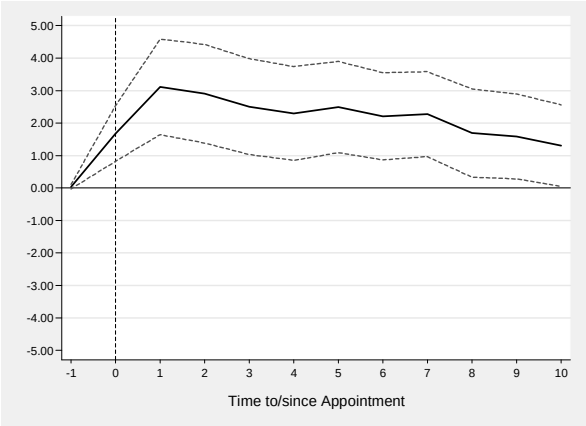
A. Articles Published Between 3 & 10 Years Before Appointment



B. Articles Published Two Years Before Appointment



C. Articles Published One Year Before Appointment

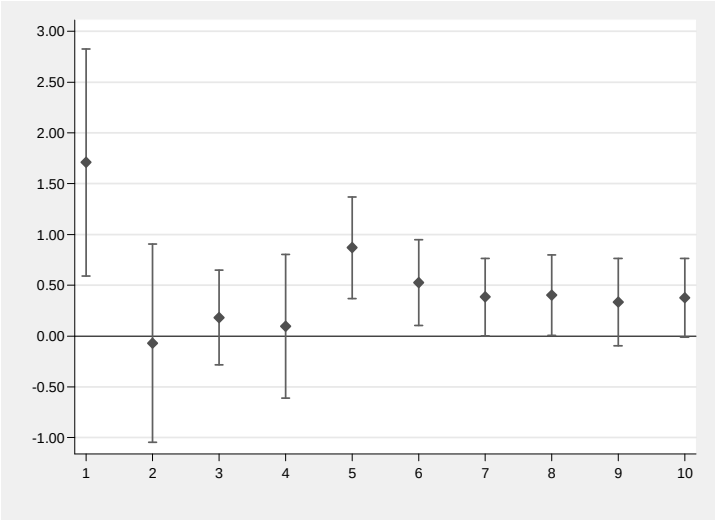


Note: Dynamics for the difference in yearly citations between HHMI-ECPW matched articles. Articles in each pair appeared in the same year and journal, and are also matched on focal scientist position on the authorship list, as well as overall number of authors. Further, control articles are selected such that the sum of squared differences in citations between control and treated article up to year t_0-1 is minimized — where t_0 is the year of (possibly counterfactual) appointment.

In Panel A, the sample is limited to articles published between year t_0-3 and year t_0-10. In Panel B, the sample is limited to articles published in year t_0-2. In addition to being matched on journal, focal scientist position on the authorship list, and overall number of authors, the articles in each pair appeared in the same quarter. In Panel C, the sample is limited to articles published in year t_0-1. In addition to being matched on journal, focal scientist position on the authorship list, and overall number of authors, the articles in each pair appeared in the same month.

The solid dark lines correpond to the sample mean difference in citations in each pair; the dashed grey lined correspond to 95% confidence intervals based on sample standard errors.

Figure 4
Interaction of HHMI Appointment Effect with Article Vintage in the Year of Appointment



Note: The grey diamonds in the above plot correspond to coefficient estimates stemming from fixed effects specifications in which the difference in the level of citations within a treated/control article pair is regressed onto year effects, career age indicator variables for the focal HHMI investigator, as well as interaction terms between the treatment effect and the vintage of each article at the time its author was appointed. Since related articles in the sample are published between one and ten years before their associated appointment event, there are ten such interaction terms. The 95% confidence interval (corresponding to robust standard errors, clustered around case codes) are denoted by the vertical bars.

Appendix I. Linking Scientists with their Journal Articles

The source of our publication data is PubMed, a publicly available bibliographic database maintained by the U.S. National Library of Medicine. PubMed contains over 14 million articles from 4,800 journals published in the United States and more than 70 other countries from 1950 to the present. We have mined these data using PubHarvester, an open-source software tool that automates the process of gathering publication information for individual life scientists (Azoulay, Stellman, and Graff Zivin 2006).

There are two major challenges that must be overcome to accurately link scientists to their publications. The first relates to what one might term "Type I Error," whereby we mistakenly attribute to a scientist a journal article actually authored by a namesake. The second relates to "Type II error," whereby we conservatively exclude from a scientist's publication roster legitimate articles:

- *Namesakes and popular names.* PubMed does not assign unique identifiers to the authors of the publications they index. They identify authors simply by their last name, up to two initials, and an optional suffix. This makes it difficult to unambiguously assign publication output to individual scientists, especially when their last name is relatively common.
- *Inconsistent publication names.* The opposite error—recording too few publications—occurs because scientists often are inconsistent in the choice of names they choose to publish under. By far the most common source of error is the haphazard use of a middle initial. Other errors stem from inconsistent use of suffixes (Jr., Sr., 2nd, etc.), or from multiple patronyms due to changes in spousal status.

To address with these measurement problems, we designed individual search queries that rely on relevant scientific keywords, the names of frequent collaborators, journal names, as well as institutional affiliations. Althogh the process of query design is very time consuming, it is feasible because we have scientists' CVs and biosketches. PubHarvester provides the option to use such custom queries in lieu of a completely generic query. For example, one can examine the publications of Scott A. Waldman, an eminent pharmacologist located in Philadelphia, PA at Thomas Jefferson University. Waldman is a relatively frequent name in the United States (with 208 researchers with an identical patronym in the American Association of Medical Colleges faculty roster); the combination "waldman s" is common to 3 researchers in the same database. A simple search query for `"waldman sa"[au] OR "waldman s"[au]` returns 302 publications at the time of this writing. However, a more refined query, based on Professor Waldman's biosketch returns only 210 publications.[1]

The above example also makes clear how we deal with the issue of inconsistent publication names. PubHarvester gives the end-user the option to choose up to four PubMed-formatted names under which publications can be found for a given researcher. For example, Louis J. Tobian, Jr. publishes under `"tobian l"`, `"tobian l jr"`, and `"tobian lj"`, and all three names need to be provided as inputs to generate a complete publication listing.

We are very confident that such a degree of customization ensures the accuracy of treated and control scientists' bibliomes.

[1] ((((`"waldman sa"[au] NOT (ether OR anesthesia)) OR ("waldman s"[au] AND (murad OR philadelphia[ad] OR west point[ad] OR wong p[au] OR lasseter kc[au] OR colorectal))) AND 1980:2010[dp]`)

Appendix II. Construction of Article Control Group

We detail the procedure implemented to identify the set of control articles from among the set of articles published by early career prize winning (ECPW) scientists (Figure I).

The sample of control articles is constructed such that the following two conditions are met:

1. treated articles exhibit no differential citation trends relative to control articles up to the time of appointment;
2. treated and control articles match on a number of time-invariant article characteristics;

We identify controls based on the following covariates: (1) year of publication; (2) specific journal (e.g. *Cell* or the *New England Journal of Medicine*); (3) number of authors (the distribution is coarsened into six bins: one, two, three, four or five, between six and nine, and ten or more authors); (4) focal-scientist position on the authorship list (first author, middle author, or last author). In the case of articles published in the year immediately preceding HHMI appointment, the list of matching covariates is expanded to also include the month of publication. In the case of articles published two years before appointment, the list of matching covariates is expanded to also include the quarter of publication. To ensure that pre-appointment citation trends are similar between articles written by HHMIs and their twins drawn from EPCW-authored papers, we also match on cumulative number of citations at the time of appointment, coarsened into 7 strata (0 to 10^{th}; 10^{th} to 25^{th}; 25^{th} to 50^{th}; 50^{th} to 75^{th}; 75^{th} to 95^{th}; 95^{th} to $99^{\text{th;}}$ and above the 99^{th} percentile).

We create a large number of strata to cover the entire support of the joint distribution of the covariates mentioned above. Each observation is allocated to a unique stratum. We then drop from the data all observations corresponding to strata in which there is no treated article and all observations corresponding to strata in which there are less than 5 potential controls. We have found that matching on cumulative citations at baseline/time of treatment is not enough to eliminate pre-move citation trends. To ensure that citation dynamics coincide for treated and control observations, we select among potential matches a single article that further minimizes the sum of squared differences in the number of citations between treated and control articles up until the year that precedes the appointment year.

The procedure is coarse because we do not attempt to precisely match on covariate values; rather, we coarsen the support of the joint distribution of the covariates into a finite number of strata, and we match a treated observation if and only if a control observation can be recruited from this stratum. An important advantage of CEM is that the analyst can guarantee the degree of covariate balance *ex ante*, but this comes at a cost: the more fine-grained the partition of the support for the joint distribution (i.e., the higher the number of strata), the larger the number of unmatched treated observations.

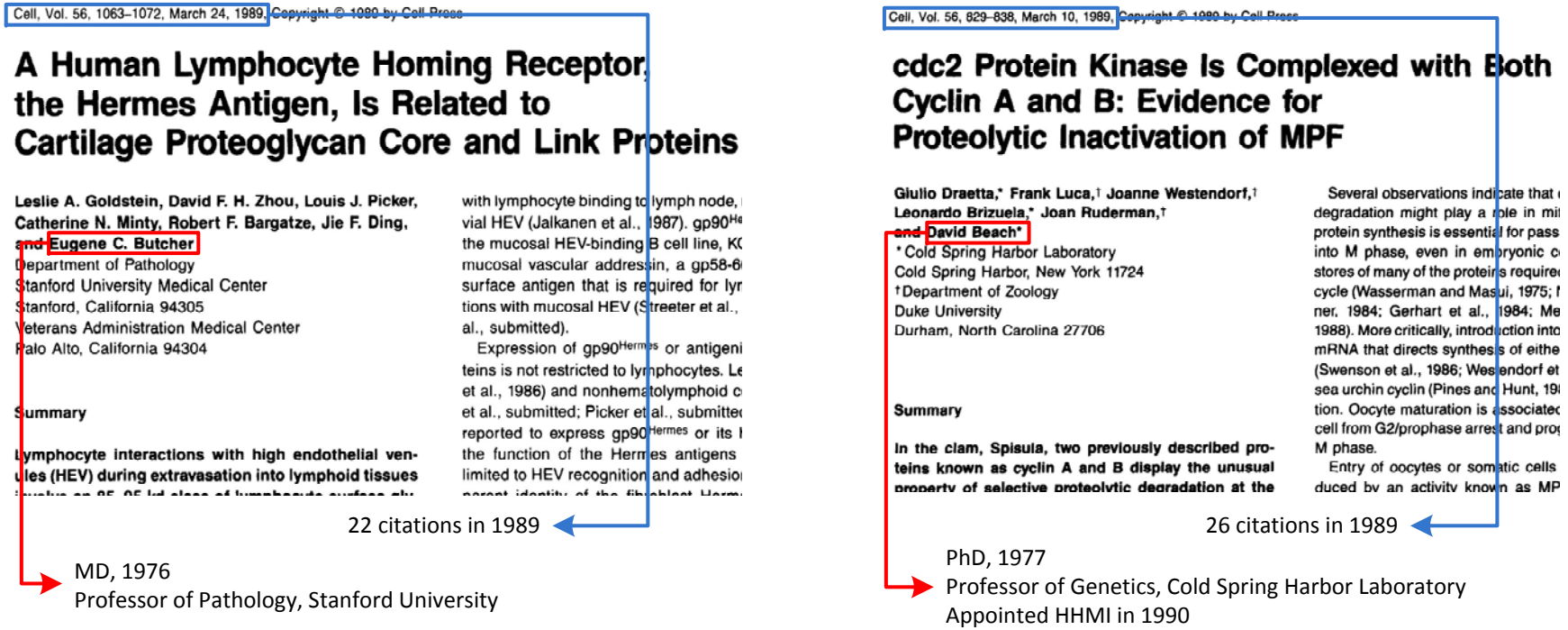
We implement the CEM procedure year-by-year, without replacement. Specifically, in year of appointment t, $1984 \leq t \leq 2003$, we:

1. eliminate from the set of potential controls all articles published by ECPW scientists who have collaborated with HHMI scientists prior to year t;
2. for each year of publication $t-k$, $1 \leq k \leq 10$;
    (a) create the strata;
    (b) identify within strata a control for each treated unit; break ties at random;
    (c) repeat these steps for year of publication $t-(k+1)$.
3. repeat these steps for year of appointment $t+1$.

Based on the descriptive statistics (Table 1 in the main body of the paper), four facts merit attention. First, article-level, time-invariant characteristics are very closely matched between treated and control groups. For some covariates (e.g., number of authors, focal author position, article age), this is a mechanical reflection of the CEM procedure, but for others (such as the article's novelty, as assessed by the average vintage of the keywords that tag the article), the close match occurs incidentally. Second, the

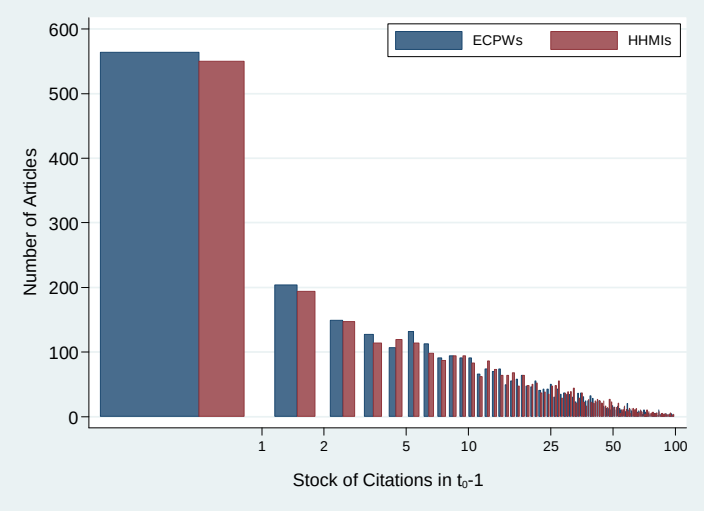
distribution of citations received at baseline is also very similar between the treated and control papers, as can be seen in Figure II. Third, as we would expect when we create a paper-level control group, balance does not extend to scientist characteristics, such as gender and graduation year, though the two groups appear well-balanced on the number of "hit articles" they have previously published at baseline. Fourth, one can at best discern a small difference in the number of citations received cumulatively up until 10 years after appointment – the dependent variable of interest.

Figure I



Note: The two articles above illustrate the essence of the Coarsened Exact Matching procedure. These two articles appeared in the journal *Cell* in March 1989. They received a very similar number of citations in the birth year (1989): 22 citations for Goldstein et al.; 26 citations for Draetta et al. David Beach, the PI on the article on the right-hand side, was appointed a Howard Hughes Medical Investigator in 1990.

Figure II
Cumulative Citations at Baseline
[N=2×3,636 Pre-Appointment Articles]



Note: cumulative number of citations for treatment and control articles, respectively, up to the year that immediately precedes the year of appointment.

Appendix III. PubMed/Web of Science Crosswalk

Our crosswalk procedure utilizes a simple search and scoring technique to best match *PubMed* articles published between 1965 and 2007 with their corresponding records in Thomson Reuters' *Web of Science, Science Citation Index Expanded*.

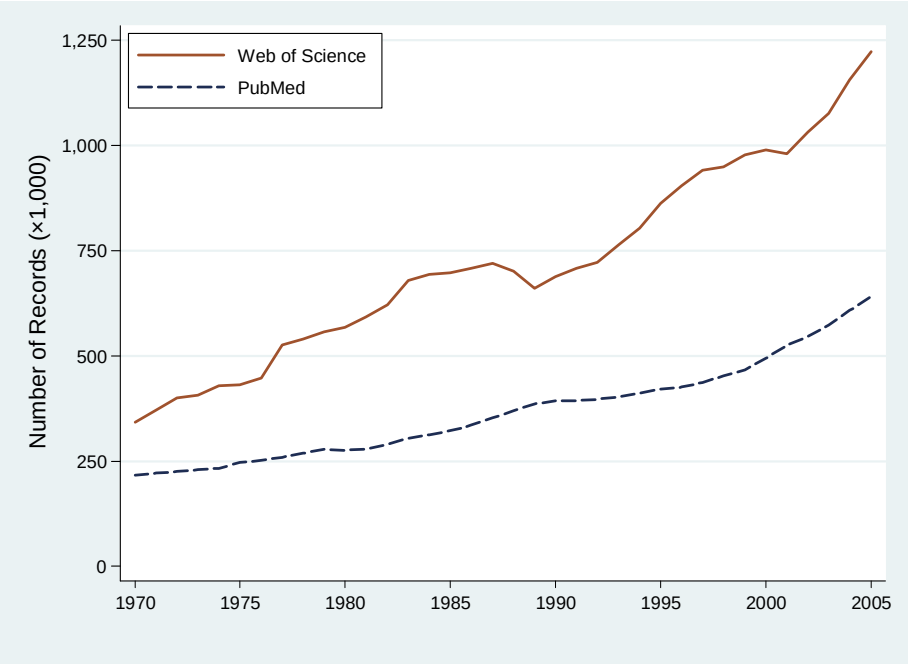
The procedure takes each *PubMed* article and finds a set of potentially matching *Web of Science* records. The program then sorts, scores, and selects the best match from within this group. The search and retrieval of articles used the publication year, title, journal name, journal ISO abbreviation, ISSN, volume, issue number, author names, and beginning and end page numbers from the *PubMed* record to find a large set of potential matches in *Web of Science*. The scoring mechanism placed a higher weight on overlap in the title and author fields, gave less weight to exact page number matches, and demanded strict equality in the publication year fields. The highest scoring record, above a threshold, was chosen as the correct match.

In order to evaluate the recall of our crosswalk program, we determined the number of journals during each year available in *PubMed* as well as the *Web of Science* database. We calculated this overlap by matching a journal's full title, an ISO abbreviation, and its ISSN number. If our subset of the *Web of Science* data included all articles from the matching journals during those years, we would expect a theoretical matching threshold of approximately 65% in 1965 and as high as 93% in 2007. The dramatic rise in the potential upper bound of the match rate is due to Thomson Reuters' efforts to index scientific journals, which substantially increased starting in the late 1960s and has outpaced *PubMed*'s expansion during the past four decades (cf. Figure III). Social science and humanities journals remain excluded from *Science Citation Index Expanded*, however, and therefore papers from the fields of health economics, social work, education, and other areas available in *PubMed* could not be matched in the *Web of Science*.

Ultimately, we were able to bridge an average of 84.99% of the articles theoretically available for matching. Overall, this accounts for 7,195,192 matched records out of 8,839,865 *PubMed* records published during these years. In Figure IV, we show the efficacy of our matching procedure, in terms of the percentage of *PubMed* articles for which our algorithm matched to a single *Web of Science* record. Efficacy improves over the years as a function of the extent of information compiled by Web of Science. Non-English language papers, which increase over time, are harder to match because titles and author names are often indexed differently in the two databases.

The result of our efforts is an accurate indexing of the majority of the physical and natural science articles of *PubMed*. While the bridge, like *Web of Science Science Citation Expanded Database*, has somewhat limited coverage of non-English articles and no coverage of social science records, it provides the first effective means of doing empirical analyses that require bibliometric information from both of these databases.

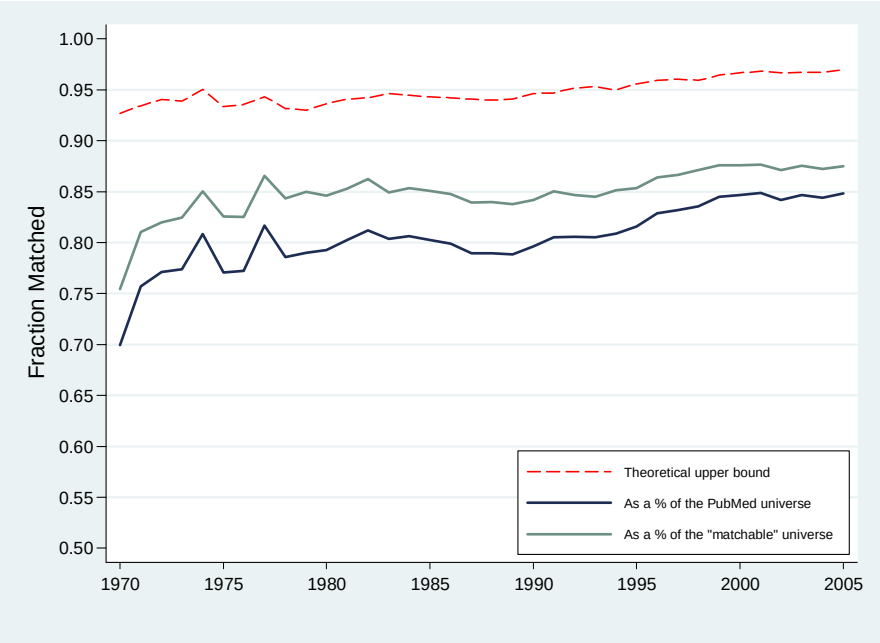
Figure III: Number of Publications Indexed by *PubMed* and *Web of Science*



Note: The above graph depicts the number of bibliographic records in the National Library of Medicine's *PubMed* database and Thomson Reuters's *Web of Science: Science Citation Index Expanded*. While the clear upward trend reflects the rapidly expanding corpus of scientific publications, the publishers of these two databases have used distinctive policies to determine the quantity and scope of journals to index each year.[2] Since the *Science Citation Index*'s inception in 1964, this bibliographic database has focused on indexing English language journals in the physical and life sciences receiving the majority of academic citations.[3] In contrast, *PubMed*, which became available in 1971, has always been intended as a database primarily for health practitioners and researchers.[4] Furthermore, the National Library of Medicine has emphasized "retrospective" information retrieval, while Thomson Reuters provides more tools for searching and evaluating recent publications.[5]

[2]Web of Science's selection process is described here:
http://thomsonreuters.com/products_services/science/free/essays/ journal_selection_process/
[3]History of Web of Science, http://www.garfield.library.upenn.edu/papers/mapsciworld.html
[4]http://www.sciencedirect.com/science/article/pii/S0740624X09000471
[5]http://www.fas.org/ota/reports/8219.pdf; http://wokinfo.com/media/pdf/jcrwebfs.pdf

Figure IV: Performance of our PubMed-Web of Science Bridging Algorithm



Note: This graph shows the efficacy of our process for matching *PubMed* articles to their corresponding citations in Thomson Reuters *Science Citation Index Expanded* (a subset of *Web of Science).* The theoretical match rate is the proportion of PubMed articles whose journal and volume of publication appear at least once in *Web of Science*. The match rate is the percentage of *PubMed* articles for which our algorithm matched to a single *Web of Science* record. We expect the match rate to be lower than the theoretical limit for at least three reasons. First, *Web of Science* covered fewer journals between 1965 and 1985 than in more recent years. In addition, the *Expanded Science Citation Index* contains only articles from the physical and life sciences, while *PubMed* incorporates articles on biomedical topics as well as, "Analysis of philosophical, ethical, or social aspects of the health professions or biomedical sciences."[6] Second, the publishers of *Web of Science* significantly truncated some parts of the bibliographic records during the first decades of the database's existence. For example, the omission or truncation of author names and article titles makes matching records between the databases more challenging. Lastly, the increasing prevalence of non-English language articles in both databases pose problems for our algorithm as titles and author names are often transliterated or translated and indexed in contrasting ways. Ultimately, our matching process bridges a significant portion of the *PubMed* data without an obvious bias.

[6]PubMed's policies are summarized at the following URL: `http://www.nlm.nih.gov/pubs/factsheets/jsel.html`

Appendix IV. Robustness of Graphical Analysis

To perform the graphical analysis in the main body of the paper, we create a dataset at the article-pair level, project the data onto "experimental time" (i.e., years before or after appointment, ignoring the influence of calendar time) and compute the raw mean and standard deviation in each year using the standard formulas. The means and 95% confidence intervals are then displayed in a graph like Figure Va below, which is an exact reproduction of Figure 3C in the main body of the paper.

While this approach has the merit of simplicity, there are also good reasons to prefer a graph displaying the coefficient of very simple regressions that replicate the spirit of the year-specific sample means analysis, but with minor adjustments.

First, there is the question of the standard errors that form the basis of the 95% confidence intervals. Those in Figure Va below are naïvely computed from the raw data. As a result, the confidence bands ignore the clustered structure of our data – the median number of articles per scientist is nine. For comparison, Figure Vb replicate Figure Va, except that the coefficients are estimated from an extremely simple regression that does no more than compute the sample means, but with the correct (i.e., clustered) standard errors.

Second, it would be trivial to purge the estimates of the effect of calendar time, by including as covariates in these regression models a full suite of calendar year indicator variables.

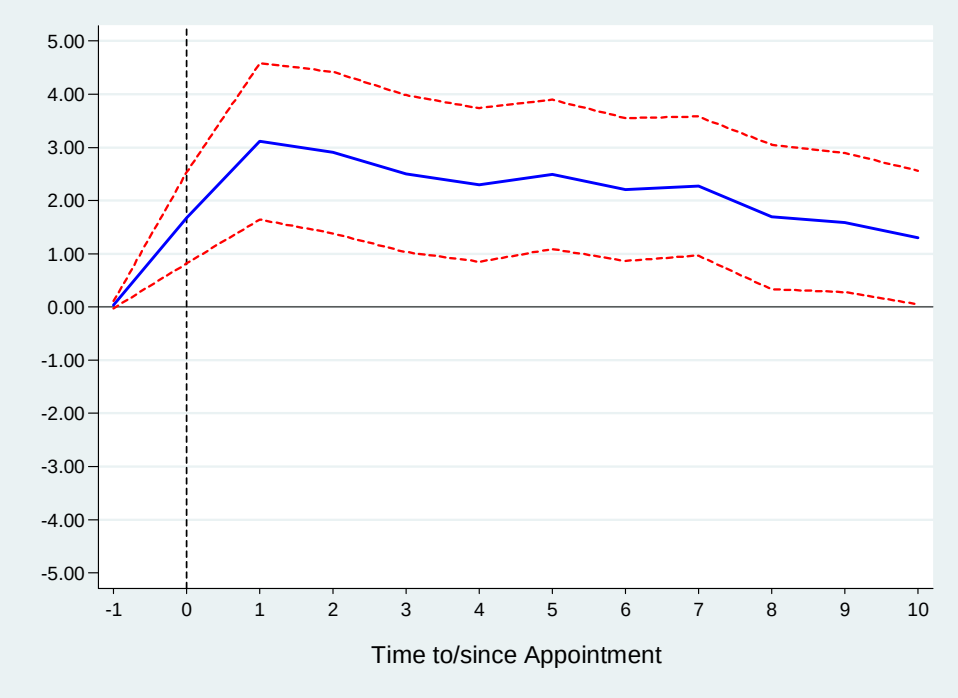
Third, Figure Vb still does not quite correspond to a dynamic version of the specification in Table 2, column 1, since that specification includes article-pair fixed effects. It is worth pausing to ask why these fixed effects could add value to our graphical approach since the pre-award mean of the outcome variable is approximately zero by construction. But recall that our matching procedure is coarse, not exact, so that this pre-award mean is not in fact exactly equal to zero. Moreover, the fixed effect transformation subtracts the mean of the outcome variable within the sample over the entire sample period from the contemporaneous level. And there is no presumption that this mean should be zero, even if our matching procedure was more "exact" than it is. In fact, for most of the analysis where we focus on the papers born one year before HHMI appointment, the fixed effect could theoretically have quite a major influence on the results.

We probe the implications of including fixed effects in Figure Vc, in which we graph the coefficient estimates corresponding to a dynamic version of Table 2, column 1. Once again, we are careful to cluster the standard errors at the level of the scientist, and the confidence intervals reflect this clustering.[7]

As can be seen below, these three figures tell a substantively similar story, though it is clear that the fixed effects contribute to tighten the confidence bounds around the estimates. The data and STATA code necessary to repliacte these analyses can be downloaded from `http://pazoulay.scripts.mit.edu/Data.html`.

[7] Note that the specication does not include year effects or career age effects as in Table 2.

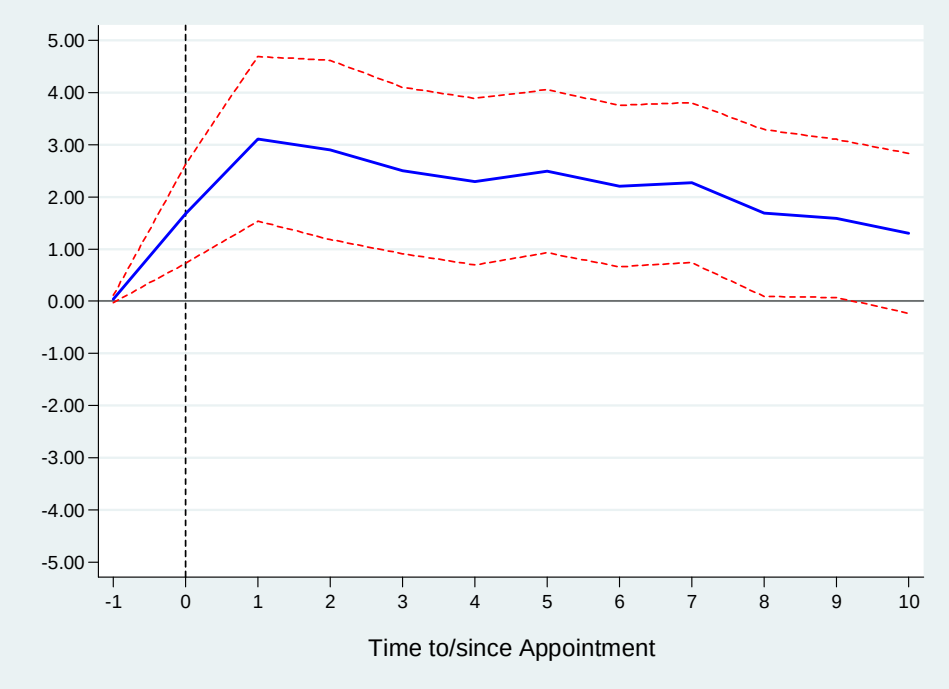
Figure Va: Effect of HHMI Appointment on Citation Rates
Pre-Appointment Article Pairs, Raw Means w/o Clustering



Note: Dynamics for the difference in yearly citations between HHMI-ECPW matched articles published in the year before appointment. Articles in each pair appeared in the same month, year and journal, and are also matched on focal scientist position on the authorship list, as well as overall number of authors. Further, control articles are selected such that the sum of squared differences in citations between control and treated article up to year t_0-1 is minimized — where t_0 is the year of (possibly counterfactual) appointment.

The solid blue line correpond to the sample mean difference in citations in each pair; the dashed red lines correspond to 95% confidence intervals based on sample standard errors.

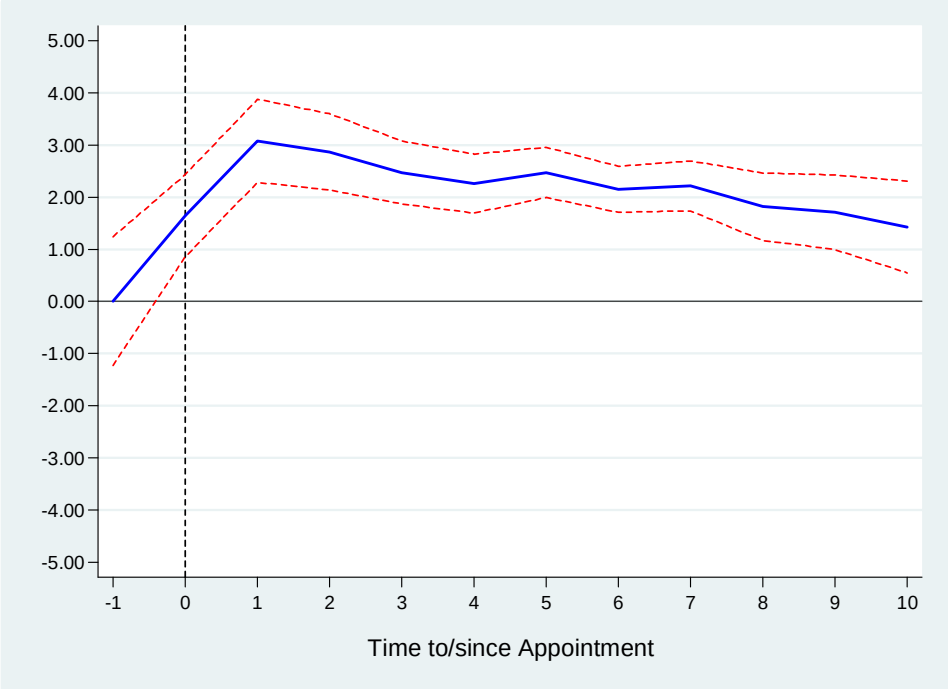
Figure Vb: Effect of HHMI Appointment on Citation Rates
Pre-Appointment Article Pairs, Raw Means w/ Clustering



Note: Dynamics for the difference in yearly citations between HHMI-ECPW matched articles published in the year before appointment. Articles in each pair appeared in the same month, year and journal, and are also matched on focal scientist position on the authorship list, as well as overall number of authors. Further, control articles are selected such that the sum of squared differences in citations between control and treated article up to year t_0-1 is minimized — where t_0 is the year of (possibly counterfactual) appointment.

The solid blue lines correspond to coefficient estimates stemming from a simple OLS regression in which the difference in citation rates for HHMI articles and their ECPW-controls are regressed onto 12 interaction terms between treatment status and the number of years before/elapsed since the appointment event (t_0-1, t_0, t_0+1,..., t_0+10 – the model does not include a constant term). The 95% confidence interval (corresponding to robust standard errors, clustered around HHMI investigator) around these estimates is plotted with dashed red lines.

Figure Vc: Performance of our PubMed-Web of Science Bridging Algorithm
Pre-Appointment Article Pairs
Estimates from Fixed Effects Specification



Note: Dynamics for the difference in yearly citations between HHMI-ECPW matched articles published in the year before appointment. Articles in each pair appeared in the same month, year and journal, and are also matched on focal scientist position on the authorship list, as well as overall number of authors. Further, control articles are selected such that the sum of squared differences in citations between control and treated article up to year t_0-1 is minimized — where t_0 is the year of (possibly counterfactual) appointment.

The solid blue lines correspond to coefficient estimates stemming from a simple linear, article-pair fixed effects regression in which the difference in citation rates for HHMI articles and their ECPW-controls are regressed onto 12 interaction terms between treatment status and the number of years before/elapsed since the appointment event (t_0-1, t_0, t_0+1,..., t_0+10 – the model does not include a constant term). The 95% confidence interval (corresponding to robust standard errors, clustered around HHMI investigator) around these estimates is plotted with dashed red lines.