HARVARD | BUSINESS | SCHOOL

Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs

Pierre Azoulay
Christopher C. Liu
Toby E. Stuart

Working Paper

09-136

Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs[•]

Pierre Azoulay
Christopher C. Liu
Toby E. Stuart

[•] The authors would like to thank Jim Baron, Matt Bothner, Ron Burt, James Evans, Fiona Murray, and Ezra Zuckerman for discussions on earlier drafts of this paper, as well as seminar participants at Yale, University of Chicago, Georgia Tech and UCLA. We thank David Caruso of the Chemical Heritage Foundation for assistance with the Pew Oral History Collection. Direct correspondence to cliu@hbs.edu or tstuart@hbs.edu.

Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs

ABSTRACT

Actors often match with associates on a small set of dimensions that matter most for the relationship at hand. In so doing, they are exposed to unanticipated social influences because counterparts have more interests, attitudes, and preferences than would-be associates considered when they first chose to pair. This implies that some apparent social influences (those tied to the rationales for forming the relationship) are endogenous to the matching process, while others (those that are incidental to the formation of the relationship) may be conditionally exogenous, thus enabling causal estimation of interpersonal influence effects. We illustrate this idea in a new dataset tracking the training and professional activities of academic biomedical scientists. In qualitative and quantitative analyses, we show that scientists match to their postdoctoral mentors based on two dominant factors, geography and scientific focus. Although they do not match on this dimension, they then adopt their advisers' orientations toward commercial science as evidenced by the transmission of patenting behavior. We demonstrate this in two-stage models that adjust for the endogeneity of the matching process, using a modification of propensity score estimation and a sample selection correction with valid exclusion restrictions. Furthermore, we draw on qualitative accounts of the matching process recorded in oral histories of the career choices of the scientists in our data. All three methods—qualitative description, propensity score estimators, and those that tackle selection on unobservable factors—are potential approaches to establishing evidence of social influence in partially endogenous networks, and they may be especially persuasive in combination.

I. Introduction

People select partners in relationships for many reasons. They match based on similarities in sociodemographic characteristics, positions in status hierarchies, spatial proximity, interest compatibility, power dependencies, and referrals from mutual acquaintances. We believe that most relationships arise from a matching process in which individuals pair on a limited number of high-priority dimensions. Although the importance of any particular factor differs across dyads, settings, and types of relationships, the actual ties that emerge among the immense array of connections that possibly could occur do so because individuals are complementary on a small set of meaningful characteristics.

Though people often match on just a few attributes, in totality actors possess many characteristics. This basic fact of social interaction introduces an element of randomness into the social influence process and, therefore, presents a strategic research site. If we consciously match to our associates on just a small set of carefully considered (or merely convenient) dimensions, we then expose ourselves to unanticipated social influences that arise from the attributes of our associates that never entered our calculus when we chose to interact. In other words, when two actors connect because they are compatible on a small set of attributes *X*, it may be that some set of additional characteristics *Z*, which was not considered when a choice was made to develop the relationship, results in the social transmission of attitudes and behaviors. Moreover, we argue that even those relationships formed for instrumental purposes often result in de facto chance interactions; if actors are rendered as discrete attributes, the fact that matching takes place on just a few characteristics means that potential influences in the rest of the attribute vector are left to chance. For this reason, we refer to social matching as being only “partially” deliberate.

We observe this phenomenon in an analysis of the origins and consequences of the matching of postdoctoral candidates to their faculty advisers. The group we study is Pew and Searle Scholars (hereafter, often “Scholars”), a set of prominent, young academic life scientists. Exploiting an extensive quantitative database and a qualitatively rich archive of oral histories, we find that two factors dominate the matching process between postdoc candidates and their advisers in both data sources: compatible scientific interests and geographic location.

In a second-stage analysis, we then show that whether a Scholar's postdoctoral adviser was a patenter during or before the time the Scholar arrives in his or her adviser's lab has a large effect on whether the advisee subsequently becomes a patenting scientist. By estimating this effect in two-stage models that adjust for the endogeneity of adviser-advisee pairings, and by relying on the qualitative evidence in the oral histories, we show that postdoctoral candidates *do not* appear to consider their advisers' patenting behavior when establishing the match. Thus, the evidence indicates that the transmission of behavior is a true social influence effect: it is causal, rather than stemming from common interests that underlie the match between candidates and advisers.

Our work relates to three literatures in sociology. First, our findings contribute to a burgeoning literature on the interface between academic and commercial science (Audretsch and Stephan 1996; Etzkowitz 1998; Evans 2004; Murray 2009; Owen-Smith and Powell 2001a; Owen-Smith and Powell 2004; Stuart and Ding 2006; Zucker, Darby and Brewer 1998). As these authors have observed, faculty members' decision to patent scientific findings and to start or advise companies is influenced by many factors, including scientific norms, peer and employer effects, and the reach of their social networks across the porous boundary demarcating academic and commercial science. Our empirical findings contribute to this work by showing the imprints of postdoctoral advisers on the subsequent choices of the scientists-in-training that travel through their laboratories. Especially in the formative years of a career in which individuals are apprenticed into a professional community's social norms and identities, they are likely to be susceptible to the influence of the role models to whom they are exposed. We find this to be the case for Pew and Searle Scholars, and we believe it has implications for the long-term stratification structure in commercial science.

The core empirical result also dovetails with the literature on career sequences (Abbott 2001; Abbott and Hrycak 1990; Stovel, Savage and Bearman 1996). Our findings suggest that the mentors one encounters early in a career have consequences not only along the anticipated dimensions that give rise to mentorship dyads, but they also cause unplanned detours in career trajectories. In this sense, the findings expose one type of "turning point" in academic scientists' career trajectories (Abbott 1997; Elder 1985). This result is interesting not simply because postdocs' career paths are shaped by the professional relationships they form, but because on the dimension on which we assess mentors' influence, the matches we study are neither deliberately created nor are they the outcomes of a standard assortative matching process. Therefore, despite the agency displayed in the creation of these important professional relationships, the consequences of the ties actors form extend well beyond the narrower rationales that first drove their creation. However strategic actors may be in forming ties, healthy doses of bounded rationality and incomplete information prevent interacting parties from predicting ahead of time the myriad ways in which they may come to influence one another.

A third contribution is a novel methodology for empirically establishing evidence of a social network effect. A growing chorus of authors in sociology has critiqued the social networks literature because of its inattention to the challenge of empirically establishing *causal* network effects (Mouw 2003; Reagens, Zuckerman and McEvily 2007; Stuart and Sorenson 2009; Winship and Morgan 1999). This stems from the fact that actors' positions in social networks rarely are exogenous to the outcomes that interest researchers. Indeed, sociologists have long observed that evidence of true social influence can be very difficult to disentangle from the mutual selection of like-minded individuals into relationships (Kandel 1978; Newcomb 1961). If we have any reason to believe that actors are deliberate in seeking relationships and that they have at least some discretion in the matches they form, then underlying individual differences—in intelligence, charisma, strategic orientation, gender, race, age, religion, socioeconomic status, …—will influence and therefore correlate with network positions. The same is true for any unobserved dimensions of commonality among the actors that choose to form ties (Van den Bulte and Lilien 2001). In studies of social influence processes, it is difficult to separate the true effect of social ties from the factors that cause the ties to come to be in the first instance.

We have a four-pronged approach to the challenge of showing that Pew and Searle Scholars match to their postdoc advisers on a few primary attributes, but a secondary dimension that *does not* shape the likelihood of a match subsequently *does* influence Scholar behavior. First, we have coded 62 comprehensive oral history transcripts of interviews conducted with Pew Scholars and find that *none* of the 62 ever mention would-be advisers' commercial activities as a factor in their selection of postdoctoral fellowship. The oral histories are, however, consistent in describing scientific topic and geography as drivers of the matching process. Second, we have constructed dyad-level matching regressions between PS Scholars and postdoc advisers. These regressions show that Scholar-adviser pairing is independent of advisers' commercial activities. Third, after generating estimates of the probability that protégés match to specific advisers, we then employ a variant of propensity score estimation (Imbens 2000) to assess the post-match effect of advisers' commercial orientation on Scholar patenting. But because the central assumption of propensity score estimators could be violated in out data, we implement a final analysis: we use Heckman's (1979) two-stage estimator in which we regard the observation of only actual (versus potential, but never-formed) Scholar-adviser matches as an instance of a sample selection problem. This approach is valid only if there are one or more "instrumental variables" that predict pairing between Scholars and advisers but can be legitimately excluded from the outcome equation. We have collected two instrumental variables – the first based on proximity between a scholar's undergraduate institution and the location of his/her postdoc employer, the second based on shared national heritage between the Scholar and his/her postdoctoral mentor —that enable us to recover estimates of advisers' influence on Scholars' behavior even in the presence of residual selection on unobserved factors.

Before further developing the argument, we note that each of these tactics—using qualitative data to establish the relevance of some but not other dimensions of matching, constructing a first-stage matching equation, using two-stage selection-on-observables estimation techniques, and using exogenous variables to instrument for the formation of matches—represent general strategies for identifying non-spurious network effects. And although none of these may be conclusive on their own in any particular study, they can be persuasive in combination.

II. Matching and Social Relations

Many studies in sociology consider the origins of social relations. The seminal works of Lazarsfeld and Merton (1954) and Blau (1977), for instance, describe the roles of homophily and population demographics in driving interactions. It is well documented that social relationships cluster among those who are categorically similar along a core set of ascribed attributes and who share status characteristics. Conceiving of the social world as a multidimensional space that is delineated by salient sociodemographic characteristics (McPherson, Popielarz and Drobnic 1992; McPherson, Smith-Lovin and Cook 2001), individuals who are proximate in this space are much more likely to be socially connected.

The literature on homophily intertwines with a long history of research on the spatial geography of relationship formation. Because the likelihood of chance interactions increases with spatial propinquity and the cost of maintaining relationships declines with proximity, social interaction depends on geographic nearness. In relationships as varied as marriage (Bossard 1932), the workplace (Allen 1977), board directorships (Kono et al. 1998), and investment syndicates (Sorenson and Stuart 2001), spatial proximity has proven to be a central determinant of interaction probabilities. In fact, separating people by even very small distances—for instance, a floor in a housing complex—can dramatically decrease the chance that relationships will emerge (Festinger, Schachter and Back 1950).

Of course, there are also instrumental theories of relationship formation and these may be especially pertinent in labor market contexts such as the one we study. In rational choice models of network formation, actors often are assumed to enter relationships only in anticipation of some specific benefit (Jackson 2008). Although sociological accounts of networks in the labor market have tended to focus on how network structures affect work and career outcomes rather than on the origins of the edges that form these networks, many theories are compatible with agentic perspectives on the creation and exploitation of social ties. Certainly, agency in tie formation is strongly implied in Burt's (1992) theory of structural holes, as it is in studies of the mentorship and other types of support and buy-in networks in the workplace (Podolny and Baron 1997).

A common denominator in this large literature is that one cannot assume that relationships emerge randomly. This creates the well-known challenge of distinguishing true peer effects from their likely correlates; namely, the processes that lead actors to particular matches may mirror, influence, or cause, relationship-based outcomes of interest. This challenge is marked. How may we be confident that any similarity in attitudes or behaviors between an adviser and those she advises reflects a causal influence of the former on the latter, rather than the mutual choices of two similarly pre-inclined individuals to work together?

In addressing this question, we begin with an assumption, which we believe to be broadly characteristic of the process of relationship formation. Whether occurring in social- or market-based contexts, actors often strike up matches based on a small set of important characteristics for the relationship at hand. This is likely to be true of casual ties in which the convenience of interaction matters most; bonds formed in the context of common social foci (Feld 1981); or market exchanges formed for instrumental purposes. In all these instances of matching—even those in which actors are strategic in seeking partners—we contend that actors do not optimize matches over a high dimensionality attribute space. Rather than an algorithmic search, matching typically occurs on the basis of the few factors that matter most to participants; people do not, in general, consider complementarities across their full ranges of characteristics when forming ties. Thus, our central theoretical claim and the goal of the empirical analysis is to show that, *when actors form relationships based on characteristics X but do not match on some additional set of attributes Z, we have an opportunity to study social influence along attributes Z in a context that may be relatively untainted by the process leading to the assignment of actors to matches*.[1]

As we will show, this situation is particularly valuable for identifying causal social influence effects when we have a thorough understanding of—and ability to model—the factors at the core of the matching process. In such instances, selection-on-observables estimation techniques may reliably

[1] As we will describe in more detail, it poses no challenge for the empirical strategy if the elements of X and Z are correlated *as long as matching only takes place on the Xs*. Thus, our primary methodology is suitable to situations in which individuals match on primary sociodemographic variables (ethnicity, age, education, …) and are then exposed to unanticipated social influences in attitudes that may be correlated with these attributes, such as political views or preferences for leisure activities.

uncover network effects. But independent of its advantage for methodological considerations, we believe that the dynamic whereby actors match on a few attributes but are then exposed to a larger set of characteristics is fundamental to understanding myriad, unforeseen social influences on actors' attitudes and behaviors.

III. Context: Adviser-Advisee Pairings Among Postdoctoral Fellows

Sociologists have had a long-running interest in postdoctoral fellows. Because of their prevalence, postdocs are integral to the everyday fabric of laboratory life (Knorr-Cetina 1999). The postdoc system also reinforces the status system in science. Not only are next-generation scientific leaders far more likely to complete postdocs with the elite of the current generation, but from an adviser's standpoint, the successful placement of postdocs is itself a core dimension of status accrual in science (Long, Allison and McGinnis 1979). The postdoctoral period is also considered to be a primary locus of socialization in the profession (Hagstrom 1965). It is the time during which young scientists engage in anticipatory socialization in preparation to assume the role of laboratory head. More generally, apprentices undergo long periods of exposure to the general professional values and more idiosyncratic opinions and scientific "styles" of their particular mentors (Zuckerman 1977). Because of the length of the postdoctoral training period (Stephan and Ma 2005), the interdependence of the work of candidates and advisers, and the hierarchical aspects of the relationship, this period is a time in which apprentices are highly susceptible and heavily exposed to the attitudes, behaviors, and styles of mentors.

Given postdoctoral fellows' essential role in academic laboratories, one might expect an efficient labor market to govern the pairing of candidates and advisers. In reality, the market for postdocs is not orderly—there is no central clearinghouse to optimally match candidates to available positions. Indeed, it would be reasonable to regard the postdoc hiring process as the antithesis of the highly structured National Resident Matching Program, which matches graduate medical residents to open positions on a single day. As a result, postocs typically find advisers through a process in which local search and chance events loom large in the connections that ultimately are established.

III.A. Study Population: Pew & Searle Scholars

Among academic life scientists, we study individuals who have been selected as Pew Scholars or Searle Scholars. These awards are granted to "young investigators of outstanding promise in the basic and clinical sciences relevant to the advancement of human health".[2] Unlike other accolades such as the Nobel Prize or a National Academy of Sciences nomination, these awards are granted on the basis of the future promise of nominees' research agendas rather than their past achievements. When the awards are bestowed, recipients have minimal track records of non-supervised research.

PS Scholars are broadly distributed across US research institutions. This is a function of the eligibility requirements for the Award—the right to nominate Scholars is granted to institutions. In 2007, for example, the Pew Foundation solicited a single nominee from each of 148 US research institutions. Twenty Pew Scholars were ultimately selected from these nominees. For Searle Scholars, 120 universities nominated 182 newly appointed assistant professors, 15 of whom were selected. Since the inauguration of the program, a per-year average of 35 Pew and Searle Scholars has been named.

For a number of reasons, PS Scholars are an attractive group for our analysis. First, because the Award is granted at the time that scientists begin their independent academic careers, we can construct a prospective dataset vis-à-vis the commercial orientation of the Scholar *after* s/he enters an independent research career. Second, the emphasis of the Award on the "advancement of human health" means that the research trajectories of most PS Scholars will straddle the academic-industry boundary; many Scholars will engage in potentially commercializable research, although not all choose to pursue this aspect of their work. This means that the decision to patent in this group may be as influenced by scholarly priorities and values as by the commercial significance of the underlying research. Likewise, studies show that the commercialization of academic research in the life sciences is heavily concentrated among the academic elite (Zucker, Darby and Brewer 1998).

[2] Quoted from the Pew Scholars Program Description at http://www.futurehealth.ucsf.edu/biomed/scholdes.html; accessed 9/30/07. These awards confer significant status to recipients, but the monetary component is generally insufficient to change the recipient's scientific research trajectory.

There is thus reason to be interested in the variance in behavior among leading scientists, versus the well-documented, steep decline in commercial activity among less productive scientists (Azoulay, Ding and Stuart 2007).

Lastly, there is one important advantage of studying the population of Pew Scholars specifically. Each recipient of a Pew Scholar Award is asked to participate in an oral history, with interviews conducted and transcribed at the cessation of the Award period. These transcripts, which we describe in detail next, are rich accounts of scientists' professional experiences and values, as well as the rationales for their career choices.

IV. Oral Histories

The Pew Scholar Oral History and Archives Project has collected the life histories of more than 200 Pew Scholars. In the literature for the program, the stated purpose is to record, "… not only the science being pursued, but also the lives of scientists. While each oral history is unique, many of them explore issues related to the Scholars' childhood, college experiences, time training in various labs, their time as a principal investigator, and broader social, political, and cultural issues related to science."[3]

The oral histories help us to understand candidates' decisions to pursue postdoc positions with particular mentors. Because the insights gained from these documents inform the matching equation in the statistical analyses, we will first report findings from them. As we will see, the oral histories buttress the argument that postdoc-adviser matches are fashioned around a limited set of dimensions.

We randomly chose 62 interview transcripts, which ranged in length from 98 to 411 pages. To analyze these documents, we first read five volumes to inductively generate criteria that were cited by Scholars as being important in the search for a postdoctoral adviser. These categories were scientific focus, geographic considerations, adviser status, and interpersonal rapport. We then added a fifth category, commercial considerations, although none of the five interviews expressed a preference for matching on this criterion.

[3] http://www.chemheritage.org/exhibits/ex-nav2-pew.asp; accessed 12/30/08

A coder then read each transcript to identify the section(s) describing the Scholar's choice of a postdoctoral adviser. For each transcript, the coder indicated if a given category was cited as a determinant for pursuing a particular postdoctoral position. The coder then excerpted relevant quotations, and also recorded any additional factors that fell outside the five primary categories. For example, Susan Birren, who received a Pew Award in 1996, earned her Ph.D. from UCLA and then transitioned to a postdoc at CalTech. In describing her search for a position, Dr. Birren recalled,

> "He [husband] had been in his postdoc for a couple of years, didn't want to leave, and so I again looked locally, and also ended up at Caltech…"
> "…At that point professionally, I was looking for a change, because what I had been doing as a graduate student was pretty straightforward transcriptional regulation… . So I talked to several people and ended up going to David Anderson's lab. He was a developmental neuroscientist … it seemed like a major problem that you could spend a long time working on."[4]

From these and related passages, the coder determined that this Scholar chose her postdoctoral adviser based on scientific interest and geographic constraints.

Findings from the oral histories are presented in Table 1, which records the percent of Scholars who describe the attribute on each row as a critical factor in pursuing a particular postdoctoral adviser. The most prevalent—in fact, ubiquitous—consideration in selecting an adviser was scientific focus. Only three Scholars did not cite scientific interest as a major factor in their decision to seek a position in a particular mentor's laboratory, and these were due to exceptional circumstances.[5] Although scientific interest did not always imply that trainees intended to continue in their current line of research—some individuals, such as Dr. Birren, used the postdoc period to shift scientific trajectories—the majority of Scholars hoped to build upon the areas of expertise they had developed during graduate school.

Over half of the Scholars also singled out geography as a major factor in their search. In 19 (31%) cases, Scholars reported that geography was a binding constraint. In these instances, family considerations, most often regarding a partner's career, limited Scholars' search to particular regions.

[4] Susan J. Birren, interview by William Van Benschoten at Brandeis University, Waltham, Massachusetts, 2-4 August 2004 (Philadelphia: Chemical Heritage Foundation, Oral History Transcript # 0459)
[5] For example, one Pew Scholar was scheduled to train with David Baltimore. One month prior to the start of the fellowship, Baltimore accepted the presidency of Rockefeller University and moved from Boston to NYC. He then arranged for the Scholar to train under (fellow Nobel Prize winner) Phillip Sharp at MIT.

For example, Nancy Hollingsworth received a PhD from the University of Washington and limited her postdoc search to the Seattle region:

> "We [Hollingsworth and partner] were together when I was 25, and as I was beginning to finish, I set up my postdoc to stay in Seattle so that we could stay together. So I arranged to go to Gerry Smith's lab at the Fred Hutchinson Cancer Center. …"[6]

In another 14 (23%) cases, individuals cited a strong personal desire to reside in a particular area, rather than a binding family constraint. All told, 33 of 62 oral histories stated that geographic limitations or preferences loomed large in their search for postdoc positions.

A third factor that garnered frequent mention is a potential adviser's scientific prestige. For example, Mark Kamps reported that he first heard about his postdoctoral adviser through a fellow graduate student:

> "I remember Anna… was another graduate student in Bart's [graduate adviser] lab and she wanted to go to David Baltimore's lab as a postdoc. She was really focused on that. … So I said, 'Who's David Baltimore?' and Anna said, "Oh, David this and this. Oh, and he's got a Nobel Prize, and he worked on one of the kinases"… So I should have known his name."
>
> With his interest piqued, Mark Kamps reached out through his informal network: "So I asked Inder Verma, who was a scientist at the Salk institute, if I could meet with David [Baltimore] when he was coming out to give a talk. And Inder said, "Sure."[7]

When the interview transcripts are systematically coded, scientific interest and geographic considerations clearly are the foremost criteria in candidates' minds when they search for advisers. In a smaller proportion of cases, scientific status and interpersonal attraction were also decision criteria.[8] These results closely coincide with those of prior surveys of the motivations for postdoctoral adviser choice (Nerad and Cerny 1999).

---

[6] Nancy M. Hollingsworth, interview by William Van Benschoten at the State University of New York at Stony Brook, Stony Brook, New York, 11-13 November 2002 (Philadelphia: Chemical Heritage Foundation, Oral History Transcript # 0465)

[7] Mark P. Kamps, interview by Andrea R. Maestrejuan at the University of California, San Diego, San Diego, California, 10-12 February 1998 (Philadelphia: Chemical Heritage Foundation, Oral History Transcript # 0437).

[8] Although adviser status was far from a universal concern (only 15% of Scholars explicitly stated that they sought an adviser based upon his/her prestige), we suspect that this is due to the fact that many individuals in the dataset considered *only* high status advisers, and did not view prestige differences among the select members in their consideration set to be germane to their decisions.

Two additional points are relevant to our argument. First, there is a complete *absence* from the oral histories of any mention of the commercial aspects of science when selecting advisers. There was no instance in which any of the 62 informants reported considering future commercial activities—such as the opportunity to patent, to gain connections with industry, or any other form of engagement with commercial-sector entities—when choosing a postdoctoral adviser. Readers may be concerned that this is because subjects considered it unsavory or counter-normative to discuss the commercial aspects of science. Given prevailing academic norms, it is possible that Scholars had—but were reluctant to reveal—commercial aspirations when choosing postdoc advisers. Although it is not possible to definitively rule out this explanation for the lack of reference to commercial motivations in the postdoc matching process, a number of the oral histories did specifically address the issue of academic patenting, although not in the context of the postdoc adviser search. In one-third of the oral histories, Scholars were explicitly asked about their thoughts regarding their own later-career patenting activities (if they had patented before the time of interview) as well as the interplay between commercial interests and academic science. Although scientists' perceptions of the social value of patenting varied greatly, all Scholars' responses appeared to be candid. In no instance did a Scholar decline to respond to the question and in a number of cases, Scholars were explicitly positive about the value of patents. Therefore, we believe that at least some Pew Scholars would have discussed their commercial interests if they were germane to their search for postdoc advisers.

Second, when we decompose the postdoc adviser choice into categories of relevant factors, the data tally to the numbers presented in Table 1. However, the table does not convey the overall impression one forms when reading the oral histories in their entirety. From these documents, it appears that the confluence of different elements of chance strongly shape the career experiences of Pew Scholars. Rather than working backward from a set of well-defined career objectives to a search for an optimal adviser match, the process individuals follow to find a postdoc mentor is one of local search in delimited scientific and geographic spaces, coupled with a heavy role for chance encounters. While the search and matching process is not entirely random, neither does it encompass a large number of dimensions. For this reason, we believe the matching process is "partially" deliberate.

V. Sample, Data and Methods

Sample. We have identified the names of all Pew or Searle Scholars since the inception of the awards (1981 for Searle and 1985 for Pew Awards) to year 2000. All told, we began with 583 PS Scholars.[9] Individuals are captured in our sampling frame when they receive the PS Award. To conduct our analyses, however, we also require information on both graduate school and postdoc advisers. We therefore search backward in time to identify all advisers for these 583 PS Scholars. Ultimately, this process reduced the analyzable sample to 489 Scholars, as the majority of the remaining individuals were MDs (and therefore did not have identifiable graduate school advisers). These 489 Scholars apprenticed as postdocs in the laboratories of 333 unique advisers.

Methods. Estimating the causal effect of mentors' influence on Scholar career outcomes must address the basic selection problem that adviser "assignment" is the outcome of a non-random matching process. Our specific concern is that if (contrary to the self-reports in the oral histories) scientists-in-training choose whether they intend to pursue commercial science during graduate school, then commercially oriented graduate students may seek postdocs in the laboratories of like-minded advisers, and vice-versa. Matching on taste for commercial science could produce a spurious estimate of postdoc adviser commercial propensity on Scholar career outcomes in any estimation that does not account for the endogeneity of the match. Therefore, standard statistical techniques, which assume that mentor assignment is exogenous, may not recover causal effects.

We contend that postdoc-adviser matching is indeed deliberate, but only partially so because it is driven by the few, primary factors highlighted in Table 1. In addition to using the oral histories to better understand the matching process, we employ two statistical approaches that adjust for matching to estimate a causal effect of adviser imprinting. First, we use a variant of propensity score

[9] There has been a total of 642 Pew or Searle Scholars. From this population, we dropped 57 individuals from disciplines that are peripheral to biomedicine, such as populationbiology and clinical psychology. The rate of patenting in the dropped group was similar to the retained sample, but because we rely on the Medline database to construct many of the covariates, we limited the sample to Scholars for whom the vast majority of publications were indexed in Medline. We also dropped one individual due to a precipitous retirement and another who succumbed to cancer within two years of receiving his Award.

estimation, which is known as a "selection on observables" approach because it is valid only under the (not testable) assumption that the outcome of interest is independent of assignment to treatment conditional on observable factors. Second, non-random matching between Scholars' and advisers can be considered to be an instance of a sample selection problem because we observe actual matches but do not observe potential matches that did not—but could—occur. Framing the problem this way, we can analyze the data in Heckman's (1979) two-stage sample selection framework, in which the first stage is a binary choice matching equation consisting of observed and counterfactual matches, and the second stage examines the probability of Scholar patenting.

A. Selection on Observables: Inverse Probability of *Exposure* Weights.

Consider a scenario in which each Scholar i (I =1, …, N) is assigned a particular mentor j from a pool of J potential mentors. As explained by Imbens (2000), one can think of mentor assignment as a multi-valued treatment $T \in \{1, \ldots, J\}$. In the pre-assignment period, we measure X_i^k, a set of prognostic factors for assignment. In our case, these are dyad-level measures of the fit of a particular Scholar with each possible mentor k=1, …, J. An outcome of interest y_i is then measured. In our case, treatment will occur when a Scholar is assigned to (i.e., matches with) an adviser who patents prior to or during the time the Scholar is a trainee in the mentor's lab, and the outcome is whether the Scholar files for a patent later in his/her career.

Let y_i^k be the value of y that would have been observed had Scholar i been assigned to mentor k (where assignment can be counterfactual, i.e., $k \neq j$; the Scholar need not be pared with his/her own mentor). Our key assumption is that mentor assignment is *unconfounded*, that is, the match between Scholar and adviser is statistically independent of y_i^k conditional on the observable factors X.

We model the effect of a particular adviser trait, patenting, on the mean of y^k conditional on assignment and exogenous scholar characteristics Z, as:

$$E[y_i^k \,|Z_i,PATENT_k] = \beta_0 + \beta'_1 Z_i + \beta_2 PATENT_k \qquad (1)$$

where $PATENT_k$ is an indicator variable capturing whether the Scholar would have been exposed to that particular trait had s/he, possibly contrary to the fact, been assigned to mentor k. Imbens (2000) shows that under unconfoundedness, β_2, the causal effect of adviser patenting, is identified and can be recovered by estimating:

$$E[y_i^j \mid Z_i, PATENT_j] = \beta_0 + \beta'_1 Z_i + \beta_2 PATENT_j \qquad (2)$$

by weighted least squares or weighted maximum likelihood (depending on the distribution of y), where the weights correspond to the inverse probability that i is assigned to his/her actual adviser j. Note that (2) differs from (1) in that the observed assignment j and outcome y^j have been substituted for the counterfactual assignment and outcome (k; y^k). Another difference is that the expectation in (1) is taken over the sample of all possible dyads. In other words, it includes all realized matches between Scholars and advisers as well as counterfactual matches. In contrast, all variables in (2), the second stage regression, are only defined for the sample of actual mentor-trainee dyads.

Implementation of IPEW is straightforward. Under unconfoundedness, the selection bias can be removed by weighting the regression by:

$$w_i = \frac{1}{Prob\left(T_i = j \middle| X_i^j\right) \cdot \prod_{k \neq j}^{K} [1 - Prob(T_i = k | X_i^k]} \qquad (3)$$

The denominator of w_i represents the conditional probability that a scholar was assigned his or her actual mentor j, which is composed of two terms: the probability of a match with j, and the probability that s/he does not match with any of the other mentors k, $k \neq j$. Suppose, as we will assume, that all relevant factors determining matches are observed and included in X. Then, weighting by w_i effectively creates a pseudo-population of scholars in which X no longer predicts assignment and the causal association between adviser patenting and outcome is the same as in the original population. We refer to β_2 when equation (1) is weighted by w_i as the Inverse Probability of *Exposure* Weighted (IPEW) estimator of β_2.[10]

[10] We have coined the term Inverse Probability of *Exposure* Weighted Estimation to refer to this variant of the more widely known Inverse Probability of *Treatment* Weighted Estimation (e.g., Robins et al. 2000), which is itself a close cousin of the propensity score (Rosenbaum and Rubin 1983). The two techniques are similar in that they both assume that treatment (in the IPTW case) or assignment to a match (in the IPEW case) is unconfounded, conditional on observables. They differ in that IPTW estimation fits situations in which an individual or firm *self-selects* into a single, well-defined treatment. By modeling the probability of self-selection for control and treated individuals, the social scientist can recover the causal effect of treatment. In contrast, IPEW is suited to the related case in which the set of observed matches reflects a prior *sample selection* process. With IPTW estimation, the first and second stages are estimated using the same sample. In contrast, with IPEW the second stage is estimated on the sample of realized

To create the weights in equation (3), we first estimate a pooled cross-sectional logit at the Scholar-adviser dyad level:

$$Prob(T_i = k) = \alpha_0 + \alpha_1 X_i^k + \delta_t \quad (4)$$

where $Prob(T_i = k)$=1 for actual Scholar-adviser matches and =0 for all counterfactual matches, X_i^k includes dyad-level covariates predicting matches between Scholars and advisers, and δ_t represents match year indicator variables. The estimates from equation (4) are also used to create the denominator of the weights in equation (3), which are computed by a simple multiplication of fitted probabilities as implied by the equation. Of course, this matching regression is of substantive interest in its own right; it reveals correlates of postdoc-mentor pairings.

IPEW estimation is simple to implement, but the unconfoundedness assumption (that the observed determinants of mentor assignment are exhaustive) is a strong one, and its validity cannot be tested. We do know that techniques assuming selection on observables perform best when it is possible to include a comprehensive list of covariates to model the probability of treatment or assignment (Dehejia and Wahba 2002). In most samples, determinants of this nature typically would not be available to the researcher. However, we have chosen a study population that may satisfy the selection-on-observables assumption required for IPEW. But because we cannot formally test the unconfoundedness assumption, we also utilize an alternative approach.

B. Selection on Unobservables: Heckman Selection Correction.

Although the oral histories suggest that commercial opportunities do not drive the choice of mentors at the postdoctoral stage, there still may be a residual factor that influences both mentor assignment and contact with the commercial sector once a scholar has secured an independent position. The existence of any such unobserved factor would undermine the validity of the IPEW estimates.

An alternative to IPEW to estimate the causal effect of adviser patenting is to isolate influences on the matching process that occur quasi-randomly, and to rely solely on this variation to

---

matches, whereas the first stage is estimated using the sample comprising both realized and counterfactual matches. Therefore IPEW estimation can be regarded as the selection-on-observables counterpart to Heckman's (1979) sample selection correction, whereas IPTW can be thought of as the selection-on-observables counterpart to the estimation of treatment effects using two-stage least squares (Angrist, Imbens, and Rubin 1996).

estimate the effect of adviser imprinting. To implement this approach, we require instrumental variables—variables (exclusion restrictions) that are *relevant* for assignment, in that they strongly predict pairing, but can be assumed to be orthogonal to unobserved determinants of the outcome of interest, and therefore legitimately *excluded* from the outcome regression.

We propose two exclusion restrictions. The first is the proximity between Scholars' undergraduate institutions and the universities where they might become a postdoctoral fellow. The second is shared nationality between the Scholar and a potential mentor, conditional on being born outside the US. The relevance of these instruments ultimately is an empirical question, and we will show below that these two variables predict the likelihood of specific scholar/mentor pairings. The validity of the instruments, respectively, rests on the assumptions that, (1) Scholars' choice of undergraduate institution does not reflect later-career commercial dispositions; and (2) national background is not systematically correlated with entrepreneurial interests. We believe these assumptions to be plausible in this setting.

Neither of these instruments is relevant for the full sample of Scholars because they generate variation in pairing in two distinct subpopulations. Specifically, shared national background with a potential postdoc adviser cannot explain variation in pairing among U.S.-born scholars, since in that subpopulation, this variable measures only whether the adviser is foreign-born. Conversely, for foreign-born Scholars, variation in proximity between U.S.-based postdoc institutions and one's undergraduate university is unlikely to be informative. Therefore, we will perform the sample selection analysis separately on these two subpopulations, and there is no presumption that the different instruments should yield identical treatment effects.

Formally, we assume that Scholar-adviser pairings arise from an unobserved matching process, during which some matches are accepted, while others are not. The specific form of endogeneity that concerns us is that we observe only the realized matches, and not those that were possible but did not come to be. Formally, we assume the existence of the underlying relationship:

$$y_i^k = \beta_0 + \beta_1' W_i^k + \beta_2 PATENT_k + \varepsilon_{ik} \qquad (5)$$

The dependent variable, however, is only observed for realized pairing (i.e., we do not observe later-career patenting behavior for Scholars who were "assigned" to any mentor other than their actual postdoc adviser). We model the probability of a match—the selection equation—as follows:

$$Prob(T_i = j) = \alpha_0 + \alpha_1 X_i^j + \delta_t + \eta_{ij} \quad (6)$$

where $Prob(T_i = j)$=1 for realized matches between Scholars and advisers and =0 for counterfactual matches, η and ε are both assumed to be standard normal random variables with correlation coefficient ρ. y_i^j is observed if and only if $\alpha_0 + \alpha_1 X_i^j + \delta_t + \eta_{ij} > 0$.

Just as in the first stage of the IPEW regressions (equation (4)), to estimate the sample selection equation arising from this data generating process, we create a sample of mentor-Scholar matches that might have occurred. This allows us to correct for sample selection by first estimating the probability that Scholar-mentor matches occur and then the likelihood that the Scholar will patent, conditional on the existence of the match. In effect, we are drawing a sample of mentor-Scholar pairs that chose not to match. Since we cannot know the 'true' rejection rate of matches in our sample, we perform robustness checks by varying the degree to which we sample counterfactual matches relative to realized ones.

While the selection model is formally identified through the nonlinearity of the selection equation, non-parametric identification relies on the two exclusion restrictions discussed above. In practice, shared national background and proximity to undergraduate institution will be included in the vector of variables X in the first-stage selection equation (6), but excluded from the vector of variables W in the outcome equation (5).

Finally, in contrast to IPEW, the Heckman approach does not require the assumption of unconfoundedness, but it does depend on the validity and relevance of the exclusion restrictions. The attractiveness of this approach is its ability to identify the causal effect of mentor imprinting even in the presence of residual selection based on unobservable influences.

C. Data Construction. Our analysis utilizes four primary data sources. First, we requested CVs from all Scholars to identify dates of training periods and degrees, as well as information on advisers and the location of undergraduate institutions.[11] Second, we supplemented the information on graduate school training with the Proquest Dissertation Abstracts database. Third, we obtained patents by matching scientist names to data from the US patent office.[12] Fourth, to construct measures of scientific outputs and content, we collected all 251,800 papers published by PS Scholars and their graduate and postdoc advisers appearing in the Medline database.

First-stage/Dyad-level Covariates. As described in the methods section, we analyze two dependent variables, each at a different level of analysis. In the first stage, we model the occurrence of a match in a dataset of realized and counterfactual ties between Scholars *i* and eligible postdoc mentors *k.* In the second stage, we analyze the discrete time hazard that Scholar *i* files for a patent in year *t* as a function of whether the Scholar was exposed to a patenting postdoc adviser.

We run the dyad regression in a dataset with all 489 actual adviser-advisee matches, along with many counterfactual matches. We create the counterfactuals by pairing each Scholar in the year that s/he began postdoc training with every adviser who mentored a Scholar in that year. For instance, in the year 1990, 25 individuals who later received a Pew or Searle Scholar Award started their postdocs, and these individuals joined the labs of 23 distinct postdoc advisers (two advisers, Douglas Melton and Charles Zuker, each mentored two future PS Scholars that year.) For this year, we create a dyad-level dataset consisting of the 25 actual matches and the 550 potential matches that did not occur.

There are two reasons to define the risk set of counterfactual dyads by creating hypothetical pairings with other, active mentors in a given year. First, this definition of the risk set insures that all potential postdoc mentors are actively engaged in advising in the year in which a graduating Scholar searches for a position. Second, as the descriptive statistics will indicate, the postdoc advisers to PS Scholars are highly accomplished scientists. This implies that the appropriate set of potential advisers for these individuals is not the average academic biomedical scientist chosen at random; it is the elite

[11] For non-responders, we exhaustively searched public databases to reconstruct career histories. No Scholars were dropped due to a non-response from our CV request.
[12] We collect all issued patents through 2007. Both Scholar and adviser names were matched to the USPTO on a case-wise basis to correct for numerous misspellings in the database.

members of the profession. By restricting counterfactual matches to other PS Scholar mentors, we believe we create a representative sample of the members of Scholars' actual choice sets.

Building on the findings from the oral histories, we assess whether scientific interest, geography, status and commercial interests influence matching in mentor-trainee dyads. However, the time of matching occurs before Scholars establish a track record of independent research, which is necessary to compile the bibliometric data that is the basis for many of the covariates. To circumvent this problem, we measure detailed characteristics of Scholars' graduate school advisers, which we then assign to Scholars themselves. The idea is that graduate school advisers have a meaningful impact on the development trajectories of the students they train, and therefore Ph.D. advisers' characteristics proxy for the scientific trajectories of their students. Specifically, we measure the level of scientific similarity between a given Scholar *i*'s Ph.D. adviser in the year the Scholar earns his/her doctorate, and all potential postdoc advisers in the dataset in that year. We also generate two measures of the similarity/dissimilarity between Scholars' graduate advisers and potential postdoc advisers in the commercial orientation of research. Our operative assumption is that the scientific expertise of graduate advisers and their stance on the commercializing of academic science will correlate with those of their students.

Graduate/postdoc adviser scientific similarity. To assess scientific similarity in mentor-trainee dyads, we use Medical Subject Heading (MeSH) article keywords. MeSH headings are expert-curated keywords comprising the National Library of Medicine's "controlled vocabulary thesaurus." In 2008, there were ~25,000 keywords to index journal articles in Medline. Given all actual graduate advisers' and all potential postdoc advisers' publications, we generate for each dyad in each year *t* a count of the number of overlapping, unique MeSH keywords, which we denominate by the sum of the two advisers' total MeSH headings. This quantity—the proportion of common scientific keywords in each graduate-postdoc adviser dyad—is a symmetric measure of scientific similarity. To allow for a flexible specification of scientific proximity in the regressions, we then generate four dummy variables corresponding to each quartile of the distribution of scientific overlap. We anticipate that a Scholar is more likely to match with a postdoc adviser when his/her graduate adviser works in the same scientific area(s) as does the potential postoc mentor.

Graduate/postdoc adviser commercial similarity. We construct two measures of Scholars' similarities in commercial science to eligible postdoc advisers. First, for each graduate and postdoc adviser, we create an indicator equal to one if the adviser was listed as an inventor on one or more patents *prior to the year* that the Scholar transitions from the graduate to the postdoc adviser's lab. For all potential Scholar-postdoc adviser matches, we then create three dummy variables: graduate and potential postdoc adviser both hold patents; Ph.D. adviser patents but eligible postdoc adviser does not; and potential postdoc adviser patents but graduate adviser does not. The omitted category is that neither adviser patents. If we find a statistically significant coefficient on any of these patenting similarity covariates, it would indicate assortative (or disassortative) matching on commercial inclination. Statistically insignificant coefficients would support our claim that matching does not occur based on commercial interests.

In a second measure of compatibility in commercial interests, we use MeSH keywords to account for the underlying "patentability" of each scientist's research. The idea behind this measure is that scientists who choose to work in particularly patentable fields of research are more likely to be oriented toward commercial science. Specifically, we adopt the approach in Stuart and Ding (2006) to identify the time-varying, inherent patentability of each MeSH keyword. We collected all keywords used in the papers of the 9,000 academic life scientists with the highest NIH grant totals (excluding PS Scholars). We then matched these scientists to the inventor rosters on all US patents and identified all scientist-years in which members of this set had patented. MeSH keywords associated with either patenting or non-patenting scientists were then assigned a weight proportional to their frequency of occurrence in the patenting sample relative to their overall occurrence. A higher weight indicates that a given MeSH keyword is more prevalently used in the articles of patenting scientists than in those of non-patenters.

We apply these weights to the keywords on all articles of graduate and postdoc advisers in all years prior to the current one to construct a time-changing variable, research patentability, which is the average patentability of each scientist's keyword vector prior to each year. We then convert this to three indicator variables: graduate and potential postdoc adviser both in the top quartile of research patentability; graduate adviser is top quartile but potential postdoc adviser is not; and potential postdoc adviser is top quartile but graduate adviser is not. The omitted category is that neither adviser is in the top quartile. Once again, if we find statistically significant effects on any of these indicator variables, it would indicate assortative matching on commercial inclinations. Statistically insignificant coefficients would support our claim that Scholars do not match to postdoc advisers on the basis of commercial focus of their respective scientific trajectories.

Scholar/postdoc adviser geographic proximity. To account for the role of geography in the formation of matches, we construct an array of measures of the spatial proximity of Scholars and advisers. Two dummies indicate the relative location of a postdoc adviser vis-à-vis a Scholar's graduate school program. One indicator =1 when the Scholar and an eligible postdoc adviser are at the same university, and a second =1 when the Scholar and potential postdoc adviser are located in the same state.

Next, we have coded the state of the undergraduate institution of each Scholar who completes secondary education in the US. We then create an indicator variable =1 if a potential postdoc adviser is located in the same state as the Scholar's undergraduate institution. Obviously, this variable only captures variation within the subpopulation of scholars with a baccalaureate degree from a US university. Finally, we generate two covariates that gauge commonality in birth country. The first variable =1 when a Scholar and an eligible adviser are born in the same, *non-US* country. For comparative purposes (and because the oral histories lead us to suspect that matching on birth country will be stronger for those born outside the US), we construct a similar covariate indicating that the US is the common birth country. As described in the methods section, undergraduate institution/adviser location match and same, non-US birth country are the two exclusion restrictions in the Heckman regressions. We expect both covariates to influence the likelihood of matching but to be exogenous with respect to Scholars' later-career probability of patenting.

Graduate/postdoc adviser status similarity. The oral histories show that a number of Scholars sought high status advisers. We account for most of the status-based matching through the construction of the risk set in the dyadic dataset; because the counterfactual matches are exclusively formed between a Scholar's graduate adviser and all of the actual advisers of PS Scholars in a given year, there are only high status postdoc advisers in the risk set for potential matches. However, to capture any residual status matching in the data, we include a polynomial function of publication differences that will account for any such effect.

Additional Controls. We coded the gender of all Scholars and postdoc advisers from CVs and websites. We include a female =1 indicator in the second stage patenting regression, and we also add to the first-stage matching equation dummies designating that the Scholar and potential adviser are the same gender, and both are female. Because the norms regarding academic entrepreneurship have changed between 1980 and 2000 (Owen-Smith and Powell 2001b), we anticipate temporal effects. All regressions therefore include cohort indicator variables.[13] As proximity to clinical practice may increase the probability of academic entrepreneurship, we include a MD/PhD = 1 indictor in the second stage regression (Stuart and Ding 2006).

VI. Results

We begin with a brief overview of the individuals in the dataset. The median PS Scholar received his award in 1991. He is male and holds a PhD in biology. He began his doctoral studies in the early 1980s and received his doctorate in 1986. Between 1986 and 1991, he trained in one postdoctoral laboratory for five years. Because they begin their assistant professorships in different years, the PS Scholars in the dataset are "at risk" of patenting for different periods of time. The modal Scholar is observed for 19.4 years and 35% file for one or more patents before the data are censored in 2007.

---

[13] We also tracked the year-by-year employment of each PS Scholar to create an extensive list of controls for employer characteristics, including university-level patenting and NIH grant totals.

Table 2 presents summary statistics on the graduate and postdoc advisers of PS Scholars. The table plainly illustrates the achievements of this elite group.[14] Almost half of the graduate advisers are members of the US National Academy of Sciences (NAS), with significant representation of Howard Hughes Medical Institute (HHMI) members and a few Nobel Laureates. These membership tallies increase for postdoc advisers. Amazingly, more than 1 in 8 postdoc advisers has won a Nobel Prize by year 2008. A significant proportion of advisers have also engaged the world of commerce, as proxied by patenting.

Upon closer inspection, advisers who train multiple Scholars clearly are among the most prominent scientists of their generation (Table 3). Prolific advisers are all members of the NAS, with an increased representation of Nobel Laureates and HHMI members. Lastly, we note that the scientific foci of these individuals span many sub-disciplines. Although all practice in areas of inquiry that inform the biopharmaceutical industry, active engagement in translational research is not a prerequisite to being selected as a postdoc adviser.

Multivariate results: The pairing process. Table 4 presents the determinants of matches between Scholars and postdoctoral advisers. These specifications are estimated using a probit regression at the adviser/Scholar level of analysis (12,775 pairs, of which 12,286 are counterfactual).[15] The specification in column (1) includes all controls and the measures of alignment in commercial interests between graduate and potential postdoc advisers. Consistent with the oral histories, the regressions fail to uncover any evidence of matching on commercial interest, whether assessed by graduate and postdoctoral advisers' patenting histories, or by the patentability of research. Specifically, patenting graduate advisers are no more likely to send their students to patenting postdoc advisers than they are to non-patenting ones; nor are advisees of graduate mentors

[14] Adviser statistics are presented at the Scholar-adviser level. Advisers who train multiple Scholars are counted multiple times.

[15] Person-level variables (such as postdoc adviser's publication or patent count) are excluded from the regressions because they would have a negligible effect on the estimates. In a dyad-level model with year dummies to absorb across-period differences in the ratio of actual-to-counterfactual observations, node-level covariates only will be meaningfully identified to the extent that some actors are involved in more than one dyad in a given year. For Scholars, this is impossible by definition; all Scholars match to a single postdoc adviser. Postdoc and graduate advisers do sometimes mentor two eventual Scholars in a single year, but we account for this effect directly in the regressions (coefficient not reported).

in the top quartile of the research patentability distribution any more or less likely to preferentially join the labs of postdoc advisers who have conducted patentable research. When combined with findings from the oral histories, we conclude that Scholars and postdoc advisers do not match on orientations toward commercial science.

In column (2) we add the covariates that assess common scientific interests between Scholars' graduate and eligible postdoc advisers. As described, we include a flexible specification of indicator variables designating the three bottom quartiles of scientific overlap. Again consistent with the oral histories, the effects on the measures of scientific proximity are strong and highly statistically significant. Specifically, compared to a potential pairing in which a Scholar's graduate and would-be postdoc advisers are in the top quartile of the distribution of overlaps in scientific keywords, the matches in the bottom quartile of the overlap distribution are 93% less likely to occur. This finding indicates that graduate advisers are much more likely to send their PhD students to the laboratories of scientifically similar postdoc mentors.

The results of spatial geography parallel those of scientific proximity in consistency with the oral histories. In column (3) we find strong evidence of geographic sorting, with actual pairings more likely to involve a postdoctoral adviser from the Scholar's Ph.D.-granting institution. Similarly, net of the propensity to remain at their current universities, Scholars are more likely to match to mentors at other universities within the same state. These results persist in column (4), which includes the most comprehensive set of covariates; this specification is the one used to create the weights in the IPEW analysis reported below.

Finally, recall that although 15% of the oral histories explicitly cite the status of a potential postdoc adviser as a consideration in the search for a mentor, the sampling methodology (as confirmed in Tables 2 and 3) limits the risk set to prominent advisers. Nonetheless, in each of the matching regressions we include the sum and difference in publication counts for the graduate and postdoctoral adviser, as well as the square and cube of these variables. We do not report their coefficients because we failed to uncover any systematic pattern of matching on relative publication counts (beyond the control that is inherent in the definition of the risk set).

Table 5 provides evidence pertaining to the exclusion restrictions for the Heckman selection correction. The baseline specification is column (4) in Table 4. (We do not report the coefficients corresponding to the commercial variables because they are statistically insignificant.) We separately analyze the determinants of pairing for Scholars who come from outside the US and for those scholars with a US-based undergraduate institution. In column (1), we find that among the 121 foreign-born scholars, there is a greater propensity to pair with a postdoctoral adviser who hails from the same country. To the extent that homophilious preferences based on national origin are orthogonal to Scholars' commercial leanings, this result can be used as an instrument to disentangle mentors' influence from selection effects. In column (2), we find that among US scholars, there is a propensity to choose a postdoctoral lab located in the same state as one's undergraduate institution. Moreover, as can be seen in column (3), this finding persists even after excluding 56 Scholars who earned their undergraduate degrees in the state of California, which has been an important locale in the birth and development of the biotechnology industry. We will assume that this pattern of geographic attachment is uncorrelated with residual commercial dispositions, and we will use this variable to identify the causal effect of advisor patenting in the subsample of US scholars.

IPEW Results. Table 6 reports results of postdoc adviser patenting on Scholars’ propensity to patent using inverse probability of exposure weights. Observations are Scholar-years in which the Scholar holds a professorship and the specification is a discrete-time hazard of first patenting event. The variable of central interest is an indicator =1 if the Scholar’s postdoc mentor had patented before the Scholar finished training. In column (1), we present the "naive" estimates that do not include exposure weights to adjust for the matching process. The coefficient implies that patenting behavior is indeed subject to adviser “imprinting”; the hazard of patenting is 69% higher among scholars whose postdoc advisers are themselves patenters. Column (2) inversely weights each observation based on the fitted probabilities from column (4) in Table 4 to perform IPEW estimation. As explained in the methods section, under unconfoundedness, inversely weighting Scholar observations by the probability of pairing with mentors effectively creates a pseudo-population of Scholars in which the dyad-level observables no longer predict mentor assignment, but the causal association between adviser patenting and Scholar behavior is the same as in the original population.

Unexpectedly, the magnitude of the postdoc adviser patenting coefficient in the IPEW results (column 2) is more than two-thirds larger than the naive estimate. This seems surprising given that we have already empirically shown that commercial interests—at least to the extent that they are captured by observable covariates—do not influence the matching process. Why, then, might the coefficient on adviser patenting increase in the IPEW regressions? Effectively, the weights inflate the importance of Scholars with "unlikely" mentors, given observables. In turn, each observation's weight is most influenced by the covariates that have the greatest effect on the probability of a Scholar-adviser match, and in both the oral histories and the dyad regressions, scientific proximity between graduate and postdoctoral mentors' research interests is the dominant predictor of pairing. Thus, the larger effect of mentor imprinting in the IPEW estimates likely is driven by up-weighting the contribution of Scholars with postdoctoral mentors whose research is a significant departure from Scholars' specializations in graduate school.

We verify this conjecture in column (3). In this specification, weights are computed using the fitted probabilities from column (3) in Table 4, which omits the measures of shared scientific interests. When we recalibrate the weights this way, the magnitude of the IPEW estimate is much reduced, and only slightly larger than the "naive" estimate in column (2). The presence of this shift has a substantive interpretation: it indicates that Scholars who are field switchers (in the sense that their postdoc advisers differ in scientific focus from their graduate advisers) appear to be more susceptible to the influence of their postdoctoral mentors. Or, stated differently, Scholars with less-well defined scientific interests upon completion of their PhDs are more likely to adopt the commercial orientation of their postdoctoral advisers.

Heckman Sample Selection Results. As described, we have two exclusion restrictions to implement the Heckman procedure. The first, shared national background between Scholar and adviser, is most relevant for foreign-born Scholars. The second variable, the Scholar's undergraduate and potential postdoc advisers' institutions are in the same state, is most relevant for the subsample of US-born Scholars. As a result, we perform separate analyses on these two subsamples.

Results are in Table 7. The estimation sample for the second-stage regressions in the Heckman procedure is just the 2007 cross-section,[16] and the specification is a Probit with sample selection (Van de Ven and Van Praag 1981). Columns (1), (2), and (3) ignore the prior mentor selection stage and report naive estimates for the overall, US, and foreign-born samples respectively. The imprinting effect of adviser patenting is statistically significant in all cases. Columns (4) and (5) report the adjusted results using the Heckman selection correction. In both subsamples, this does not shift significantly the magnitude of the effect of mentor patenting, though the coefficient is only statistically significant at the 10% level in the sample of scholars with US undergraduate degrees.[17] In fact, consistent with our overarching claim that matching is only partially deliberate, likelihood ratio tests indicate that the estimates of ρ, the correlation between the error terms in the selection and outcomes equations, is not statistically different from zero in either Column (4) or (5). In other words, the Heckman results indicate that the selection process in which Scholars match to mentors can be safely ignored in the analysis of the probability that Scholars patent later in their careers.

Robustness Checks. As emphasized in the methodology section, the difficulty in establishing causality in our setting is that advisee-adviser matching is purposeful. To address this issue, the two statistical techniques we have employed rely on different assumptions. IPEW estimation hinges on unconfoundedness and the sample selection method depends on the validity of the exclusion restrictions. It is reassuring that the two techniques yield qualitatively similar results, but to further buttress the causal interpretation of the effect of adviser imprinting on Scholars' incidence of later-career patenting, we conduct four robustness checks.

First, we undertake a form of a falsification test—we examine whether adviser patenting influences other career outcomes. Second, we test the sensitivity of the IPEW estimate to

---

[16] It is not possible to estimate the outcome equation in pooled cross sections as we do in the IPEW regressions.
[17] In general, one would not expect the Heckman sample selection estimates and IPEW estimates to be identical because they produce different measures of a treatment effect. Under unconfoundedness, IPEW identifies the average treatment effect. In contrast, instrumental variables estimators identify local average treatment effects; that is, an effect relevant for the observations whose behavior changes because of the instruments.

assumptions about the composition of the risk set in the matching equation. Third, we examine the relative propensities of patenting versus non-patenting Scholars to continue along the scientific trajectories of their postdoc advisers. Finally, we investigate whether adviser patenting *after* the Scholar departs from the adviser's lab influences the likelihood of Scholar patenting.

Beginning with the falsification tests, we ask whether adviser patenting influences three career outcomes at the Scholar level: publication and citation rates, and NIH grants (for brevity, we report results only for Scholars' publications, but we also find null results for citation and grant totals). The motivation for these analyses is that *if* postdoc adviser patenting affects career outcomes that are unrelated to commercial activities, we would worry that mentor patenting in fact captures some unobserved dimension of Scholar talent that makes scientists more likely to succeed, whether in the commercial or open science spheres. Columns (4) and (5) of Table 6 report, respectively, naïve and IPEW estimates from quasi-maximum likelihood (QML) Poisson regressions of Scholars' annual publication rates. These results indicate that there is absolutely no effect of adviser patenting on unrelated measures of scientific achievement.

Next, we examine the sensitivity of adviser patenting to changes in the construction of the counterfactual dyads in the first-stage analysis. The coefficients in Table 6 are based on a risk set of counterfactual matches to other postdoc advisers who were active mentors in the year the Scholar transitioned to a postdoctoral fellowship. Here, we expand the set of counterfactual matches. First, we construct pairings between Scholars in year t and all postdoc advisers in either the current, preceding or subsequent year (i.e., we define the potential postdoc adviser dyads using a three-year moving window centered on the Scholar's graduation year). This results in 36,010 counterfactual dyads. Second, we further expand the set of potential adviser matches in year t to include any adviser who previously mentored one or more PS Scholars. This results in 95,251 counterfactual matches. We then re-estimated the IPEW-adjusted effect of adviser patenting in these two datasets and found that the coefficient varied only slightly from that in Table 6, Column (2).[18] Thus, within the tolerance

[18] We also re-estimated the baseline IPEW model in Table 6, column (3) after trimming observations in the highest and lowest 5% of the IPEW weight distribution. This attenuates the IPEW-induced increase in the postdoc adviser patenting coefficient relative to the naïve estimate, with no decrease in statistical significance.

we can explore without collecting a great deal of additional data, the results are insensitive to alternative methods of constructing the risk set of non-occurring dyads.

Third, our findings show that exposure to a patenting postdoc adviser significantly increases a Scholar's subsequent propensity to patent. Some readers still may worry that this propensity merely reflects the adoption by the Scholar of the focus of an adviser's research, but not the social transmission of advisers' stance toward patenting.[19] To further address this alternative interpretation, we examine whether Scholars who exhibit similar commercialization behaviors to their postdoc advisers are demonstrably *more* similar to their advisers' scientific trajectories than Scholars who deviate from past mentors' behavior with respect to patenting. We generated the MeSH keyword overlap (our measure of scientific proximity) between postdoc advisers' publications at the time the Scholar departed from their laboratories and Scholars' subsequent publication stocks at the 5th, 10th, and 15th years of their independent careers. The idea is to compare the relative scientific proximity of former postdocs who adopt their advisers' stance on patenting to those who deviate from it. Specifically, are trainees of patenting advisers who themselves patent later in their careers more scientifically proximate to their advisers than trainees of patenting advisers who do not patent, and therefore depart from their adviser's behavior? Conversely, are trainees of *non*-patenting advisers who do not patent more scientifically proximate to their advisers than trainees of non-patenting advisers who do patent? If the findings are driven by the differential transmission of advisers' research interests, we would expect to see *less* keyword overlap between those Scholars who deviate from their postdoc advisers patenting behavior than among those whose future actions conform to those of their advisers. This would suggest a scientific explanation for the core finding, rather than a sociological one.

Representative data for 15 years after the Scholar completed his postdoc are presented in box and whiskers plots in Figure 1. We report the distribution of scientific similarity scores between postdoc advisers and Scholars broken out by whether or not the adviser was a patenter and whether the Scholar becomes a patenter. The informative comparisons are between the two distributions

---

[19] Recall that the Scholar patenting regressions in Table 6 already address this concern by directly controlling for the flow and stock of the patentability of each Scholar's research.

within adviser type; that is, are patenting trainees of patenting advisers more scientifically similar to them than are non-patenting advisees? We see no evidence for this in Figure 1 or in any formal comparisons of distributions we have examined. In other words, the "inheritability" of scientific focus is constant across pairs in which advisees do/don't adopt the patenting practices of their advisers.

The fourth robustness test also addresses the question of whether the effect of patenting advisers represents a true social influence, versus just a transmission of advisers' scientific focus. In this analysis (for brevity, coefficients are not reported), we limit the sample to postdocs who trained under advisers who had yet to patent prior to the time the postdoc left their labs. In the regressions of Scholar patenting in this restricted sample, we then include a time-changing dummy =1 if a Scholar's adviser patents *after* the Scholar had already departed from his lab. Reassuringly, this dummy is statistically insignificant in the regressions of Scholar patenting. If adviser patenting after postdoc departure had an effect, it would indicate that patenting is transmitted even without direct exposure to advisers' behavior, which would be cause for concern that unobserved scientific factors drive the result. The fact that first-hand exposure is required buttresses our claim that the core result is a causal influence of advisers on trainees.

VII. Discussion and Conclusion

We present two central findings. First, in scientists' autobiographical accounts and in a novel database, we show that Pew and Searle Scholars match to their postdoctoral advisers based on two primary factors: scientific compatibility and geography. Second, the primary result is that postdoctoral advisers' patenting behavior is "imprinted" on their trainees. Through the use of inverse probability of exposure-weighted estimations and an instrumental variables approach, as well as from knowledge of the matching process gained from scientists' oral histories, we demonstrate that the social influence of advisers on trainees is real; it is not endogenous to trainee-adviser matching dynamics. Moreover, the social influence effect is statistically large. To put the magnitude into context, we find (Table 6) that female scientists in academe are much less likely than men to patent. However, if a female postdoc by chance matches with a patenting adviser, the adviser's estimated influence on her probability of later-career patenting almost fully offsets the very large, negative effect of gender.

On one hand, the findings from the second stage analysis are to be expected; few will be surprised that the attitudes of the most important mentors in a foundational period of advisees' professional development matter, especially in a training period as lengthy as a postdoctoral fellowship. However, the interesting finding is not the lasting influence of the mentor, but that the consequent is unanticipated by the antecedent. Specifically, advisees are significantly influenced by advisers on a dimension that appears not to have entered their thoughts at the time they initiated the search for a mentor. The development of scientists' commercial orientations does not appear to follow predetermined career objectives that direct the search for an adviser. Rather, the end result seems to arise by chance; Scholars conduct a local search for an adviser in bordered scientific and geographic spaces. Whether or not an adviser is a commercialist is orthogonal to the search process, but it is highly relevant to the development of the advisee's career. In this way, chance exposures to patenting advisers appear to induce transition points in individuals' careers.

Generalizing this observation, the paper's central theotetical claim is that when actors connect based on a small set of attributes *X*, it is often the case that some set of additional characteristics *Z*, which was never considered when a choice was made to develop a relationship, nevertheless become socially transmitted. Perhaps one reason why sociologists have not investigated this dynamic is that in the contexts that interest the discipline, the exposure effect (the *Z*s) generally is correlated with the matching variables (the *X*s). For example, political attitudes (*Z*) may be correlated with religious beliefs (*X*), which may in turn drive associations. We show that this is not the case in our setting—matching factors (science, geography) are independent of the exposure effect (adviser patenting). But even in the general case of correlation between *X* and *Z*, as long as the list of factors in the pairing model is exhaustive, such correlation does not invalidate the estimates in IPEW regressions. Nor is it problematic to omit from the list of *X*s predictors of pairing that are statistically independent of the exposure effect—such omission will, at most, inflate the standard errors. However, the assumption of unconfoundedness will fail, and estimates will be biased, if attributes that are relevant for pairing and are correlated with the second stage outcome are omitted from the specification of the pairing model.

The fact that matching is only partially deliberate clearly opens avenues for the unforeseen transmission of attitudes and behaviors. In the majority of cases, unanticipated exposures are likely to be of insignificant consequence. All of us can call to mind instances in which an associate shared some unexpected point of view that had nothing to do with how our relationship with that individual came into being—but was also inconsequential for how we behave or think. In certain circumstances, however, the attributes to which we are unexpectedly exposed can matter. Particularly when these exposures take place in the context of relationships with long durations or ones in which there are notable status or experience differentials between partners, chance exposures can fundamentally change individuals' points of view. In long running, asymmetric relationships (such as those between protégés and postdoc advisers), the length of interaction provides ample opportunity for the standard pathways of influence to take hold. And when these experiences occur in the process of professional development as we have seen in this study, they may result in turning points that reorient actors' career trajectories.

Finally, we believe that the argument in the paper will generalize to other settings. As one example, the literature on board interlocks suggests that the nomination committees of corporate boards select new directors based on patterns of inter-corporate resource dependencies and spatial proximity. But the individual directors who are elected may have preconceived views of corporate governance issues or strategic imperatives that were not criteria in the selection process. Even if these views were independent of the match, the opinions of new directors may, through their influence on other members of the board, shape subsequent, board-level decisions. In fact, we believe that partially deliberate matching may permeate the sociology of the economy, as many social relationships in market contexts arise from a limited set of economic imperatives, but subsequently become pipes for social influence.

Table 1: Summary of Oral Histories- Determinants for Postdoc Adviser Choice

Category (%/N)	Representative Quotes
1. Science (95%/59)	
-Extension of Prior Knowledge	[Emerson; pg 149-150] "Well, I wanted to expand on my graduate work in that I wanted to add the element of chromatin structure to the study of gene regulation... Gary Felsenfeld was the king of chromatin".
-Moving Away from Base	[Greenberg; pg 44-45] "Basically, at Harvard, we had really no exposure to plant research. It was really the chance reading of an article from Ausubel's lab where they talked about this plant, Arabidopsis, that I work on now… if one wanted to study adaptation to the environment… one could do it in a plant, and then it would get around all the ethical problems that I had with killing a lot of animals."
-Moving Towards Frontier	[Horowitz; pg 73] "…after my work on murine leukemia viruses, I wanted to work on oncogenes because it became really apparent while I was doing my graduate work that that's where the action was for most human cancers."
2. Geography (53%/33)	
-Personal Constraints	[Horowitz; pg 73] "… my wife, Barbara, decided she wanted to work for him [Bernard Fields at Harvard]. She applied and was pretty much quickly accepted so it then became necessary for me to find a postdoc in Boston."
-Personal Preferences	[Julius, pg 203] "… by the time my time was up there, I was ready to leave. Berkeley can be a very sort of uniform-seeming community… I was ready to see what living on the East Coast was like again…"
3. Advisor Status (15%/9)	[Hirano, pg 29] "Tim Mitchison was another young assistant professor at that moment. But he did a very famous discovery when he was in graduate school. And he was very young, but he was already famous. And it was clear he was one of the brightest cell biologists at his age… "
4. Interpersonal Rapport (12%/7)	[Jardetzky; pg 58] "And he [Don Wiley] was an incredible person, and just sitting with him for an hour, I realized that that was where I wanted to be. I just wanted to be working with somebody like that who had that kind of insight, that kind of drive, that kind of creative energy. He was a really impressive guy."
5. Commercial Opportunities (0%/0)	N/A

Table 2: Descriptive Statistics

Panel A: Scholar Characteristics (N = 489)

	Mean	Std. Dev	Min.	Max.
Female	.233	.423	0	1
US	.753	.432	0	1
PhD	.867	.340	0	1
Highest degree year	1986	4.88	1973	1998
Year of First Academic Appt.	1990	5.20	1977	2000
Member-NAS	.061	.240	0	1
HHMI	.161	.368	0	1
Nobel Laureate	.002	.045	0	1
Stock of Publications (2007)	70.71	49.02	11	381
Is a Patentor (2007)	.360	.480	0	1
# of patents (2007)	1.37	4.13	0	57
Patentability Stock (2007)	.516	.574	0	4.49

Note: 489 unique Scholars

Panel B: Graduate Adviser Characteristics (N = 489)

	Mean	Std. Dev	Min.	Max.
Female	.067	.251	0	1
Member-NAS	.411	.493	0	1
Member-HHMI	.123	.328	0	1
Nobel Laureate	.063	.244	0	1
At end of Scholar training				
Stock of Publications	88.61	81.40	1	513
Is a Patentor	.194	.396	0	1
Avg # of patents	.620	2.54	0	45
Patentability Stock	.151	.246	0	2.48

Note: 415 unique graduate advisers

Panel C: Postdoc Adviser Characteristics (N = 489)

	Mean	Std. Dev	Min.	Max.
Female	.061	.240	0	1
Member-NAS	.601	.490	0	1
Member-HHMI	.321	.467	0	1
Nobel Laureate	.135	.342	0	1
At end of Scholar training				
Stock of Publications	108.42	100.89	0	729
Is a Patentor	.438	.497	0	1
Avg # of patents	2.08	5.41	0	73
Patentability Stock	.433	.508	0	3.13

Note: 333 unique postdoc advisers

Table 3: Characteristics of Prolific Advisers

Graduate Advisers with Four or More Trainees

# of Trainees	Name	Nobel	HHMI	NAS	Research Program
4	Eric Davidson	No	No	Yes	Sea Urchins Development
4	Robert Baldwin	No	No	Yes	Protein Folding
4	Gunter Blobel	Yes	Yes	Yes	Yeast Nuclear Transport
5	David Botstein	No	No	Yes	Yeast Genetics
5	Philip Sharp	Yes	No	Yes	RNA Splicing
5	Jack Szostak	No	Yes	Yes	Yeast Chromosomes

Postdoc Advisers with Five or More Trainees

# of Trainees	Name	Nobel	HHMI	NAS	Research Program
5	Ronald Davis	No	No	Yes	Molecular Immunology
5	Harold E. Varmus	Yes	No	Yes	Viral Oncology
6	Marc Kirschner	No	No	Yes	Developmental Biology
6	Stanley Falkow	No	No	Yes	Microbial Pathogenesis
6	Robert Tjian	No	Yes	Yes	Biochemistry of Transcription
6	H. Robert Horvitz	Yes	Yes	Yes	C. elegans Development
6	Randy Schekman	No	Yes	Yes	Yeast Vesicle Transport
8	Thomas Cech	Yes	Yes	Yes	Transcription and Splicing
8	Gerald Rubin	No	Yes	Yes	Fruitfly Genetics
8	Thomas Maniatis	No	No	Yes	MolecularGene Regulation
9	Richard Axel	Yes	Yes	Yes	Molecular Olfaction
11	David Baltimore	Yes	No	Yes	Molecular Virology

Table 4: Determinants of Pairing by Scholars and potential Postdoc Advisers (Probit).

	(1)	(2)	(3)	(4)
Dataset	All Scholars	All Scholars	All Scholars	All Scholars
Grad/Pdoc Keyword Overlap 0-25th pctile		-1.173**		-1.182**
		(0.079)		(0.078)
Grad/Pdoc Keyword Overlap 25-50th pctile		-0.746**		-0.751**
		(0.062)		(0.062)
Grad/Pdoc Keyword Overlap 50-75th pctile		-0.481**		-0.482**
		(0.055)		(0.055)
Grad & Pdoc at same university			0.328**	0.368**
			(0.090)	(0.097)
Grad & Pdoc in same state, different university			0.140*	0.144*
			(0.063)	(0.067)
Scholar & Pdoc are the same gender	0.011	0.032	0.012	0.035
	(-0.049)	(-0.052)	(-0.050)	(-0.052)
Scholar & Pdoc are both female	0.051	-0.014	0.044	-0.020
	(-0.135)	(-0.135)	(-0.138)	(-0.139)
Grad & Pdoc both patent	0.114	0.094	0.099	0.075
	(-0.085)	(-0.091)	(-0.086)	(-0.093)
ONLY Grad patents	-0.074	-0.090	-0.084	-0.101
	(-0.074)	(-0.077)	(-0.074)	(-0.077)
ONLY Pdoc patents	-0.003	-0.048	-0.005	-0.049
	(-0.035)	(-0.039)	(-0.036)	(-0.040)
Grad & Pdoc research patentability -top quartile	-0.043	-0.156	-0.044	-0.161
	(-0.096)	(-0.099)	(-0.097)	(-0.100)
ONLY Grad research patentability -top quartile	-0.033	-0.022	-0.036	-0.025
	(-0.069)	(-0.073)	(-0.070)	(-0.073)
ONLY Pdoc research patentability -top quartile	0.014	-0.036	0.018	-0.034
	(-0.044)	(-0.048)	(-0.044)	(-0.048)
Constant	-1.340**	-0.603**	-1.337**	-0.605*
	(0.183)	(0.241)	(0.183)	(0.241)
Log likelihood	-2042	-1888	-2034	-1879
Observations	12775	12775	12775	12775
# of scholars	489	489	489	489
# of pdoc advisers	333	333	333	333

Note: Estimates are displayed as raw coefficients. Errors are clustered at the Pdoc adviser level. All models include Scholar-cohort dummies, and an indicator variable if the Grad or Pdoc adviser had sent/received multiple students within that cohort-year. All models also include the sum and absolute difference of Grad and Pdoc adviser publication counts, as well as their square and cube, respectively. These variables are not shown. For Grad/Pdoc Keyword Overlap, the excluded quartile (75-100th pctile) includes those dyads that are most scientifically similar. Robust standard errors in parentheses below; + significant at 10%; * significant at 5%; ** significant at 1%.

Table 5: Determinants of Pairing by Scholars and potential Postdoc Advisers, by Subsample (Probit).

Dataset	(1) Only foreign Scholars	(2) Only US Scholars	(3) Only US Scholars- no CA
Scholar & Pdoc born in same foreign country	0.803**		
	(0.215)		
undergrad & Pdoc university in same state		0.212**	0.263**
		(0.076)	(0.098)
Grad/Pdoc Keyword Overlap 0-25th pctile	-1.322**	-1.177**	-1.292**
	(0.166)	(0.096)	(0.109)
Grad/Pdoc Keyword Overlap 25-50th pctile	-0.977**	-0.696**	-0.689**
	(0.139)	(0.070)	(0.076)
Grad/Pdoc Keyword Overlap 50-75th pctile	-0.486**	-0.481**	-0.473**
	(0.102)	(0.063)	(0.072)
Grad & Pdoc at same university	0.450**	0.285*	0.272*
	(0.163)	(0.119)	(0.124)
Grad & Pdoc in same state, different university	0.348*	0.039	0.099
	(0.139)	(-0.082)	(-0.089)
Scholar & Pdoc are the same gender	0.002	0.044	0.025
	(-0.113)	(-0.058)	(-0.062)
Scholar & Pdoc are both female	0.151	-0.143	-0.099
	(-0.286)	(-0.189)	(-0.221)
Constant	-0.689	-0.563+	-0.208
	(-0.577)	(0.316)	(-0.411)
Log likelihood	-441	-1421	-1198
Observations	3097	9678	8201
# of scholars	121	368	312
# of pdoc advisers	333	333	333

Note: Estimates are displayed as raw coefficients. Errors are clustered at the Pdoc adviser level. All models include Scholar-cohort dummies, and an indicator variable if the Grad or Pdoc adviser had sent/received multiple students within that cohort-year. All models also include the sum and absolute difference of Grad and Pdoc adviser publication counts, as well as their square and cube, respectively. These variables are not shown. For Grad/Pdoc Keyword Overlap, the excluded quartile (75-100th pctile) includes those dyads that are most scientifically similar. Robust standard errors in parentheses below; + significant at 10%; * significant at 5%; ** significant at 1%.

Table 6: Impact of Postdoc Adviser Patenting on Scholar Patenting and Publication Outcomes (IPEW)

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Scholar First Patenting Event			Scholar Publication Count	
Model	Discrete-Time Hazard Rate-Logit			QML-Poisson	
IPE Weights	NO	YES	YES-w/o Scientific Distance	NO	YES
Pdoc was a patentor	0.526**	0.832**	0.548**	-0.021	0.005
	(0.172)	(0.207)	(0.179)	-0.043	-0.047
Research patentability flow, no lag	4.167**	4.688+	4.051*	4.192**	4.560**
	(1.483)	(2.791)	(1.626)	(0.477)	(0.462)
Research patentability stock, 1-year lag	0.914*	0.212	0.942+	0.440**	0.429**
	(0.464)	(0.673)	(0.512)	(0.076)	(0.068)
Female	-0.706**	-0.916**	-0.771**	-0.136**	-0.090
	(0.243)	(0.303)	(0.257)	(0.046)	(0.056)
MD/PhD	0.502*	0.707*	0.524*	0.185**	0.174**
	(0.226)	(0.286)	(0.227)	(0.052)	(0.056)
Log(University NIH $)	-0.235*	-0.160	-0.325**	-0.011	0.005
	(0.095)	(-0.139)	(0.101)	(-0.025)	(-0.026)
Log(University patents)	0.097	0.109	0.134+	0.016	-0.025
	(-0.072)	(-0.099)	(0.079)	(-0.020)	(-0.030)
Constant	-0.034	-1.302	1.413	1.437**	1.349**
	(-1.664)	(-2.295)	(-1.762)	(0.429)	(0.407)
Log-pseudolikelihood	-676	-672	-676	-14349	-906497
Observations	5040	5040	5040	6587	6587
# of scholars	489	489	489	489	489
# of postdoc advisers	333	333	333	333	333

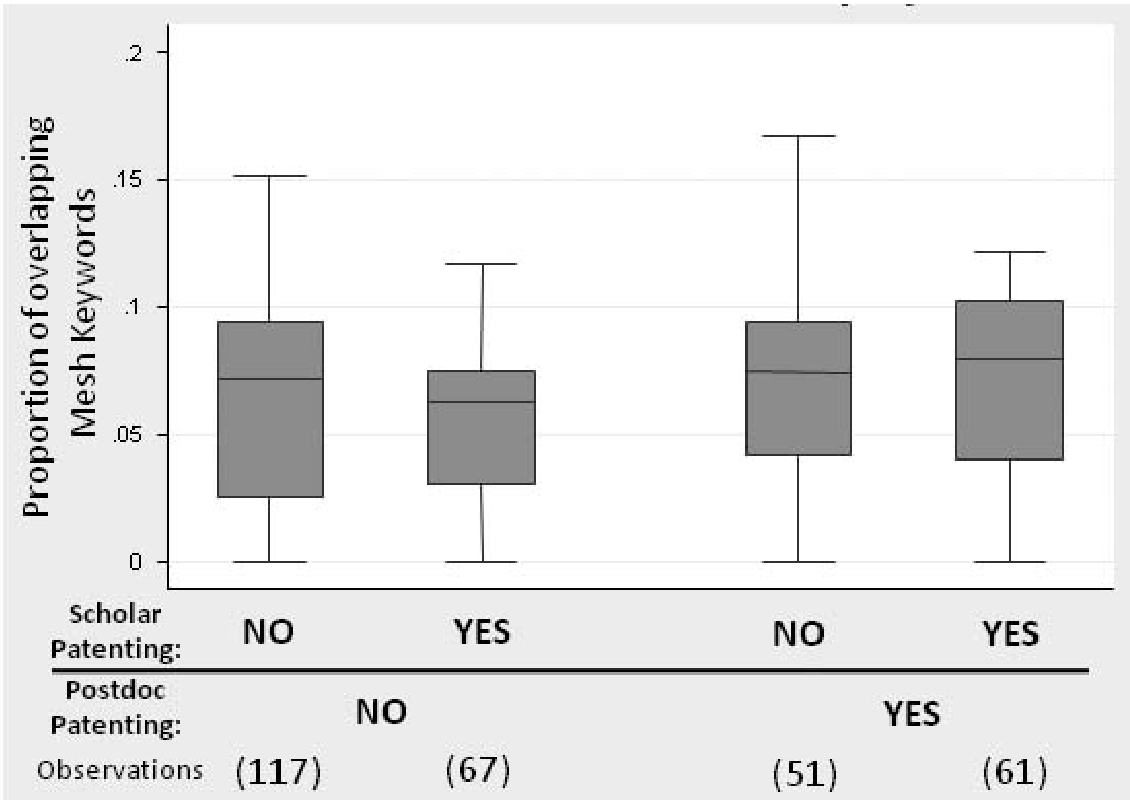
Note: Estimates are displayed as raw coefficients. All models include, but do not show, year dummies, and (every other year) cohort dummies. Robust standard errors, clustered by Scholar, are in parentheses below; + significant at 10%; * significant at 5%; ** significant at 1%

Table 7: Cross-Sectional (Heckman) Probit Model with Selection on Scholar Patenting.

	(1)	(2)	(3)	(4)	(5)
Model	Probit	Probit	Probit	Heckprob	Heckprob
Dataset	ALL	US	Foreign	US	Foreign
Exclusion Restrictions	None	None	None	Undergrad & Pdoc university in same state	Both Scholar & Pdoc same foreign country
Pdoc was a patentor	0.363**	0.338*	0.553*	0.312+	0.524*
	(0.130)	(0.157)	(0.256)	(0.172)	(0.231)
Research patentability stock	0.454**	0.341**	1.007**	0.346*	0.988*
	(0.117)	(0.129)	(0.269)	(0.168)	(0.421)
Female	-0.569**	-0.446*	-0.859**	-0.436+	-0.641+
	(0.153)	(0.177)	(0.326)	(0.226)	(0.380)
MD/PhD	0.318+	0.464*	-0.295	0.387	-0.246
	(0.184)	(0.212)	(-0.413)	(-0.246)	(-0.353)
Log(University NIH $)	-0.067	-0.038	-0.244+	-0.039	-0.273*
	(-0.053)	(-0.052)	(0.127)	(-0.055)	(0.138)
Log(University patents)	0.000*	0.000*	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grad/Pdoc Keyword Overlap 0-25th pctile				0.871	0.227
				(-0.530)	(-0.833)
Grad/Pdoc Keyword Overlap 25-50th pctile				0.225	0.066
				(-0.437)	(-0.687)
Grad/Pdoc Keyword Overlap 50-75th pctile				0.417+	0.227
				(0.240)	(-0.342)
graduate and post-doc advisor same university				0.022	-1.467**
				(-0.349)	(0.514)
Constant	1.265	0.761	3.789	0.960	5.459*
	(-1.065)	(-1.111)	(-2.306)	(-1.377)	(2.182)
atanh(ρ)				-0.442	-0.626
				(-0.719)	(-0.719)
Log-pseudolikelihood	-268	-200	-62	-1648	-509
Observations (Scholars)	489	368	121	368	121
# of postdoc advisers	333	262	102	333	333
Potential Dyads				9678	3097

Models restricted to US Scholars include 11 cohort dummies (including 1 excluded). Models restricted to foreign Scholars include 4 cohort dummies (including 1 excluded). Both Models 4 & 5 include a count of each Scholar's potential Pdoc dyads. These variables, as well as the exclusion restrictions, are not shown. Robust standard errors, clustered by Scholar, are in parentheses below; + significant at 10%; * significant at 5%; ** significant at 1%

Figure 1: Scholar and Postdoc Adviser Scientific Proximity-By Adviser Patenting



Note: Representative box and whiskers plot for the proportion of postdoc adviser (year *t*) and Scholar (year *t* + 15) MeSH keywork overlap; *t* is the last year of Scholar training. 296 Scholars are presented.

References

Abbott, Andrew. 1997. "On the Concept of Turning Point." Pp. 85-106 in *Comparative Social Research*, edited by Grete Brochmann, Fredrik Engelstad, Ragnvald Kalleberg, Arnlaug Leira, and Lars Mjoset. Greenwich, CT: JAI Press Inc.

—. 2001. *Time matters: On theory and method*: University of Chicago Press.

Abbott, Andrew, and Alexandra Hrycak. 1990. "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96:144-185.

Allen, Thomas J. 1977. *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the Research and Development Organization*. Cambridge, MA: MIT Press.

Angrist, Joshua D., Guido. W. Imbens, and Donald. B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91:444-455.

Audretsch, David. B., and Paula. E. Stephan. 1996. "Company-Scientist Locational Links: The Case of Biotechnology." *The American Economic Review* 86:641-652.

Azoulay, Pierre, Waverly Ding, and Toby E. Stuart. 2007. "The Determinants of Faculty Patenting Behavior: Demographics or Opportunities?" *Journal of Economic Behavior & Organization* 63:599-623.

Blau, Peter M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York NY: Free Press

Bossard, James H. S. 1932. "Residential Propinquity as a Factor in Marriage Selection." *American Journal of Sociology* 38:219-224.

Burt, Ronald S. 1992. *Structural Holes: The Social Structure of Competition.* Cambridge, MA: Harvard University Press

Dehejia, Rajeev. H., and Sadek Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84:151-161.

Elder, Glen H. 1985. *Life Course Dynamics: Trajectories and Transitions, 1968-1980*. Ithaca, NY: Cornell University Press.

Etzkowitz, Henry. 1998. "The Norms of Entrepreneurial Science: Cognitive Effects of the New University-Industry Linkages." *Research Policy* 27:823-833.

Evans, James A. 2004. "Sharing the Harvest? The Uncertain Fruits of Public/Private Collaboration in Plant Biotechnology." Stanford University.

Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86:1015-1035.

Festinger, Leon, Stanley Schachter, and Kurt W. Back. 1950. *Social Pressures in Informal Groups*. New York, NY: Harper.

Hagstrom, Warren O. 1965. *The Scientific Community*. London: Basic Books Inc.

Heckman, James J. 1979. "Sample Selection Bias as a Specifiction Error" *Econometrica* 47 153-161

Imbens, Guido W. 2000. "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika* 87:706-710.

Jackson, Matthew O. 2008. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Kandel, Denise B. 1978. "Homophily, Selection, and Socialization in Adolescent Friendships." *American Journal of Sociology* 84:427.

Knorr-Cetina, Karin D. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
Kono, Clifford, Donald Palmer, Rroger Friedland, and Matthew Zafonte. 1998. "Lost in Space: The Geography of Corporate Interlocking Directorates." *American Journal of Sociology* 103:863-911.
Lazarsfeld, Paul F. and Robert K. Merton, 1954. "Friendship as a Social Process: A Substantive and Methodological Analysis." Pp18-66 in *Freedom and Control in Society*, edited by Morroe Berger, Theodore Abel, and Charles Page. New York: Van Nostrand
Long, J. Scott, Paul D. Allison, and Robert McGinnis. 1979. "Entrance into the Academic Career." *American Sociological Review* 44:816-830.
McPherson, J. Miller, Pamela A. Popielarz, and Sonja Drobnic. 1992. "Social Networks and Organizational Dynamics." *American Sociological Review* 57:153-170.
McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415-444.
Mouw, Ted. 2003. "Social Capital and Finding a Job: Do Contacts Matter?" *American Sociological Review* 68:868-898.
Murray, Fiona E. 2009. "The Oncomouse that Roared: Resistance & Accomodation to Patenting in Academic Science." *American Journal of Sociology* Forthcoming.
Nerad, Maresi, and Joseph Cerny. 1999. "Postdoctoral Patterns, Career Advancement, and Problems." *Science* 285:1533.
Newcomb, Theodore Meade. 1961. *The Acquaintance Process*. Oxford, UK: Holt, Rinehart and Winston
Owen-Smith, Jason, and Walter W. Powell. 2001a. "To Patent or Not: Faculty Decisions and and Institutional Success at Technology Transfer." *Journal of Technology Transfer* 26:99-114.
—. 2004. "Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community." *Organization Science* 15:5-21.
Owen-Smith, Jason, and Walter. W. Powell. 2001b. "Careers and Contradictions: Faculty Responses to the Transformation of Knowledge and Its Uses in the Life Sciences." *Research in the Sociology of Work* 10:109-140.
Podolny, Joel, and James Baron. 1997. "Resources and Relationships: Social Networks and Mobility in the Workplace." *American Sociological Review* 62:673-693.
Reagens, Ray E., Ezra Zuckerman, and Bill McEvily. 2007. "On Firmer Ground: The Collaborative Team as Strategic Research Site for Verifying Network-Based Social-Capital Hypotheses." Pp. 147-182 in *The Missing Links*, edited by James E Rauch. New York: Russell Sage Foundation.
Robins, James M., Miguel A Hernán, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology*, 11: 550-60.
Rosenbaum, Paul R., and David B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41-55.
Sorenson, Olav, and Toby E. Stuart. 2001. "Syndication Networks and the Spatial Distribution of Venture Capital Investments." *American Journal of Sociology* 106:1546-1588.
Stephan, Paula. E., and Jennifer Ma. 2005. "The Increased Frequency and Duration of the Postdoctorate Career Stage." *American Economic Review* 95:71-75.

Stovel, Katherine, Michael Savage, and Peter Bearman. 1996. "Ascription into Achievement: Models of Career Systems at Lloyds Bank, 1890-1970." *American Journal of Sociology*:358-399.

Stuart, Toby E, and Waverly Ding. 2006. "When do Scientists Become Entrepreneurs? The Social Structural Antecedents of Commercial Activity in the Academic Life Sciences." *American Journal of Sociology* 112:97-114.

Stuart, Toby E. , and Olav Sorenson. 2009. "Strategic Networks and Entrepreneurial Ventures." *Strategic Entrepreneurship Journal* 1:211-227.

Van de Ven, Wynand P.M.M., and Bernard M.S. Van Praag. 1981. "The Demand for Deductibles in Private Health Insurance." *Journal of Econometrics* 17:229-252.

Van den Bulte, Christophe, and Gary L. Lilien. 2001. "Medical Innovation Revisited: Social Contagion versus Marketing Effort." *American Journal of Sociology* 106:1409-1435.

Winship, Christopher, and Stephen L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25:659-706.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of US Biotechnology Enterprises." *American Economic Review* 88:290-306.

Zuckerman, Harriet. 1977. *Scientific Elite: Nobel Laureates in the United States*. New York, NY: Free Press.