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

Pierre Azoulay
MIT & NBER
Sloan School of Management
100 Main Street, E62-487
Cambridge, MA 02142

Christopher C. Liu
University of Toronto
Rotman School of Management
105 St. George Street
Toronto, ON M5S 3E6

Toby E. Stuart
University of California
Haas School of Business
2200 Piedmont Ave. – F455
Berkeley, CA 94720

January 14, 2016

Abstract

Actors and associates often match 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 contacts considered when they first chose to pair. This implies that some apparent social influences (those tied to the rationales for forming a relationship) are endogenous to the matching process, while others (those that are incidental to the formation of the relationship) may be exogenous, thus enabling *causal* estimation of social influences on exogenous-to-the-match attributes. We label as "partially deliberate" social matching that occurs on a small set of actor attributes, and we present empirical methods for identifying influence effects when relationships follow this generative logic. In a dataset tracking the training and professional activities of academic biomedical scientists, we show that two factors, geography and scientific focus, are very important to the match between scientists-in-training and their postdoctoral mentors. Although they do not match on it, young scientists then adopt their advisers' orientations toward commercial science as evidenced by adviser-to-advisee transmission of patenting behavior. We demonstrate this in two-stage models that account for the endogeneity of matching, first using inverse probability of treatment weights (a "selection on observables" approach), then using Heckman-style sample selection estimators (a "selection on unobservables" approach). We also draw on qualitative accounts of how candidates matched to their advisers, which are recorded in oral histories taken from the scientists in the data. Overall, we present a theory and methods to establish evidence of social influence when tie formation is only partially deliberate.

Keywords: matching, social influence, social networks, postdoctoral training, patents, sociology of science.

*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 `tstuart@haas.berkeley.edu`. All errors are our own.

1 Introduction

People select partners in relationships for many reasons. They match based on similarities in sociodemographic characteristics, spatial locations, interests, and referrals from trusted associates. Substantial bodies of theory suggest that many 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 will differ across particular pairs, settings, and types of relationships, the actual ties that emerge from the vast set of possibilities often do so because individuals are complementary on a small set of meaningful characteristics.

Though we often match on salient attributes, in totality people possess very many characteristics. This prevalent aspect of social matching creates randomness in the social influence process and, therefore, offers a strategic research site. If people deliberately match on a subset of carefully considered (or merely convenient) dimensions, we are then exposed to unanticipated social influences when we encounter the views and tastes that never entered our calculus when we chose a particular interaction. In other words, if two people connect because they are compatible on some attributes X, it is likely that additional characteristics Z, which were not evaluated when the choice was made to develop the relationship, are then transmitted from one contact to the other. For example, if two people strike up a companionship because they work at the same establishment, share a love of opera, and have similar-aged children (Xs), one member of the pair may later convince the second to volunteer at the local animal shelter, or of the health benefits of exercise (Zs). Importantly, the two socially transmitted behaviors, volunteer work and exercise, did not contribute to the original formation of the match. In a general framework, if each actor is construed as a vector of discrete attributes, the fact that matching takes place on sub-segments of these attribute vectors rather than their entireties implies an opportunity to identify causal social influences across non-matched-on attributes.[1]

In an empirical illustration of this theory, we examine postdoctoral candidates and faculty advisers. We study Pew and Searle Scholars (hereafter, "Scholars"), a set of prominent, young, academic life scientists. Exploiting an extensive quantitative database and a qualita-

[1]Readers may be concerned at this point that there is correlation across the elements of individual attribute vectors, which will confound estimates of social influence. We will present technical details below, but 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 (e.g., ethnicity, age, and education) and are then exposed to unanticipated social influences in attitudes that may be correlated with age attributes, such as political views or preferences for certain types of leisure activities.

tively rich oral history archive, we find that two factors (the Xs) often spur matches between postdoc candidates and their advisers: compatible scientific interests and geographic location. In a second-stage analysis, we then show that whether a Scholar's postdoctoral adviser was a patenter (the exposure effect, Z) during or before (but not after) the time the Scholar joined the adviser's lab has a large effect on the advisee's likelihood of patenting later in his or her career. By estimating this effect in two-stage models that account for the endogeneity of adviser-advisee pairings, and by relying on the oral histories, we show that postdoctoral candidates *do not* appear to consider their advisers' patenting behavior when establishing the match. Therefore, the evidence suggests that the transmission of patenting behavior truly is a causal social influence, rather than stemming from common commercial interests or other latent similarities that underlie the initial candidate-adviser match.

The primary contributions of the paper are a theory, method, and illustration of partially deliberate social matching. A further contribution concerns the substantive findings of the specific empirical case, which joins a burgeoning literature on sociological questions at the interface of academic and commercial science (e.g., Etzkowitz 1998; Evans 2010; Murray 2010; Owen-Smith and Powell 2001a; Owen-Smith and Powell 2004; Stuart and Ding 2006). The cornerstone of the theoretical assertion is that people are complex and multi-dimensional. Therefore, at the time of inception of a new relationship, we cannot know *all* of a would-be counterpart's attributes, attitudes, tastes, and preferences. Moreover, when individuals are faced with even a relatively small number of features in a choice context, the literature on the psychology of choice demonstrates that subjects employ strategies to eliminate attributes from consideration to reduce the complexity of the decision. By extension, when we establish a new relationship, people are unlikely to know—and match on—the full complement of an associate's political views, musical tastes, cultural preferences, skills and knowledge, friendships, attitudes, and so forth. Moreover, given the immense number of potential connections that might occur and the short time horizons over which many relationships gestate, individuals often follow a boundedly rational, satisficing approach in the initial selection of associates. This theoretical assertion directly implies an empirical approach to identifying causal social influences in settings in which such "partially deliberate" matching occurs and data on pairings and outcomes are available.

To preview the empirical analysis in the paper, we have a four-pronged approach to the challenge of showing that Scholars match to advisers on a few primary attributes (geography, scientific interest), but a secondary dimension (adviser patenting) that *does not* shape the likelihood of a match subsequently does influence Scholar behavior. First, we code 62 comprehensive "oral histories" of Pew Scholars and find that *none* of the 62 transcripts mention would-be advisers' commercial activities as a factor in their selection of postdoctoral fellowship. Conversely, the oral histories consistently describe scientific topic and geography as drivers of the matches that form. Second, we estimate dyad-level matching regressions between Scholars and postdoc advisers. These regressions both show that Scholar-adviser pairing is independent of advisers' commercial activities, and they strongly reinforce the qualitative evidence that geography and scientific focus are core to the matching calculus. 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. Because the assumptions of propensity score estimators could be violated in these data, we implement a fourth 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 that allow us to recover estimates of advisers' influence on Scholars' behavior even in the presence of residual selection on unobserved factors.

2 Social Influence in Partially Deliberate Matches

The study of social influence is of primary concern in sociology. It is foundational in the social networks literature (e.g., Marsden 1981; Friedkin 1993), in social psychology (e.g., Hogg and Abrams 2002), and in the literature on diffusion (e.g., Coleman, Katz, and Menzel 1957). It is also relevant in many other areas of sociological inquiry, including socialization processes (Stouffer 1949; Merton 1957) and institutional theory (DiMaggio and Powell 1983).

Because social influence is a theoretical edifice in multiple subfields of the discipline, a growing chorus of authors has critiqued the empirical literature for its inattention to the challenge of establishing evidence of *causal* social influences in observational data (e.g., Winship and Morgan 1999; Van den Bulte and Lilien 2001; Mouw 2003; Reagans, Zuckerman, and McEvily 2007; Stuart and Sorenson 2009; Aral, Muchnik, and Sundararajan 2009; Shalizi and Thomas 2010). The principle empirical challenge arises because the outcomes that interest researchers often are endogenous to the factors that spur the formation of social ties. Indeed, early contributors to the literature noted that the mutual selection of like-minded individuals into relationships mimics a social influence process in observational data, even when no such process occurs (Kandel 1978; Newcomb 1961). Thus, the mechanisms of social matching, most notably homophily, often masquerade as social influence. Moreover, this is not merely an academic distinction: the two processes, social influence and homophily, generally have different implications for policies and strategies to influence social outcomes. Whether one is interested in the diffusion of health behaviors, the spread of agricultural technologies, or the commercialization of academic science, alternative mechanisms have different implications for the dynamics of influence and the interventions that may impact them. Therefore, it is important that we attempt to distinguish between these processes, and to attempt to pinpoint the precise mechanisms of social influence.

Contrasting the mechanisms of homophily and social influence belies the fact of a temporal separation in the two. For the most part, homophily concerns how ties come to be; social influence often occurs after relationships are in place.[2] Beginning with the former—the inception of new relationships—a rich body of work has illuminated the guiding hand of social similarity in social interaction. Lazarsfeld and Merton (1954) and Blau (1977) develop the social foundations that lead us to anticipate homophilous interaction. In current work on the subject, it is very well documented that social relationships cluster among categorically similar individuals who share a core set of ascribed attributes and status characteristics (McPherson, Popielarz, and Drobnic 1992; McPherson, Smith-Lovin, and Cook 2001), although the relative contribution of preference-based or opportunity-based motives for social similarity continues to animate empirical work.

The literature on homophily closely aligns with research on the spatial geography of relationships. Because chance interactions are more likely between spatially co-located actors and the cost of maintaining relationships is higher at a distance, social interaction depends on geographic nearness. In relationships as varied as marriages (Bossard 1932), workplace collaborations (Allen 1977), board directorships (Kono et al. 1998), and investment syndicates (Sorenson and Stuart 2001), proximity is a main determinant of the likelihood of interaction.[3]

[2]Of course, this is a simplification. Social influence also occurs in the absence of direct relationships, as when prominent individuals' shape the opinions and views of others.

[3]In addition to a shared interest in the origins of social relationships, the intertwining of work on homophily and propinquity stems from the fact that geographic proximity and social similarity co-occur. If neighborhoods are racially segregated, for example, then as long as residential propinquity has some impact on the friendships that form in society, it will appear as if people have a preference for within-race interactions. (Of course, it may be precisely a preference to affiliate with same-race companions that gives rise to racial segregation of neighborhoods, but this need not be true.) Stated differently, if we randomly choose a pair of geographically proximate individuals and compare them to a randomly chosen dyad in which

A common denominator across these lines of work is that relationships do not emerge randomly from the vast set of feasible ties. The non-randomness in social relationships, whether driven by peoples' proclivity toward homophilous interaction or any other mechanism of attachment, is one source of the empirical difficulty in distinguishing a true social influence from its possible correlates. Namely, regularities in the formation of social relationships—what we might think of as the rules that generate the network data we record as observers of the social world—can lead to the appearance of social influence even when it does not occur. This challenge is marked.

In addressing this issue, we begin with an assumption. Our premise is that actors often strike up matches based on a small set of important characteristics for the relationship at hand. This is particularly likely to be true of casual ties in which a premium is placed on convenience (Feld 1981), but we believe it to be true even when actors seek significant, instrumental relationships. In most instances of social matching, we contend that actors do not optimize partner selection over a high-dimensional attribute space. Rather than arising from an algorithmic search across a vast sample of potential partners' individuating characteristics, matching typically occurs on just a few factors that matter most to would-be connections. People often halt their search for a partner when they find one who is judged to be suitable enough. If this is a fair characterization of the process of relationship inception in some contexts, the central assertion in the paper is this: *when actors form relationships based on characteristics X but do not match on other attributes Z, we can study social transmission along an attribute Z in a context that may be relatively untainted by the process leading to the assignment of actors to matches.*

Why is it reasonable to postulate that people form relationships according to the logic of partially deliberate matching? Much of the answer lies in the psychology literature that evaluates how people make complex choices. As a decision maker confronts a greater number of options or as the information about available options increases, people respond in two ways: they entertain fewer of the feasible choices, and they process a reduced fraction of the total information available (Hauser and Wernerfelt 1990; Iyengar and Lepper 1999). Thus, as complexity increases, subjects invoke simplifying strategies (cf. Payne 1982; Payne, Bettman, and Johnson 1993). Timmermans (1993), for example, compares the decision-making strategies people use when presented with three, six, or nine choices. Even in a stylized setting that greatly simplifies many real-world decision-contexts, Timmermans finds

members are located at a significant distance, the former pair is more likely to exhibit social similarities than the latter pair.

that the fraction of participants who employed "elimination strategies" increased monotonically with the size of the choice set. In contrast, the use of available information decreased with the number of options presented.

The finding of a reduction in the amount of available information that is considered as the complexity of a choice increases resembles Simon's (1947) notion of satisficing: boundedly rational individuals typically search until they identify a satisfactory choice, rather than maximize over complicated decision spaces. The gist of the literature is that people simply stop in decision contexts when they achieve a "good enough" result. Moreover, there is also evidence that when faced with a choice that is difficult, individuals place a premium on making a decision that is compatible with a line of reasoning that grounds a compelling narrative (Shafir, Simonson, and Tversky 1993). In describing how people choose between complex but equivalent alternatives, Shafir et al. (1993) write "... people seem to be following a choice mechanism that is easy to explain and justify: choosing according to the most important dimension provides a better reason for choice." As decision makers, we appear to be concerned with the post-choice narrative we can formulate, which is how we justify difficult choices to ourselves and others.

While psychologists have focused on how people make decisions, sociologists and economists often travel in the opposite direction: they often model observed choices to infer partner selection strategies. This returns us to the literature on homophily. In this body of work, it is striking the extent to which basic dimensions of geographic proximity and the cornerstone elements of sociodemographic similarity drive the inception of new relationships, across multiple contexts. In a nice illustration, Marmaros and Sacerdote (2006) investigate the formation of friendships among incoming freshmen on a college campus. Their study is novel and empirically persuasive because it models friendships in a newly forming network. They find that micro-geography and race are the dominant factors in how people select new friends. A few other measures of similarity, such as whether the two members of a potential friendship are both varsity athletes, also have positive (but lesser) effects on matches in this study. Notably, Marmaros and Sacerdote conclude that individuals do not search across the campus to widen their pool of potential friends; rather, they quickly settle into relationships with others who were randomly assigned to the same physical space. Similar results pepper the sociology literature: spatial proximity, race, gender, socioeconomic class, age, and a few other factors appear to be pervasive determinants of the relationships we choose (cf. McPherson, Smith-Lovin, and Cook 2001).

To recapitulate, the experimental literature in psychology establishes that individuals employ information reduction and choice elimination strategies to manage the enormous complexity of challenging decision contexts. When confronting such decisions, we are known to satisfice: because it is costly and infeasible to make assessments over all potentially relevant considerations, individuals default to choices based on a more limited and higher-priority set of characteristics. We believe that the parallel to these findings at the macro level is the sociological work on a relatively narrow range of sociodemographic and physical-world similarities that appear to underlie the creation of many new associations. These arguments underlie our proposition that people match on a subset of their full attribute vectors.

3 Context: Adviser-Advisee Pairings for Postdoctoral Fellows

Social scientists have had a long-running interest in postdoctoral fellows. Because of their prevalence, postdocs are integral to the fabric of laboratory life (Knorr-Cetina 1999). The postdoc system also reinforces the status ordering in science. Not only are next-generation scientific leaders more likely to complete postdocs with the elite of the current generation, but from an adviser's standpoint, successfully placing 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 when young scientists engage in anticipatory socialization in preparation for 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 significant duration of the postdoctoral training period (Stephan and Ma 2005) and the direct interdependence of the work, apprentices are deeply exposed to the attitudes, behaviors, and styles of mentors—and they are likely to be highly susceptible to these influences.

In light of postdoctoral fellows' essential role in scientific production, one might expect formalized institutions to govern the matching between candidates and advisers. In reality, the market for postdocs is not orderly. There is no central clearinghouse to pair candidates to available positions. Indeed, the postdoc hiring process might be regarded as the antithesis of the highly structured National Resident Matching Program, which matches graduate medical residents to available positions on a single day.

3.1 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."[4] Unlike other accolades such as the Nobel Prize, 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 independent 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 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, but not all will choose to pursue this aspect of their work. The decision to patent a scientific discovery in this group is likely to be as influenced by scholarly priorities and values as it is by the commercial significance of the underlying research.

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, these texts provide detailed accounts of the rationales for Award-winners' career choices.

[4] Quoted from the Pew Scholars Program Description at `http://www.pewtrusts.org/`; accessed 4/27/14. These awards confer significant status to recipients, but the monetary component is generally insufficient to change the recipient's scientific research trajectory.

3.2 Oral Histories

The Pew Scholar Oral History and Archives Project has collected the life histories of more than 200 Pew Scholars. The expressed purpose of these histories is to record, "...the lives of scientists ... many of them explore issues related to the Scholars childhood, college experiences, time training in various labs, their time as a PI, and broader social, political, and cultural issues related to science."[5]

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 matches are fashioned around a limited set of dimensions.

We randomly chose 62 interview transcripts to read, 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, geography, adviser scientific status, and interpersonal rapport. Given the focus here, we then added a fifth category, commercial considerations, although none of the original five interviews expressed a preference for matching on this criterion.

A coder then read each transcript to identify the sections describing the graduate and postdoctoral period. For each transcript, the coder indicated if a given category was cited as a determinant for pursuing a particular 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 doctorate from UCLA and then became a postdoc at CalTech. In describing her search for a postdoc 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."*[6]

[5] `http://www.chemheritage.org/discover/collections/oral-histories`

[6] 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)

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

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 single, ubiquitous consideration in selecting an adviser was scientific focus. Only three Scholars did not mention scientific interest as a major factor in seeking a position in a particular mentor's laboratory, and these were due to exceptional circumstances.[7] 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 pivot scientific trajectories—the majority of Scholars hoped to build on areas of expertise they had developed during graduate school.

More than half of the Scholars singled out geography as a major factor in their search. In 19 cases (31%), Scholars reported that geography was a binding constraint. In these instances, family considerations, most often regarding a partner's career, limited a Scholar's search to a particular region. 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."*[8]

In another 14 cases (23%), individuals cited a strong personal preference 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... 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,*

[7] For example, one Pew Scholar was scheduled to train under David Baltimore. One month prior to the start of the fellowship, Baltimore accepted the presidency of Rockefeller University and moved from Boston to NYC. Baltimore then arranged for this particular Scholar to train under (fellow Nobel Prize winner) Phillip Sharp at MIT.

[8] 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)

> *David this and that. 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.'"*[9]

Across the interview transcripts, scientific interest and geographic considerations are the two criteria that are foremost in candidates' minds when they search for advisers. In a smaller proportion of cases, adviser status and interpersonal attraction were also decision criteria.[10] These results closely coincide with those of prior surveys of the motivations for postdoctoral adviser choice (Nerad and Cerny 1999).

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, to work alongside an adviser who has connections in industry, or any other form of engagement with commercial-sector entities—when choosing a postdoctoral adviser. 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 quite a few elements of chance contour the career experiences of Pew Scholars. Rather than working backward from well-defined career objectives to a search for an optimal match, the process individuals follow to find a postdoc mentor is one of local search in delimited scientific and geographic spaces, coupled with the intervention of chance encounters. While the search and matching process is not entirely random, neither does it seem to encompass a large number of reported dimensions. For this reason, we believe the matching process conforms to our notion of "partial deliberateness."

At this point, readers may be concerned that this is because interviewees considered it unsavory or counter-normative to discuss the commercial aspects of science. Given prevailing

[9] 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)

[10] 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 very select members in their consideration set to be germane to their decisions.

academic norms, it is possible that Scholars had—but were reluctant to share—commercial aspirations when choosing postdoc advisers. Although we cannot rule out any explanation for the lack of reference to commercial motivations in the postdoc matching process, a number of the oral histories did specifically address the subject of academic patenting. In one third of the histories, Scholars were directly asked for thoughts regarding their own patenting activities (if applicable) and 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 most cases, Scholars were explicitly positive about the scientific and professional benefits of patents. Therefore, we believe at least some Scholars would have discussed their commercial interests if they recalled them to be germane in the search for an adviser.

4 Sample, Data, and Quantitative Methods

4.1 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) until year 2000. All told, we began with 583 Scholars.[11] Individuals are captured in our sampling frame when they receive the Award. To conduct the analyses, however, we require information on both graduate school and postdoc advisers. We therefore search backward in time to identify all advisers for these 583 Scholars. Ultimately, this process reduced the analyzable sample to 489 Scholars; the remaining individuals were MDs who did not have identifiable graduate school advisers. These 489 Scholars apprenticed as postdocs in the laboratories of 333 unique advisers.

4.2 Methods

Estimating the causal effect of mentors' influence on Scholar career outcomes must address the basic selection problem that adviser "assignment" is non-random. Our specific concern

[11] 642 Scholarships had been awarded at the time of data collection. From this population, we dropped 57 individuals from disciplines that are peripheral to biomedicine, such as population biology and clinical psychology. The rate of patenting in the dropped group was similar to the retained sample, but because we rely on the PubMed database to construct many of the covariates, we limited the sample to Scholars for whom the vast majority of publications were indexed in PubMed. We also dropped one individual due to a precipitous retirement and another who succumbed to cancer within two years of receiving his Award.

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 will seek postdoc positions in the laboratories of like-minded advisers, and vice-versa. Matching on a taste for commercial science could produce a spurious association between postdoc adviser commercial propensity and Scholar career outcomes in any estimation approach that does not account for the endogeneity of the outcome to bases for forming matches. Therefore, standard statistical techniques, which assume that mentor assignment is exogenous, may not recover causal effects.

We contend that postdoc-adviser pairing is indeed deliberate, but only partially so because of the heavy influence of the 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 account for matching to estimate a causal effect of adviser influence. First, we use a variant of propensity score estimation, which is known as a "selection on observable" approach because it is valid only under the untestable assumption that the outcome of interest is independent of assignment to treatment conditional on observed factors. Second, non-random matching between Scholars' and advisers can be considered to be an instance of a sample selection problem because we witness actual matches but do not observe potential matches that did not—but could have—occured. 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 as a function of postdoc adviser behavior.

Selection on observables: Inverse Probability of Treatment Weights (IPTW). Consider a scenario in which each Scholar i ($I = 1, \ldots, N$) is assigned a mentor j from a pool of J potential mentors. One can think of mentor assignment as a multi-valued treatment $T \in 1, \ldots, J$ (*cf.* Imbens 2000). In the pre-assignment period, we measure X_i^k, a set of prognostic factors for assignment to a particular match. These prognostic factors will be dyad-level covariates that influence the likelihood that Scholar i pairs to adviser j. The outcome of interest y_i is then measured at a subsequent time. In our case, treatment occurs when a Scholar matches with an adviser who patents prior to or during the time the Scholar is a trainee in the mentor's lab. The outcome we study 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. In this framework, assignment may be counterfactual, i.e., $k \neq j$; the Scholar need not be paired with his/her own mentor. To reliably estimate the average treatment effect, we require matches to be *unconfounded*: Scholars-adviser pairs must be statistically independent of y_i^k conditional on observable factors X. The term "unconfoundedness" was coined by Rubin (1990) to refer to the situation in which conditioning on a fixed set of covariates removes all bias in comparisons between treated and control cases, thus allowing for a causal interpretation of the covariate-adjusted treatment effect. In other words, the uncofoundedness assumption states that treatment is conditionally random; given observed factors X, treatment is not confounded by unobserved covariates, which is to say that there are no omitted variables that affect both assignment to treatment and outcomes. Formally, we write the unconfoundedness assumption:

$$T \perp y_i^k \mid X \text{ for all } i \text{ and } k$$

In addition to the assumption that treatment condition is random within subpopulations defined by values of the covariates, we must also assume that, for all included values of the covariates, the likelihood of being matched to any particular mentor is positive. Formally, this assumption is known as common support. The intuition is that it is necessary to observe both treated and non-treated cases that correspond to particular values of X. The assumption of common support can be formally written as:

$$0 < Prob(y_i^k = 1|X = x) < 1$$

Intuitively, given a value $X = x$, it must be possible to estimate both $E[y_i|X_i = x, T_i = 1]$ and $E[y_i|X_i = x, T_i = 0]$, which we can do only if there are observations in both the treatment and control groups at each value of X.

Both of these assumptions are non-trivial. The unconfoundedness assumption is not testable and it places strong demands on the data generating process. We 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 assignment to treatment (Dehejia and Wahba 2002). In many samples, determinants of this nature are not available. However, we have chosen a study population for which we were able to carefully investigate and measure pre-treatment variables that we believe to be most likely to confound comparisons between units assigned to different treatment conditions. As a result, we believe that the unconfoundedness assumption provides a reasonable starting point in our context.

The common support assumption is testable. It implies that we should limit our comparisons to sets of values for which there is sufficient overlap in the match probabilities between actual and counterfactual matches (Barber, Murphy, and Verbitsky 2004). Below, we will provide graphical evidence that the region of common support is very wide in our specific case.

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:

$$E[y_i^k|Z_i, PATENT_k] = \beta_0 + \beta_1^{'} Z_i + \beta_2 PATENT_k \tag{1}$$

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 the assumption of unconfoundedness, β_2, the causal effect of adviser patenting, is identified and can be recovered by estimating:

$$E[y_i^j|Z_i, PATENT_j] = \beta_0 + \beta_1^{'} Z_i + \beta_2 PATENT_j \tag{2}$$

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 $(k; y^k)$. A second 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.

The implementation of this estimation technique is straightforward. Under unconfoundedness, selection bias can be removed by weighting the regression by

$$w_i = \frac{1}{Prob(T_i = j|X_i^j)} \tag{3}$$

The denominator of w_i is the conditional probability that a Scholar was assigned his or her actual mentor j. 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 the outcome variable is unchanged from the original population.[12] We refer to β_2 when

[12] We can now return to a previous point: if the unconfoundedness assumption of IPTW estimation is met, it poses no problem for causal influence if the social influence variables Z are correlated with the matching variables X. This is because in the pseudo-population of Scholars (i.e., Scholars weighted by the inverse probability of treatment), the Xs are uncorrelated with mentor assignment. Therefore, so too is any function of X, or the projection of some other variable Z on the vector X. This is true by the assumption of unconfoundedness.

equation (1) is weighted by w_i as the Inverse Probability of Treatment Weighted (IPTW) estimator of β_2.

We face a specific challenge in estimating the weights in the data. The treatments considered here are assignments to particular mentors. These are qualitatively distinct treatments that are devoid of any logical ordering. A natural approach would be to estimate the probability of assignment to each specific mentor in a multinomial logit or probit framework.[13] This is not feasible in our case, since the population of mentees and the population of potential mentors are of similar size (489 and 333 respectively).

As a result, we do not model the probability that a mentee matches with a specific mentor. Rather, we model the probability of pairing with his or her *own* mentor. The difference is subtle, but important. Concretely, we estimate a single probit regression that pools the observations corresponding to each actual matches ($n = 489$) with the observations corresponding to the counterfactual matches ($n = 12,286$):

$$Prob(T_i = k) = \alpha_0 + \alpha_1 X_i^k + \delta_t \tag{4}$$

where $Prob(T_i = k) = 1$ for actual Scholar-adviser matches and $= 0$ for all counterfactual pairs, X_i^k includes dyad-level covariates predicting matches between Scholars and advisers, and δ_t represents match year indicator variables. Of course, equation (4) is of substantive interest in its own right; it reveals correlates of postdoc-mentor pairings.

One issue with this modeling choice is that it fails to constrain the match probabilities *for a given mentee* to sum to one. Rather, it simply guarantees that the sum of match probabilities *for the entire mentee sample* will sum to one. Therefore, to construct weights for the second stage, we normalize the fitted probabilities that emerge from this specification by dividing them by the sum of probabilities for all matches (actual or counterfactual) for each mentee.[14] Formally:

$$w_i = \frac{\sum_{k \in J_i} Prob(T_i = k | X_i^k)}{Prob(T_i = j | X_i^j)} \tag{5}$$

where J_i is the set of potential postdoctoral advisers for Scholar i. IPTW estimation is very simple to implement, but the unconfoundedness assumption is a strong one, and its validity cannot be tested. As a result, we also utilize an alternative approach.

[13] This is the approach adopted by Huang et al. (2005), who first model the probability that an asthma patient will match to a given physician group, before asking whether this choice natters for health outcomes.

[14] The correlation between the normalized and "raw" weights is 0.99. Through visual inspection of Epanechnikov kernel densities for the two distributions, we have also verified that there are no worrisome differences in the upper tails of the normalized and raw IPT weight distributions. We conclude from this that the renormalization does not alter the character of the pseudo-population.

Selection on unobservables: Heckman selection correction. Although the oral histories suggest that commercial opportunities do not drive the choice of postdoctoral mentors, 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 IPTW estimates. A potential alternative to IPTW to estimate a causal social influence is to isolate quasi-random factors that shape the matching process, and to rely solely on this variation to estimate the effect of treatment. To implement this approach, we require instrumental variables—quantities 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 in our setting. The first is the proximity between Scholars' undergraduate institutions and the universities where they might become postdoctoral fellows. The logic for this instrument comes from the findings in the oral history: we anticipate that geography will drive postdoc matching in a manner that is independent of the propensity to patent. The second exclusion restriction is shared nationality between the Scholar and a potential mentor, *conditional on being born outside the US*. Here, we believe that common birth country and native language will promote mutual awareness and interest in matching. The relevance of these instruments ultimately is an empirical question, and we will provide below statistical evidence 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 commercial activities. We believe these assumptions to be plausible in this setting, and we will describe a number of robustness tests that bolster them.

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 US-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 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.

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 never came to be. Formally, we assume the existence of the underlying relationship:

$$y_i^k = \beta_0 + \beta_1' W_i^k + \beta_2 PATENT_k + \varepsilon_{ik} \tag{6}$$

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:

$$Prob(T_i = j) = \alpha_0 + \alpha_1 X_i^j + \delta_t + \eta_{ij} \tag{7}$$

where $Prob(T_i = j) = 1$ for realized matches between Scholars and advisers and $= 0$ for counterfactual matches, and 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 IPTW regressions [equation (4)], in order 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 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, it is well known that relying on functional form assumptions to estimate average treatment effects in the Heckman framework provides poor identification (LaLonde 1986). In our case, 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 (6). To implement the Heckman approach, we have adopted a parametric approach, that of Probit with sample selection (Van de Ven and Van Praag 1981). We also explore a more flexible, semiparametric approach (Newey, Powell, and Walker 1990; Gerfin 1996). Because our substantive conclusions are unaffected by the estimation technique, we limit a detailed exposition of the semi-nonparametric approach to an online Appendix (Part II).

There are two noteworthy differences between the IPTW and Heckman analyses. First, in contrast to IPTW, the Heckman framework does not require the assumption of unconfoundedness. It does, however, depend on the validity and relevance of the exclusion restrictions. The attractiveness of the latter approach is its ability to identify the causal effect of mentor imprinting even in the presence of residual selection based on unobservable influences. Second, the Heckman sample selection and IPTW approaches are unlikely to yield identical coefficient estimates because they produce different measures of a treatment effect. Under unconfoundedness, IPTW identifies the average treatment effect. In contrast, instrumental variables estimators identify the local average treatment effect; that is, an effect only relevant for the cases whose behavior changes because of the instruments.

4.3 Data Construction

Our analysis relies on four primary data sources. First, we requested CVs from all Scholars to identify dates of training periods, degrees, advisers, and undergraduate institutions.[15] 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 US patent office data.[16] 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 PubMed 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 pairs. 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

[15]For non-responders, we exhaustively searched public databases to reconstruct career histories. No Scholars were dropped due to a non-response to our CV request.

[16]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.

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 remarkably 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 comprises the elite members of the profession. By restricting the set of counterfactual matches to other active PS Scholar mentors, we believe we create a representative sample of the members of Scholars' actual choice sets. Likewise, we believe that the postdoc candidates in the sample are representative of the quality of the individuals who are legitimate contenders for positions in the labs of the elite mentors in the data. Moreover, as we will report below, this sample selection choice will meet the assumption of common support.

Building on the findings from the oral histories, we assess whether scientific interest, geography, social status and commercial interests influence matching in mentor-trainee dyads. The ideal approach would be to have direct measures of graduate students' scientific trajectories and commercial aspirations. Because we cannot survey Scholars at the time of matching, we instead use bibliometric data to proxy for scientific foci and commercial orientation. The challenge with this approach, however, is that at the time matching occurs graduate students have yet to establish a track record of independent research, which is what generates the bibliometric data. To address this problem, we instead 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 PhD advisers' characteristics proxy for the scientific trajectories of their students. Specifically, we measure the level of scientific similarity between a given Scholar's PhD 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.

Graduate/postdoc adviser scientific similarity. To assess the scientific similarity between focal Scholar i and potential Postdoc Mentor j, we use Medical Subject Heading (MeSH) keywords. MeSH headings are expert-curated keywords comprising the National Library of Medicine's controlled vocabulary thesaurus. There are approximately 25,000 keywords to index journal articles in PubMed. 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. Thus, we anticipate scientific similarity between actual graduate advisers and potential postdoc mentors will predict student exchange.

Graduate/postdoc adviser commercial similarity. Before describing the specific measures of commercial similarity between graduate students and potential postdoc advisers, it is important to restate a central assumption of our measurement strategy. We assume that, *if* graduate students give significant thought to commercial science, then it will be the case that students who complete their PhD studies under the supervision of commercially oriented graduate advisers will be more likely to themselves hold an interest in commercial science. Therefore, if commercial interests enter into the equation in the search for postdoc advisers, students who are matriculating from commercially oriented graduate mentors will be more likely to pair to postdoc advisers of like mind. Specifically, if commerce invades the matching process, we expect to observe greater proximity in the respective commercial orientations of graduate $\Rightarrow$ postdoc adviser pairs that mentor the same student, relative to pairs that do not exchange a student. If there is no indication of adviser matching on commercial science, we will take it as evidence that this dimension falls outside the matching calculus.

The evidence gleaned from the oral histories suggests that commercial science is not a significant factor in matching, and therefore (in our measurement strategy) we will not in fact observe its transmission from graduate advisers to graduate students. This begs an important question: why do we hypothesize that postdoc advisers' patenting behavior will transmit to mentees, but we simultaneously do not anticipate a similar social influence in graduate school? The answer lies in the differing roles of the graduate student and the postdoc in the scientific process. Graduate students and postdocs have broadly similar career objectives, which is to publish high-profile papers. However, the two types of personnel play different roles in the laboratory. Graduate students often have minimal laboratory experience. In our highly selective cohort, these students might be described as clever, but not yet wise. They lack the experience and craft necessary to choose and execute a research program, and are as-yet unfamiliar with the ecosystem of journals and reviewers.

By contrast, postdocs have much more experience and, in many respects, become seasoned scientists before they complete their training. In the majority of cases, middle-term and more senior postdocs have brought a number of research projects to fruition. Building on this experience, postdocs are able to be strategic with regards to the scientific trajectory of their career. Preparing for the transition to run their own laboratory is often front-of-mind for postdocs.

Given the differences between these two roles, we anticipate that graduate students are unlikely to be influenced by the commercial practices of their advisers, whereas postdocs may be. Why? First, in the hierarchical structure of academic laboratories (the largest of which employ dozens of students, postdocs, and technicians), postdocs spend much more time directly working with the lab head than do graduate students. Second, postdocs are mentored in many facets of scientific careers, whereas graduate students are more narrowly focused on conducting bench science. For example, postdocs must learn to handle managerial issues including laboratory personnel, grant raising, and, in the case of this paper, the possibility of engaging in entrepreneurial activities. Acquiring these "advanced" skills typically is not on the mind of most graduate students, who participate at a lower level of the organization. Finally, postdocs are much more likely to work with the PI to set the laboratory's research agenda. In so doing, they gain exposure to the problem space that drives the research agenda and, in more commercially oriented labs, they are likely to join conversations about patenting alongside the traditional scientific outputs of papers and conference presentations.

Turning to the covariates, we construct two measures of Scholars' graduate advisers' similarities in commercial science to eligible postdoc mentors. 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 applied for *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; PhD 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. Small and 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 followed by Stuart and Ding (2006) and Azoulay, Ding, and Stuart (2009) 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.[17]

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. We construct an array of measures of the spatial proximity of Scholars and advisers. Two dummies indicate the relative location

[17]We collect all issued patents through 2007. The names of all 9,000 scientists were matched to the USPTO and hand-checked to correct for numerous misspellings in the database. Further details on the construction of these keyword weights can be found in the online Appendix (Part I).

of a postdoc adviser vis-à-vis a Scholar's graduate school program. One indicates when the Scholar and an eligible postdoc adviser are at the same university, and a second indicates 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 equal to one 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 equals one 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 university/adviser location match and same, non-US birth country are the two exclusion restrictions in the Heckman-style analyses. We expect both covariates to influence the likelihood of matching but to be exogenous with respect to Scholars' later-career probability of patenting.[18]

Graduate/postdoc adviser status similarity. The oral histories show that a number of Scholars sought high-status advisers. We implicitly account for 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, only high-status postdoc advisers populate the risk set for potential matches. To capture any residual status matching in the data, however, we include a polynomial function of publication differences between graduate advisers and potential postdoc mentors.

Dependent variable. We match the patent output of the Scholars and their advisers to the records of the US Patent and Trademark Office (USPTO), and their publication output to PubMed, which is maintained by the National Library of Medicine. One must remember that the bulk of the output of the academics we study is in publications, rather than patents. Over 60 percent of the Scholars never apply for a patent, and the majority of those who do

[18]It is possible that there is a correlation between certain ethnicities and the propensity to patent. However, even in the presence of this correlation, the Heckman selection equation actually is a matching model in which we predict the likelihood that a would-be postdoc i matches to a potential mentor j. Therefore, the instrument is not the nationality or ethnicity of the candidate per se; rather, it is whether the candidate and the postdoc adviser *share the same national origin.*

patent have only one or two inventions to their credit. The primary dependent variable is the rate of patenting in the post-training careers of Scholars, as a function of the treatment effect of training under a commercially oriented postdoc adviser.

Additional controls. We coded the gender of all Scholars and postdoc advisers from CVs and websites. We include a female indicator in the second-stage patenting regressions, and we 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 commercializing academic science have changed between 1980 and 2000 (Owen-Smith and Powell 2001b), we anticipate temporal effects. All regressions therefore include cohort indicator variables.[19] As proximity to clinical practice may promote academic entrepreneurship, we include an indicator for joint degree holders, $MD/PhD = 1$ (Stuart and Ding 2006).

5 Results

We begin with a description of the individuals in the dataset. The median 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 a five-year postdoc. Because they begin their assistant professorships in different years, the 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 percent file for one or more patents before the data are right censored.

Table 2 presents summary statistics for graduate and postdoc advisers. The table illustrates the achievements of this group.[20] 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 were Nobel Prize winners by year 2008. A significant proportion of advisers also have patented. On 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.

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

[20] Adviser statistics are presented at the Scholar-adviser level. Advisers who train multiple Scholars therefore are counted multiple times so that the reported averages reflect the mean exposure of the mentees in the data.

Multivariate results: the pairing process. Table 4 presents the determinants of matches between Scholars and postdoctoral advisers from probit regressions at the adviser/Scholar level of analysis (12,775 pairs, of which 489 are realized).[21] 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. Likewise, advisees of graduate mentors in the top quartile of the research patentability distribution are no more or less likely to 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 percent 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 for spatial geography appear in column (3). We find strong evidence of geographic sorting, with actual pairings more likely to involve a postdoctoral adviser from the Scholar's PhD-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.[22] These results persist in column (4), which includes the most comprehensive

[21] In these regressions, person-level variables, such as postdoc adviser publication count, have negligible effects. 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. In our data structure, this is impossible by construction; 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).

[22] With a significantly larger population of Scholars, one could imagine a nested-logit modeling approach that brings the estimates of match probabilities in closer alignment with the evidence of geographic matching

set of covariates; this specification is the one used to create the weights in the IPTW analysis reported below.

Finally, recall that although 15 percent 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.

The assumption of common support. Figure 1 displays the distribution of match probabilities, separately for the cases that correspond to actual pairs and for the cases that correspond to counterfactual assignments. Inspection of the histograms shows that the region of common support is extremely wide. In fact, it is so wide that even for the least and most likely actual matches, we are able to find counterfactual matches with similar odds.

From a substantive perspective, this overlap between the two distributions is unsurprising. Almost all Scholars and advisers work in somewhat related subfields of biology. In other words, the types of counterfactuals contemplated in these matching equations do not involve pairing physicists and biologists. Rather, a scientifically distant pairing in the data might include a molecular biologist working with worms as a model organism, matched to one who does mouse genetics. Empirically, these pairings are infrequent in the data, but they occur. To provide further evidence that it is reasonable to conceive of the mentors in the data as a cohesive population through which our Scholars could match, we have characterized fully the coauthorship network in the group of postdoc mentors. This network exhibits a high degree of closure; among the 333 postdoc advisers, only 32 have no coauthors within the network, and 80 percent of these isolates are based outside North America.

Validating the exclusion restrictions for the Heckman-style analysis. 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 small in magnitude and statistically insignificant; these covariates, however, are included in the specifications.) We separately

gleaned from the oral histories. Specifically, with enough data, one could model the process of matching as unfolding within geographic regions. With a sample of fewer than 500 Scholars, however, we can only specify a single matching equation, which we saturate with covariates based on the oral histories.

analyze the determinants of pairing for Scholars who come from outside the US and for those who attended US-based undergraduate institutions. In column (1), we find that among the 121 foreign-born Scholars, there is a greater propensity to match with a postdoctoral adviser from the same country. To the extent that homophily based on national origin is orthogonal to Scholars' commercial leanings, this result can be used as an instrument to disentangle mentors social influences from selection effects. A counterargument is that individuals from particular countries might have systematically greater proclivities to engage in commercially relevant science, while also displaying a greater tendency to create native-language matches. We single out Chinese Scholars (mainland-born) because a recent study (Gaulé and Piacentini 2013) found that native Chinese graduate students in chemistry both are more likely to train with a Chinese PI for their PhD and that they are more productive than domestic students. Column (2) in Table 5 replicates our matching equation in the subsample of foreign-born Scholars that exclude those born in mainland China. The results are qualitatively similar in this smaller sample.

Column (2) shows that, among US Scholars, there is a propensity to match with a postdoctoral lab located in the same state as one's undergraduate institution. One concern with relying on this source of variation for identification is that students who attended colleges located in "biotech-heavy" states acquire their taste for commercially relevant science before graduate school, maybe through exposure to the local entrepreneurial ecosystem, internships, etc. If that were the case, the effect of doing one's postdoc in the same state would be expected to influence patenting outcomes directly, and not only through one's choice of mentor. The exclusion restriction would be clearly invalid in that case. This concern is why we included column (4) in Table 5: our results hold even when we exclude California, a state in which 56 (15.2%) of our American-born Scholars went to college and which has been an important locale in the birth and development of the biotechnology industry.[23] 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 adviser patenting in the subsample of US Scholars.

IPTW results. The first three columns in Table 6A report results of postdoc adviser patenting on Scholars' propensity to patent using inverse probability of treatment weights. Observations are Scholar-years in which the Scholar holds a faculty position and the specifi-

[23] We cannot exclude all the "biotech-heavy" states (CA, MA, WA, PA, NY, MD), since they account for 193 (52.4%) of the sample of American-born Scholars by college location. But columns (3) and (4) include state-specific intercepts for these six states.

cation is a discrete-time hazard of the first patenting event. The variable of central interest is a dummy indicating whether the Scholar's postdoc mentor had patented before the Scholar completed training. In column (1), we present the "naïve" estimates that do not include weights to adjust for the matching process. The coefficient implies that patenting is indeed subject to adviser "imprinting"; the hazard of patenting is 69 percent higher for Scholars whose postdoc advisers were patenters. Column (2) inversely weights each observation by the fitted probabilities from column (4) in Table 4 to perform IPTW estimation. Under unconfoundedness, inversely weighting Scholar observations by the probability of pairing with mentors 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 remain unchanged from the original population.

To our surprise, the magnitude of the coefficient on postdoc adviser patenting in the IPTW results (column 2) is more than two-thirds *larger* than the naïve 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 IPTW 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 the mentor's influence on the Scholar's likelihood of patenting in the IPTW estimates likely results from up-weighting the contribution of Scholars with postdoctoral mentors whose research significantly differs 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, the magnitude of the IPTW 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 change scientific foci—those who select postdoc advisers who 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.

The unexpected finding that scientific distance is associated with a larger treatment effect requires further investigation. Specifically, it becomes important to investigate whether the results are solely driven by a subset of scientific field switchers, who may differ from the general population of postdocs. We undertake two, additional assessments of this issue. First, we run regressions in which we drop from the sample Scholars in the top percentile, ventile, and decile of the distribution of IPT weights, respectively. The (unreported) coefficient estimates of a patenting postdoc mentor do decrease when we drop these scientists from the sample, but the magnitudes of the social influence effects remain large, and very similar to the ?nave? estimates obtained in column (1) (i.e., the specification that does not weight the observations by the inverse probability of treatment).

Second, we compare the career-long level of scientific focus for postdoc candidates who bridge a greater expanse of scientific distance to those who remain close to their existing areas of emphasis. We conduct this analysis to address the concern that a significant scientific change between these two phases of training is a harbinger of a more migratory style of science over the course of a scientific career. In particular, could it be that high-IPT-weight Scholars are consummate dilettantes who will continue to experiment with various topics throughout their full careers, thus increasing the odds that they will eventually stumble on areas where patenting is a natural byproduct of their scientific research? Or, does their scientific profile stabilize once they emerge from their postdoctoral fellowship?

To shed light on this question, we consider each Scholar's entire corpus of work as independent researchers. We assemble all subsequent-to-postdoc, last-authored publications, for which we assume the focal scientist is the principal investigator. Using the MeSH keywords that tag these last-authored publications, we compute an index of scientific focus: one minus the Herfindahl index over unique keywords (we compute this measure with and without weights for each keyword's frequency of use). Figure X provides a scatterplot of the Herfindahl against the log of IPT weight, together with the implied regression line in the cross-section (using data on all the 489 Scholars this time). As can be readily observed, the relationship is weak, and if anything "wrong signed": the higher the inverse probability of treatment weight, the more concentrated is the distribution of keywords that tag the Scholar's publications in his/her role as laboratory head.

From these two analyses, we conclude that the central IPTW social influence finding is not a mere artifact of the scientific field switchers in the data.

Comparing the effect of assignment on publication vs. patent output. Having established the basic social influence result, we next ask whether adviser patenting has any effect on other Scholar-level career outcomes, such as publication and citation rates. The motivation for these analyses is that *if* postdoc adviser patenting affects career outcomes that are unrelated to commercial activities, we might worry that mentor patenting has an effect only because it 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 6A report, respectively, naïve and IPTW 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 the rate of publication output. Table 6B proceeds in the same vein, this time concentrating on the impact of the Scholars' published research as measured by citations, both in subsequent scientific journals (most of which stem from articles written by other academics) and in subsequent patents (most of which stem from patents assigned to commercial firms). Both citation measures exclude self-citations. Once again, we find that the social influence of mentor patenting is highly specific: having trained with a patenting postdoc adviser increases the rate at which a Scholar's research is cited in future patents, but not in future papers.[24]

Heckman sample selection results. Recall that we use 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, whether 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,[25] 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 naïve estimates for the overall, US, and foreign-born samples, respectively. The social influence 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 dramatically shift the magnitude of the effect of mentor patenting, though the coefficient is only statistically significant at the 10

[24]In unreported analyses, we also find that adviser patenting has no effect on NIH grant funding outcomes for these Scholars.

[25]It is not possible to estimate the outcome equation in pooled cross-sections as we do in the IPTW regressions.

percent level in the sample of Scholars with US undergraduate degrees. 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 column (5). In other words, in accordance with our understanding of the matching process in this context, 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. 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. IPTW 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 five robustness checks.

First, as reported in the previous section, we undertake a form of a falsification test—we examine whether adviser patenting influences other career outcomes. We find it does not. Second, we test the sensitivity of the IPTW estimate to 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. Fourth, we investigate whether adviser patenting *after* the Scholar departs from the adviser's lab influences the likelihood of Scholar patenting. Finally, we revisit the oral histories in an effort to determine whether the lack of discussion of commercial interests in the adviser search process results from Scholars' reluctance to disclose their preferences on this issue because of the taboo associated with commercial science.

We begin with the sensitivity of adviser patenting to changes in the construction of the counterfactual dyads in the first-stage analysis. The coefficients in Table 6A and 6B 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 IPTW-adjusted effect of adviser patenting in these two datasets and found that the coefficient varied only slightly from that in Table 6, column (2).[26] Thus, within the tolerances 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.[27] To address this 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 5^{th}, 10^{th}, and 15^{th} 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 themselves patent, and therefore depart from their adviser's behavior? Conversely, are trainees of *non*-patenting advisers who do not patent later in their careers more scientifically proximate to their advisers than trainees of non-patenting advisers who do patent, and thus deviate from adviser behavior? 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 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-whisker plots in Figure 2. We report the distribution of scientific similarity scores between postdoc advisers and Scholars broken out by whether or not the adviser

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

[27] 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.

was a patenter and whether the Scholar becomes a patenter. The informative comparisons are between the two distributions *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 2 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, 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 indicator variable that switches on if and when the Scholar's former postdoctoral mentor starts applying for patents (results available in the online appendix, Part III). Reassuringly, the coefficient for this indicator variable is much smaller in magnitude than its counterpart in Table 6A, and statistically insignificant. If post-training-period adviser patenting 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 firsthand exposure is required buttresses our claim that the core result is a causal social influence.

6 Discussion and Conclusion

The paper's central theoretical claim is that when actors connect based on a small set of attributes X, it is often the case that some additional characteristic Z, which was never considered when a choice was made to develop a relationship, becomes socially transmitted. We develop the psychological and sociological foundations of a theory we call partially deliberate matching, and we present a set of empirical methods that are generally useful for uncovering causal social influence effects in observational data.

We present two central empirical 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 causal social influence effect is that postdoctoral advisers' patenting behavior is transmitted to their trainees. Through the use of inverse probability of treatment-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 a sociological reframe, we find (Table 6A) 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 mentor in a period of advisees' intensive 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 been accorded much thought 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 largely orthogonal to the search process, but it is 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.

This core empirical result also dovetails with the literature on career sequences (e.g., 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 career trajectories. In this sense, the findings expose one type of "turning point" in academic scientists' careers (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.

The fact that matching is only partially deliberate opens avenues for the unforeseen transmission of attitudes and behaviors. In the majority of instances, unanticipated exposures are of insignificant consequence. All of us can recruit 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 think and behave. 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 with 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.

We do believe that the theory or partially deliberate matching generalizes to other settings. For instance, Marmaros and Sacerdote (2006) show that friendship formation on a college campus depends on random assignment to residence halls. This precisely sets up the pre-conditions for social influences given under partially deliberate matching. Another context in which our approach may be useful is analyses of social influences in relationships with online origins. There is increasing sociological interest in the types of relationships that are formed online, which run the gamut from romantic relationships (Wimmer and Lewis 2010) to open source communities (Piskorski and Gorbatai 2013) to social structures within online games (e.g., Burt 2012; Torfason 2012). In each of these cases, the dimensionality for matching is limited to a narrow band of information that each potential interactant presents, but these relationships often blossom into richer and more multiplex forms of interaction. In fact, we believe that partially deliberate matching may permeate the sociology of the digital economy, as many social relationships in online markets arise from a limited set of compatibilities, but evolve into wider pipes.

In general terms, we conclude with four conditions that may be necessary for researchers to document causal social influence in contexts of partially deliberate matching. (i) An ability to gather qualitative evidence that reveals the attributes that are most relevant for social matching. In its effect, the role of qualitative evidence is to provide the researcher with a theory of the data generating process. Such a theory is essential for to support the unconfoundedness assumption, which is untestable in the quantitative data and justifiable only when the researcher has a rich understanding of how matches come to be. (ii) Matching regressions that validate the qualitative evidence. (iii) IPTW regressions to recover the causal effect of a putative social influence on unselected attributes. (iv) Ideally, additional variation in the data (other dependent variables, the timing of events) that can be exploited to rule out the plausibility of results driven by latent homophily. These conditions are a tall order, but so it goes to establish persuasive, causal evidence of social influence in observational data.

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**(1): 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.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan. 2009. "Distinguishing Influence Based Contagion From Homophily-Driven Diffusion in Dynamic Networks." *PNAS* **106**(51): 21544-49.

Azoulay, Pierre, Waverly Ding, and Toby Stuart. 2009. "The Effect of Academic Patenting on the Rate, Quality, and Direction of (Public) Research Output." *Journal of Industrial Economics* 57**(4)**: 637-623.

Azoulay, Pierre, Joshua S. Graff Zivin, and Bhaven N. Sampat. 2012. "The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine." Pp. 107-155 in *The Rate and Direction of Inventive Activity Revisited,* edited by Joshua Lerner and Scott Stern. Chicago, IL: University of Chicago Press.

Barber, Jennifer S., Susan A. Murphy, and Natalya Verbitsky. 2004. "Adjusting for Time-Varying Confounding in Survival Analysis." Sociological Methodology **34**(1):163-192.

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**(2): 219-224.

Burt, Ronald S. 2012. "Network-Related Personality and the Agency Question: Multirole Evidence from a Virtual World." *American Journal of Sociology* **118**(3): 543-591.

Coleman, James, Elihu Katz, and Herbert Menzel. 1957. "The Diffusion of an Innovation Among Physicians." *Sociometry* **20**(4): 253-70.

Dehejia, Rajeev H., and Sadek Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* **84**(1): 151-161.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional and Collective Rationality in Organizational Fields." *American Sociological Review* **48**(2): 147-60.

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

Evans, James A. 2010. "Industry Induces Academic Science to Know Less about More." *American Journal of Sociology* **116**(2): 389-452.

Etzkowitz, Henry. 1998. "The Norms of Entrepreneurial Science: Cognitive Effects of the New University-Industry Linkages." *Research Policy* **27**(8): 823-833.

Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* **86**(5): 1015-1035.

Friedkin, Noah E. 1993. "Structural Bases of Interpersonal Influence in Groups: A Longitudinal Case Study." *American Sociological Review* **58**(6): 861-72.

Gaulé, Patrick, and Mario Piacentini. 2013. "Chinese Graduate Students and U.S. Scientific Productivity." *Review of Economics and Statistics* **95**(2): 698-701.

Gerfin, Michael. 1996. Parametric and Semi-Parametric Estimation of the Binary Response Model of Labour Market Participation." Journal of Applied Econometrics **11**(3): 321-339.

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

Hauser, John R., and Birger Wernerfelt. 1990. "An Evaluation Cost Model of Consideration Sets." *The Journal of Consumer Research* **16**(4): 393-408.

Heckman, James J. 1979. "Sample Selection Bias as a Specifiction Error." *Econometrica* **47**(1): 153-161.

Hogg, Michael A., and Dominic Abrams. 2002. *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*. New York, NY: Routledge.

Huang, I-Chan, Constantine Frangakis, Francesca Dominici, Gregory B. Diette and Albert W. Wu. 2005. "Application of a Propensity Score Approach for Risk Adjustment in Profiling Multiple Physician Groups on Asthma Care." *Health Services Research* **40**(1): 253-78.

Imbens, Guido W. 2000. "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika* **87**(3): 706-710.

Iyengar, Sheena S., and Mark R. Lepper. 1999. "Rethinking the Value of Choice: A Cultural Perspective on Intrinsic Motivation." *Journal of Personality and Social Psychology* **76**(3): 349-66.

Kandel, Denise B. 1978. "Homophily, Selection, and Socialization in Adolescent Friendships." *American Journal of Sociology* **84**(2): 427-436.

Knorr-Cetina, Karin D. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Kono, Clifford, Donald Palmer, Roger Friedland, and Matthew Zafonte. 1998. "Lost in Space: The Geography of Corporate Interlocking Directorates." *American Journal of Sociology* **103**(4): 863-911.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* **76**(4): 604-620.

Lazarsfeld, Paul F. and Robert K. Merton. 1954. "Friendship as a Social Process: A Substantive and Methodological Analysis." Pp 18-66 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**(5): 816-830.

Marmaros, David, and Bruce Sacerdote. 2006. "How Do Friendships Form?" *Quarterly Journal of Economics* **121**(1): 79-119.

Marsden, Peter V. 1981. "Introducing influence Processes Into a System of Collective Decisions." *American Journal of Sociology* **86**(6): 1203-35.

McPherson, J. Miller, Pamela A. Popielarz, and Sonja Drobnic. 1992. "Social Networks and Organizational Dynamics." *American Sociological Review* **57**(2): 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.

Merton, Robert K. 1957. "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* **22**(6): 635-59.

Mouw, Ted. 2003. "Social Capital and Finding a Job: Do Contacts Matter?" *American Sociological Review* **68**(6): 868-898.

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

Nerad, Maresi, and Joseph Cerny. 1999. "Postdoctoral Patterns, Career Advancement, and Problems." *Science* **285**(5433): 1533-1535.

Newcomb, Theodore Meade. 1961. *The Acquaintance Process*. Oxford, UK: Holt, Rinehart and Winston.

Newey, Whitney K., James L. Powell, and James R. Walker. 1990. "Semiparametric Estimation of Selection Models: Some Empirical Results." *American Economic Review* **80**(2): 324-28.

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**(1-2): 99-114.

———————————————. 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.

———————————————. 2004. "Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community." *Organization Science* **15**(1): 5-21.

Payne, John W. 1982. "Contingent Decision Behavior." *Psychological Bulletin* **92**(2): 382-402.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge, UK: Cambridge University Press.

Piskorski, Mikolaj Jan, and Andreea Gorbatai. 2013. "Testing Coleman's Social-Norm Enforcement Mechanism: Evidence from Wikipedia." HBS Working Paper No. 11-055.

Reagans, 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.

Rubin, Donald B. 1990. "Formal Model of Statistical Inference for Causal Effects." *Journal of Statistical Planning and Inference.* **25**(3): 279-292.

Shafir, Eldar, Itamar Simonson, and Amos Tversky. 1993. "Reason-based Choice." *Cognition* **49**(1): 11-36.

Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods & Research* **40**(2): 211-239.

Simon, Herbert A. 1947. *Administrative Behavior*. New York, NY: The Free Press.

Sorenson, Olav, and Toby E. Stuart. 2001. "Syndication Networks and the Spatial Distribution of Venture Capital Investments." *American Journal of Sociology* **106**(6): 1546-1588.

Stephan, Paula. E., and Jennifer Ma. 2005. "The Increased Frequency and Duration of the Postdoctorate Career Stage." *American Economic Review* **95**(2): 71-75.

Stouffer, Samuel A. 1949. "An Analysis of Conflicting Social Norms." *American Sociological Review* **14**(6): 707-17.

Stovel, Katherine, Michael Savage, and Peter Bearman. 1996. "Ascription into Achievement: Models of Career Systems at Lloyds Bank, 1890-1970." *American Journal of Sociology* **102**(2): 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**(3-4): 211-227.

Timmermans, Danielle. 1993. "The Impact of Task Complexity on Information Use in Multi-Attribute Decision Making." *Journal of Behavioral Decision Making* **6**(2): 95-111.

Torfason, Magnus. 2012. "With Us or Against Us? Networks, Identity and Order in a Virtual World." HBS Working Paper No. 13-019.

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**(2): 229-252.

Van den Bulte, Christophe, and Gary L. Lilien. 2001. "Medical Innovation Revisited: Social Contagion versus Marketing Effort." *American Journal of Sociology* **106**(5): 1409-1435.

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* **116**(2): 583-642.

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

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

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

Category (%/N)	Representative Quotes
1. Science (95%/59)	
Extension of Prior Knowledge	[Emerson; pp. 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; pp. 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; pp. 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; pp. 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, pp. 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. Adviser Status (15%/9)	[Hirano, pp. 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; pp. 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
MD/PhD	.133	.340	0	1
Highest degree year	1986	4.88	1973	1998
Year of First Academic Appt.	1990	5.20	1977	2000
Member of the Natl. Academy of Sciences	.061	.240	0	1
Howard Hughes Medical Investigator	.161	.368	0	1
Nobel Laureate	.002	.045	0	1
Cmltv. Nb. of Publications	70.71	49.02	11	381
Cmltv. Nb. of Citations in Publications	.039	.273	0	4
Cmltv. Nb. of Citations in Patents	.027	.161	0	1
Patenter	.360	.480	0	1
Cmltv. Nb. of Patents	1.37	4.13	0	57
Patentability Stock	.516	.574	0	4.49

Note: Citation information is current as of 2008. Publication and patent cumulative counts are computed as of the end of the year 2007, which coincides with the end of the observation period for all Scholars. Citations in publications is the total number of (forward) citations in publications made to all of the Scholar's publications. Citations in patents is the total number of (forward) citations in patents made to all of the Scholar's publications. Patenter is an indicator variable denoting whether the Scholar has applied for at least one patent by the end of 2007. Patentability stock is a measure of the underlying "patentability" of a Scholar's research.

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				
Cmltv. Nb. of Publications	88.61	81.40	1	513
Patenter	.194	.396	0	1
Cmltv. Nb. 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				
Cmltv. Nb. of Publications	108.42	100.89	0	729
Patenter	.438	.497	0	1
Cmltv. Nb. 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 Urchin Development
4	Robert Baldwin	No	No	Yes	Protein Folding
4	Gunter Blöbel	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	Yes	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	Yes	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 Mentors (Probit, All Scholars)

	(1)	(2)	(3)	(4)
Grad/Postdoc Mentors Keyword Overlap (Bottom Quartile)		-1.173** (0.079)		-1.182** (0.078)
Grad/Postdoc Mentors Keyword Overlap (Second Quartile)		-0.746** (0.062)		-0.751** (0.062)
Grad/Postdoc Mentors Keyword Overlap (Third Quartile)		-0.481** (0.055)		-0.482** (0.055)
Grad & Postdoc Mentors at same university			0.328** (0.090)	0.368** (0.097)
Grad & Postdoc Mentors in same state, different university			0.140* (0.063)	0.144* (0.067)
Scholar & Postdoc of the same gender	0.011 (0.049)	0.032 (0.052)	0.012 (0.050)	0.035 (0.052)
Scholar & Postdoc both female	0.051 (0.135)	-0.014 (0.135)	0.044 (0.138)	-0.020 (0.139)
Grad & Postdoc Mentors both patent	0.114 (0.085)	0.094 (0.091)	0.099 (0.086)	0.075 (0.093)
Only Grad Mentor patents	-0.074 (0.074)	-0.090 (0.077)	-0.084 (0.074)	-0.101 (0.077)
Only Postdoc Mentor patents	-0.003 (0.035)	-0.048 (0.039)	-0.005 (0.036)	-0.049 (0.040)
Grad & Postdoc Mentors both in top quartile of research patentability	-0.043 (0.096)	-0.156 (0.099)	-0.044 (0.097)	-0.161 (0.100)
Only Grad Mentor in top quartile of research patentability	-0.033 (0.069)	-0.022 (0.073)	-0.036 (0.070)	-0.025 (0.073)
Only Postdoc Mentor in top quartile of research patentability	0.014 (0.044)	-0.036 (0.048)	0.018 (0.044)	-0.034 (0.048)
Constant	-1.340** (0.183)	-0.603** (0.241)	-1.337** (0.183)	-0.605* (0.241)
Log likelihood	-2,042	-1,888	-2,034	-1,879
Observations	12,775	12,775	12,775	12,775
# of Scholars	489	489	489	489
# of Postdoc Mentors	333	333	333	333

Note: Estimates are displayed as raw coefficients. All models include Scholar-cohort dummies, and an indicator variable if the grad or postdoc adviser had sent/received multiple students within that cohort-year. All models also include the sum and absolute difference of grad and postdoc adviser publication counts, as well as the square and cube of this variable (coefficient estimates not reported). For Grad/Postdoc Mentors Keyword Overlap, the excluded quartile corresponds to the dyads that are most scientifically similar.

Robust standard errors, clustered at the postdoc mentor level reported in parentheses below each coefficient estimate.
† significant at 10%; * significant at 5%; ** significant at 1%.

Table 5: Determinants of Pairing by Scholars and Potential Postdoc Mentors (Probit, by Subsample)

	(1)	(2)	(3)	(4)
Subsample	Only Foreign Scholars	Only Foreign Scholars–excl. Chinese	Only US Scholars	Only US Scholars–excl. California
Scholar & Postdoc Mentor born in same foreign country	0.803** (0.215)	0.880*** (0.224)		
Undergrad & Postdoc university in same state			0.212** (0.076)	0.263** (0.098)
Grad/Postdoc Mentors Keyword Overlap (Bottom Quartile)	-1.322** (0.166)	-1.318*** (0.189)	-1.177** (0.096)	-1.292** (0.109)
Grad/Postdoc Mentors Keyword Overlap (Second Quartile)	-0.977** (0.139)	-1.025*** (0.154)	-0.696** (0.070)	-0.689** (0.076)
Grad/Postdoc Mentors Keyword Overlap (Third Quartile)	-0.486** (0.102)	-0.441*** (0.115)	-0.481** (0.063)	-0.473** (0.072)
Grad & Postdoc training at same university	0.450** (0.163)	0.397* (0.170)	0.285* (0.119)	0.272* (0.124)
Grad & Postdoc training in same state, different university	0.348* (0.139)	0.172 (0.160)	0.039 (0.082)	0.099 (0.089)
Scholar & Postdoc Mentor are of the same gender	0.002 (0.113)	-0.046 (0.123)	0.044 (0.058)	0.025 (0.062)
Scholar & Postdoc Mentor are both female	0.151 (0.286)	-0.050 (0.350)	-0.143 (0.189)	-0.099 (0.221)
Constant	-0.689 (0.577)	-0.789 (0.766)	-0.563† (0.316)	-0.208 (0.411)
Log likelihood	-441	-338	-1,421	-1,198
Observations	3,097	2,404	9,678	8,201
# of Scholars	121	93	368	312
# of Postdoc advisers	333	333	333	333

Note: Estimates are displayed as raw coefficients. All models include Scholar-cohort dummies, and an indicator variable if the grad or postdoc adviser had sent/received multiple students within that cohort-year. All models also include the sum and absolute difference of grad and postdoc adviser publication counts, as well as the square and cube of this variable (coefficient estimates not reported). For Grad/Postdoc Mentors Keyword Overlap, the excluded quartile corresponds to the dyads that are most scientifically similar. Models (3) and (4) include undergraduate university state indicator variables for MA, CA, WA, NY, MD, NJ, and PA (the states in which the great bulk of biotech entrepreneurship are located). None of the coefficients on these indicators is significant, and those results are not shown.

Robust standard errors, clustered at the postdoc mentor level reported in parentheses below each coefficient estimate.
† significant at 10%; *significant at 5%; ** significant at 1%.

Table 6A: Impact of Postdoc Mentor Patenting on Scholar Patenting Propensity and Publication Rates

Dependent Variable	Scholar First Patenting Event			Scholar Publ. Count	
Model	Discrete-Time Hazard Rate			QML-Poisson	
IPT Weights	No	Yes	Yes–excl. scientific distance covariates	No	Yes
	(1)	(2)	(3)	(4)	(5)
Postdoc Mentor was a patenter	0.535^{**}	0.854^{**}	0.521^{**}	-0.023	-0.006
	(0.171)	(0.208)	(0.177)	(0.042)	(0.045)
Research Patentability Flow, no lag	4.178^{**}	$4.910^{\dagger}$	4.071^{**}	3.893^{**}	4.264^{**}
	(1.299)	(2.705)	(1.408)	(0.524)	(0.513)
Research Patentability Stock (lagged one year)	1.013^{*}	0.230	$1.007^{\dagger}$	0.538^{**}	0.538^{**}
	(0.478)	(0.686)	(0.532)	(0.068)	(0.065)
Female	-0.716^{**}	-0.923^{**}	-0.856^{**}	-0.127^{**}	-0.077
	(0.241)	(0.295)	(0.256)	(0.046)	(0.056)
MD/PhD	0.478^{*}	0.630^{*}	0.501^{*}	0.186^{**}	0.170^{**}
	(0.224)	(0.297)	(0.227)	(0.051)	(0.055)
Log(University NIH $)	-0.234^{*}	-0.127	-0.281^{**}	-0.012	-0.003
	(0.096)	(0.137)	(0.101)	(0.026)	(0.025)
Log(University Patents)	0.095	0.071	0.115	0.016	-0.011
	(0.072)	(0.102)	(0.077)	(0.020)	(0.026)
Constant	-1.692	-3.669	-1.347	0.116	-0.070
	(1.852)	(2.559)	(1.935)	(0.821)	(0.750)
Log-pseudolikelihood	-677	-39,801	-46,137	-14,059	-792,137
Observations	5,250	5,250	5,250	6,587	6,587
# 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 Scholar-cohort dummies, and a full suite of calendar-year indicator variables (not reported). Research patentability is a measure of the underlying patentability of a Scholar's research, derived from the publication and patent records of 9,000 life scientists (see Appendix I).

Robust standard errors, clustered at the postdoc mentor level reported in parentheses below each coefficient estimate. † significant at 10%; * significant at 5%; ** significant at 1%.

Table 6B: Impact of Postdoc Mentor Patenting on Scholar Citations in Publications or Patents

Dependent Variable	Scholar Citations in Publications from Publications		Scholar Citations in Patents from Publications	
Model	QML-Poisson		Logit	
IPT Weights	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
Postdoc Mentor was a patenter	-0.041	0.005	0.329**	0.366**
	(0.084)	(0.121)	(0.091)	(0.111)
Research Patentability Flow, no lag	3.818**	4.276**	32.695**	26.621**
	(0.526)	(0.551)	(2.066)	(4.366)
Research Patentability Stock (lagged 1-year)	0.814**	0.786**	-0.092	0.098
	(0.066)	(0.084)	(0.207)	(0.338)
Female	-0.360**	-0.363**	-0.117	-0.172
	(0.087)	(0.104)	(0.106)	(0.129)
MD/PhD	0.142	0.064	0.139	0.202
	(0.096)	(0.098)	(0.144)	(0.154)
Log(University NIH $)	-0.042	0.011	-0.036	-0.071
	(0.054)	(0.069)	(0.063)	(0.068)
Log(University Patents)	0.028	0.023	0.015	0.007
	(0.041)	(0.049)	(0.045)	(0.051)
Constant	3.793**	2.673*	-4.508**	-4.688**
	(1.105)	(1.287)	(1.420)	(1.404)
Log-pseudolikelihood	-1,240,426	-6,7328,843	-3,414	-203,187
Observations	6587	6587	6579	6579
# of Scholars	489	489	489	489
# of postdoc advisers	333	333	333	333

Note: Estimates are displayed as raw coefficients. All models include Scholar-cohort dummies, and a full suite of calendar-year indicator variables (not reported). Research patentability is a measure of the underlying patentability of a Scholar's research, derived from the publication and patent records of 9,000 life scientists (see Appendix I). Citations to publications in patents are rare, relative to the incidence of citations to publications in publications. As a result, we collapse the flow of citations in patents for a Scholar in a given year (a count random variable) into an indicator variable (equal to one for at least one citation to publications in patents in a given year).

Robust standard errors, clustered at the postdoc mentor level reported in parentheses below each coefficient estimate.
† significant at 10%; * significant at 5%; ** significant at 1%.

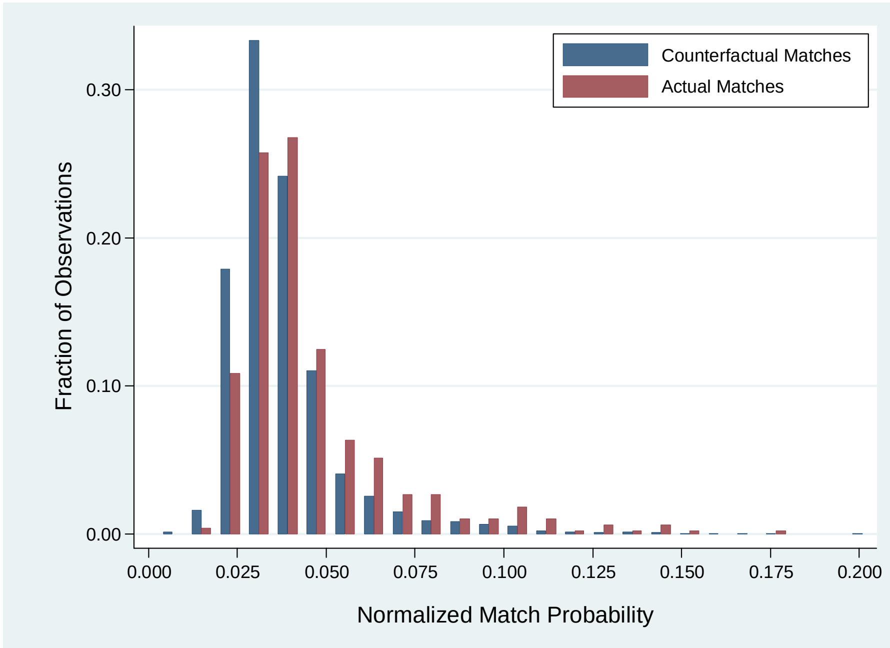
Table 7: Cross-Sectional Probit Model of Scholar Patenting with Sample Selection

Subsample	All Scholars	US Scholars	Foreign Scholars	US Scholars	Foreign Scholars
Model		Probit		Probit with Sample Selection (Van de Ven and Van Praag 1981)	
Exclusion Restrictions	N/A	N/A	N/A	Undergrad & Postdoc University in Same State	Both Scholar & Postdoc from Same Foreign Country
	(1)	(2)	(3)	(4)	(5)
Postdoc Mentor was a Patenter	0.343** (0.132)	0.323* (0.157)	0.556* (0.268)	0.296† (0.163)	0.565* (0.230)
Research Patentability Stock	0.467** (0.119)	0.346** (0.129)	1.082** (0.287)	0.344* (0.149)	1.004* (0.390)
Female	-0.597** (0.156)	-0.466** (0.179)	-0.852** (0.326)	-0.465† (0.201)	-0.705* (0.360)
MD/PhD	0.323† (0.182)	0.489* (0.214)	-0.420 (0.426)	0.501* (0.246)	-0.393 (0.361)
Log(University NIH $)	-0.134† (0.074)	-0.099 (0.067)	-0.457** (0.164)	-0.088 (0.071)	-0.420* (0.180)
Log(University Patents)	0.059 (0.058)	0.030 (0.062)	0.284† (0.166)	0.025 (0.063)	0.264 (0.159)
Constant	1.996 (1.258)	1.393 (1.195)	5.975* (2.383)	1.289 (1.661)	6.626** (2.305)
atanh(ρ)				-0.213 (1.065)	-0.650 (0.632)
Log-pseudolikelihood	-266	-199	-56	-1,642	-507
Observations (Scholars)	489	368	121	368	121
Nb. of Postdoc Mentors	333	262	102	333	333
Potential Dyads				9,678	3,097

Note: Research Patentability, university patents and university NIH funding are measured as of 2007—the end of our observation period. Models (2) and (4) include eleven cohort indicator variables. Models (3) and (5) include four cohort indicator variables. All models also include on the right-hand side all the covariates in the matching equations (cf. Models (1) and (2) in Table 5, which reports estimates for the first-stage matching equations). Though these covariates are included, we do not report the corresponding estimates since they are not of substantive interest in the second-stage outcome equations.

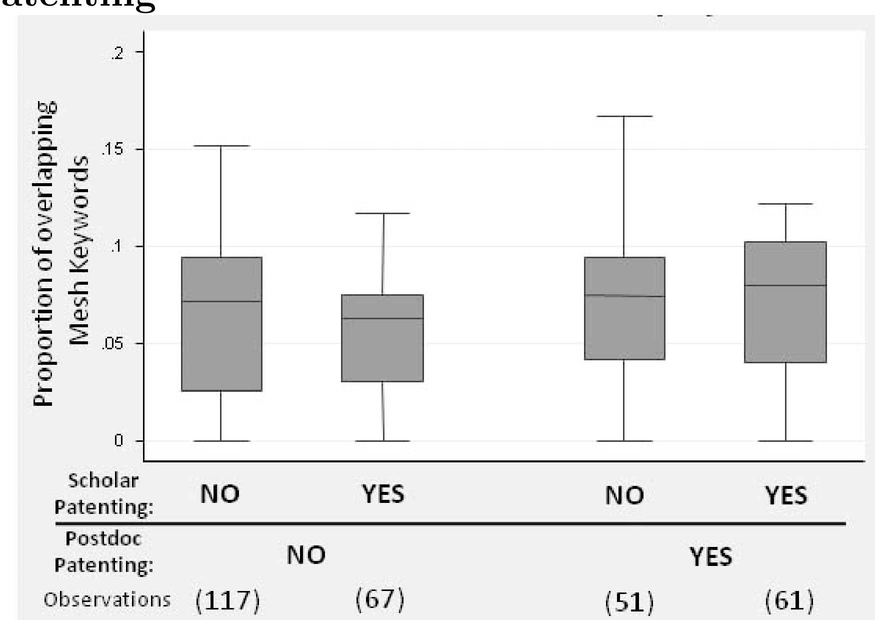
Robust standard errors, clustered at the postdoc mentor-level reported in parentheses below each coefficient estimate.
† significant at 10%; * significant at 5%; ** significant at 1%.

Figure 1: Distribution of Match Probabilities



Note: The histograms above depict the distribution of (normalized) match probability for the 489 Scholars × 333 postdoctoral advisers, broken down between 489 actual matches and 12,286 counterfactual matches.

Figure 2: 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 keyword overlap; t is the last year of Scholar training. 296 Scholars are presented.

Appendix I
Keyword Weights and Patentability Score

Not all research performed by Pew/Searle Scholars is of commercial interest, or susceptible to being patented. This could be either because the research is focused on the elucidation of fundamental biological mechanisms, or because the research, even if applied, occurs in an area where commercial applications are difficult to envisage (e.g., population biology). For each scientist, we construct a time-varying score that seeks to control for variation in the inherent patentability of their research.

The essential inputs into the computation of this score are *Medical Subject Heading* [MeSH] terms. MeSH terms constitute a controlled vocabulary maintained by the National Library of Medicine that provides a very fine-grained partition of the intellectual space spanned by the biomedical research literature. There are 26,581 descriptors in the 2012 MeSH edition (new terms are added to the dictionary as scientific advances are made). Almost every publication in PubMed is tagged with a set of MeSH terms (between 1 and 103 in the current edition of PubMed, with both the mean and median approximately equal to 11). Importantly for our purposes, MeSH keywords are assigned to each scientific publication by professional indexers and not by the authors themselves.

The construction of the research patentability (RP) score proceeds in two steps. First, we compute time-varying MeSH-level weights that capture the extent to which each individual keyword j is associated with patentable research at time t. Second, we apply these keywords to the body of work for each scientist i active in year t. We describe both steps in detail below.

Step 1: Computation of the weights

The essence of the weighting scheme we develop is that a keyword is weighted more heavily (i.e., be deemed more "patentable") when it appears disproportionately often in publications whose authors have already patented in the past. A requirement is therefore to have a large population of scientists for whom both the bibliome and the patentome are well characterized. One such population is the set of Pew/Searle Scholars. However, it seems desirable for the publications of the scientists that are the focus of the analysis NOT to influence the calculation of the sts. We therefore use as a reference population a set of 12,159 elite, US-based academic life scientists for which Azoulay, Graff Zivin, and Sampat (2012) have assigned publications and patents using an exhaustive, manual hand-coding process (cf. Appendices A, B, and C, pp. 145-149).[1]

With these preambles in mind, w_{jt} the patentability weight for each keyword j in year t is defined as:

$$w_{jt} = \frac{\sum_{s \in I_t^p} \frac{m_{sjt}}{\sum_k m_{skt}}}{\sum_{s \in I_t^{np}} m_{sjt}}$$

where m_{sjt} denotes the number of times keyword j has appeared in articles published up to year t by scientist s, I_t^p is the subset of scientists in our sample that have already applied for one or more patents as of year t, and I_t^{np} is the subset of scientists in our sample that have not yet applied for any patent as of year t.

To create the numerator of w_{jt}, we first create a row-normalized matrix with each scientist in the patenting regime listed in a row and each of the keywords used to describe their papers up to year t listed in a column. The sj^{th} cell in the matrix, $[m_{sjt}/\sum_k m_{skt}]$, corresponds to the proportion of keywords for scientist s that corresponds to keyword j. We then take the column sums from this matrix, i.e., we sum the contributions of individual patenting scientists for keyword j. Turning next to the denominator, we proceed in a similar manner, except that the articles considered only belong to the set of scientists who have not applied for patents as of year t. The numerator is then deflated by the frequency of use for j by non-patenters.

The weights w_{jt} are large for keywords that have appeared with disproportionate frequency as descriptors of papers written by scientists already in the patenting regime, relative to scientists not yet in the patenting

[1] The fraction of patenters in this population is slightly lower than in the Pew/Searle Scholar population (33% vs. 36%).

regime. $w_{jt} = 0$ for all keywords that have never appeared in the titles of papers written by scientists that have patented before t.

Step 2: Computation of the RP score

We now turn to the sample of interest, the set of 489 Pew/Searle Scholars. For each scientist i in the dataset, we produce a list of MeSH keywords in the individual's papers published in year t, calculate the proportion of the total represented by each keyword j, apply the appropriate keyword weight $w_{j,t-1}$, and sum over keywords to produce a composite score. The resulting variable increases in the degree to which keywords in the titles of a focal scientist's papers have appeared relatively more frequently in the titles of other academics who have applied for patents. This score is entered in the regressions to control for the research patentability of scientists' areas of specialization.

To illustrate the construction of the research patentability measure, Table A1 lists some representative keywords, along with their patentability weights in the year 2000.

Appendix II
Semiparametric Estimation of the Sample Selection Model

The parameters of discrete-choice models are typically estimated by maximum likelihood (ML) after imposing assumptions on the distribution of the underlying error terms. If the distributional assumptions are correctly specified, then parametric ML estimators are known to be consistent and asymptotically efficient. However, departures from the distributional assumptions may lead to inconsistent estimation. This problem has motivated the development of several semiparametric estimation procedures which consistently estimate the model parameters under less restrictive distributional assumptions.

Probit model with sample selection. Consider a bivariate binary-choice model with sample selection where the indicator Y_1 is always observed, while the indicator Y_2 is assumed to be observed only for the subsample of n_1 observations (with $n_1 < n$) for which $Y_1 = 1$. The model can be written as:

$$Y_j^* = \alpha_j + \beta_j^T X_j + \varepsilon_j \qquad j = 1, 2 \tag{8}$$

$$Y_1 = 1(Y_1^* \geq 0) \tag{9}$$

$$Y_2 = 1(Y_2^* \geq 0) \qquad \text{if } Y_1 = 1 \tag{10}$$

where α_1 and α_2 are intercept terms, β_1 and β_2 are $[1 \times k_1]$ and $[1 \times k_2]$ column vectors, respectively.

When the latent regression errors ε_1 and ε_2 have a bivariate Gaussian distribution with zero means, unit variances, and correlation coefficient ρ, Model (1)-(3) is known as a bivariate probit model with sample selection. Unlike the case of full observability, the presence of sample selection has two important implications. First, ignoring the potential correlation between the two latent regression errors may lead to inconsistent estimates of $\theta_2 = (\alpha_2, \beta_2)$ and inefficient estimates of $\theta_1 = (\alpha_1, \beta_1)$. Second, identifiability of the model parameters requires imposing at least one exclusion restriction on the two sets of exogenous covariates X_1 and X_2.

Construction of the log-likelihood function for joint estimation of the overall vector of model parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \rho)$ is straightforward after noticing that the data identify only three possible events: $(Y_1 = 1, Y_2 = 1)$, $(Y_1 = 1, Y_2 = 0)$, and $(Y_1 = 0)$. Let $\phi(.)$ denote the standardized Gaussian distribution function and $\Phi(.,.,\rho)$ denote the bivariate Gaussian distribution function with zero means, unit variances, and correlation coefficient ρ. The log-likelihood function for a random sample of n observations is

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} Y_{i1} Y_{i2} ln\pi_{i11}(\boldsymbol{\theta}) + Y_{i1}(1 - Y_{i2}) ln\pi_{i10}(\boldsymbol{\theta}) + (1 - Y_{i1}) ln\pi_{i0}(\theta_1) \tag{11}$$

where:

$$\begin{aligned}
\pi_{11}(\boldsymbol{\theta}) &= Prob(Y_1 = 1, Y_2 = 1) && = \Phi(\mu_1, \mu_2, \rho) \\
\pi_{10}(\boldsymbol{\theta}) &= Prob(Y_1 = 1, Y_2 = 0) && = \phi(\mu_1) - \Phi(\mu_1, \mu_2, \rho) \\
\pi_{01}(\boldsymbol{\theta}) &= Prob(Y_1 = 0, Y_2 = 1) && = \phi(\mu_2) - \Phi(\mu_1, \mu_2, \rho) \\
\pi_{00}(\boldsymbol{\theta}) &= Prob(Y_1 = 0, Y_2 = 0) && = 1 - \phi(\mu_1) - \phi(\mu_2) - \Phi(\mu_1, \mu_2, \rho) \\
\pi_0 &= \pi_{00} + \pi_{01} && \\
\mu_j &= \alpha_j + \beta_j^T X_j && j = 1, 2
\end{aligned}$$

The ML estimator $\hat{\boldsymbol{\theta}}$ maximizes the log-likelihood function in (4) over the parameter space $\boldsymbol{\Theta} = \mathbb{R}^{k_1+k_2+2} \times]-1; 1[$. We implement this estimator — also called the probit model with sample selection in the literature (Van de Ven and Van Praag 1981) — to produce the estimates presented in Table 7.

Semi-nonparametric estimation (SNP). The semi-nonparametric approach of Gallant and Nychka (1987), originally proposed for estimation of density functions, was adapted to estimation of univariate and bivariate binary-choice models by Gabler, Laisney, and Lechner (1993).

The basic idea of SNP estimation is to approximate the unknown densities of the latent regression errors by Hermite polynomial expansions and use the approximations to derive a pseudo-ML estimator for the model parameters. The semiparametric specification of the log-likelihood functions have the same form as (4), with the probability functions replaced by:

$$\begin{aligned}
\pi_{11}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= 1 - F_1(-\mu_1) - F_2(-\mu_2) + F(-\mu_1, -\mu_2) \\
\pi_{10}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= F_2(-\mu_2) - F(-\mu_1, -\mu_2) \\
\pi_{01}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= F_1(-\mu_1) - F(-\mu_1, -\mu_2) \\
\pi_{00}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= F(-\mu_1, -\mu_2)
\end{aligned}$$

where F_j is the unknown marginal distribution of ε_j in eqn. (1), and F is the unknown joint distribution of $(\varepsilon_1,\varepsilon_2)$. Following Gallant and Nychka (1987), the unknown joint density f is approximated by a Hermite polynomial expansion:

$$f^*(x_1, x_2) = \frac{1}{S}[P(x_1, x_2)]^2 \phi(x_1)\phi(x_2) \tag{12}$$

where $\phi(.)$ is the standardized gaussian density, $P(x_1, x_2)$ is a polynomial of order $r = (r_1, r_2)$, and S is a normalizing constant that ensures that f^* integrates to 1.

Adopting the SNP approximation to the density of the latent regression errors does not guarantee that they have zero mean and unit variance. The zero-mean condition implies that some location restriction needs to be imposed on either the distributions of the error terms, or the systematic part of the model. As a result, we set the two intercept coefficients α_1 and α_2 to their parametric estimates. The Probit with sample selection and its SNP equivalent do not yield estimates that are directly comparable because the SNP approximation does not have unit variance. However, one can still compare the ratio of the estimated coefficients across the two models.

SNP estimators can be obtained by maximizing the pseudo log-likelihood function (4) in which the unknown distribution functions F, F_1, and F_2 are replaced by their approximations F^*, F_1^*, and F_2^*. As shown by Gallant and Nychka (1987), the resulting pseudo-ML estimator is $\sqrt{n}$-consistent provided that both r_1 and r_2 increase with the sample size. In practice, we select the values of r_1 and r_2 through a sequence of likelihood-ratio tests.

Application to the mentor/Scholar data. Table A2 presents semi-nonparametric estimates which mirror the parametric estimates displayed in Table 7. No coefficients estimates are reported for the constant term, since these are set at the value implied by the equivalent parametric model in each column, as explained above. Asymptotic properties of the SNP estimators require that the degree r of the Hermite polynomial expansion increases with the sample size. In particular, SNP generalizes the Probit model only if $R \geq 3$ (see Gabler et al. [1993]). In our application, many more observations are available to identify μ_1 than are available to identify μ_2. As a result, in Models 4 and 5, the order of the Hermite polynomial is four for the matching step, but only three for the outcome equation.[2]

We cannot compare the coefficient estimates presented in Table 7 with those presented in Table A3 because of the normalization necessary to estimate the SNP models (cf. De Luca 2008). However, we use the delta method to compute the ratio of the coefficient corresponding to postdoc adviser patenting with the coefficient corresponding to the research patentability score, and we compare the value of this ratio with its Table 7 counterpart. In all five columns, the magnitudes for the ratio are very similar (though some of the estimates are fairly imprecise). We conclude that the results of the Probit models in Table 7, despite the strong distributional assumption on which they rely, imply substantive conclusions that are in line with those that can be drawn from the more involved semi-nonparametric estimation routine.

[2] The results of likelihood-ratio tests indicate that higher order terms are not necessary.

Appendix III
Impact of the Timing of Patenting Onset for the Postdoc Mentor

Our empirical approach presumes that the onset of patenting is a meaningful milestone in an academic life scientist's evolving research trajectory. In addition, our theoretical argument only applies to this setting if attitudes towards commercializable research in general, and the decision to patent in particular, can be imprinted on a junior researcher by his mentor during the time of his/her training period.

One possible falsification check is to verify that our imprinting result disappears, or is at least attenuated, when the postdoctoral mentor starts patenting only *after* the Scholar has left his/her lab to start an independent career. This is explored in Table A3, which is largely patterned after Models (1) and (2) of Table 6A. The first two columns of Table A3 use the entire Scholar sample, and include one additional covariate: an indicator variable equal to one if the postdoc mentor started patenting after the departure of the Scholar. In other words, the effect of the key covariate of interest (onset of mentor patenting before or during the training period) should now be interpreted as relative to the background patenting hazard for Scholars advised by mentors who never in the past patented, nor will patent at any point in the future. Model (1) ignores the endogeneity of matching, while Model (2) weights the model with the same inverse probability of treatment weights used in Model (2) of Table 6A. As can be observed, the magnitudes for the covariate of interest increase slightly in these new specifications (relative to what can be observed in Table 6A). Moreover, the coefficients for the post-training patenting onset are much smaller in magnitude and not significantly different from zero.

Models (2) and (4) propose a small variant of the same idea. After dropping from the sample the Scholars who were imprinted by their mentor during the training period, we simply contrast the Scholars whose mentors patent after their departure with the Scholars whose mentors have never, and will never patent. Once again, we find that the effect for the delayed onset indicator variable is small in magnitude, and imprecisely estimated.

To summarize, the specificity of the effect we estimate with respect to the precise timing for the onset of mentor patenting buttresses our claim that the estimates presented in Tables 6A correspond to a true social influence which necessitate close interaction in the mentor's lab.

References

De Luca, Giuseppe. 2008. "SNP and SML Estimation of Univariate and Bivariate Binary-choice Models." The Stata Journal **8**(2): 190-220.

Gabler, Siegfried, François Laisney and Michael Lechner. 1993. "Seminonparametric Estimation of Binary-Choice Models with an Application to Labor-Force Participation." *Journal of Business & Economic Statistics* **11**(1): 61-80.

Gallant, A. Ronald, and Douglas W. Nychka. 1987. "Semi-Nonparametric Maximum Likelihood Estimation." *Econometrica* **55**(2): 363-390.

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**(2): 229-252.

Table A1: Sample Patentability Weights in the year 2000

	Frequency of Use by Patenting Scientists	Frequency of Use by Non-Patenting Scientists	MeSH Keyword Weight
Group 1			
Telomerase	275	68	2.1407
Adenovirus E1 Proteins	12	2	2.9214
Growth Cones	10	3	12.9760
Myelin P0 Protein	10	1	79.5711
Heterocyclic Compounds with 4 or More Rings	5	1	92.3742
Group 2			
Yersinia pseudotuberculosis	15	24	0.0767
Anemia, Megaloblastic	9	11	0.1604
Vitamin K	153	166	0.4466
Small Ubiquitin-Related Modifier Proteins	3	3	0.5259
Genes, p53	508	547	0.7281
Group 3			
Satellite Viruses	1	59	0.0100
Bacteriophage P22	1	49	0.0118
Chlamydomonas	5	283	0.0124
Autistic Disorder	19	675	0.0141
Twins, Dizygotic	5	232	0.0203

Note: To illustrate the construction of MeSH keyword weights, we have chosen representative keywords in three categories. Group 1 keywords are typical of those that appear frequently in the work of patenting scientists, and infrequently in the work of non-patenting scientists. These keywords receive high patentability weights. Group 2 comprises keywords that occur frequently in the journal articles of both patenting and non-patenting scientists. Keywords in this group garner intermediate weights. Group 3 contains keywords that are very common in the research of non-patenting scientists but uncommon in the work of patenters. In consequence, these keywords receive low weight.

Table A2: Semi-Nonparametric Estimation, with and without Sample Selection

Subsample	All Scholars	US Scholars	Foreign Scholars	US Scholars	Foreign Scholars
Model	SNP Estimator (Gabler et al. 1993)			SNP Estimator with Sample Selection (De Luca 2008)	
Exclusion Restrictions	N/A	N/A	N/A	Undergrad & Postdoc University in Same State	Both Scholar & Postdoc from Same Foreign Country
	(1)	(2)	(3)	(4)	(5)
β_1: Postdoc Mentor was a Patenter	0.253†	0.696	0.571	0.266	0.750*
	(0.141)	(0.441)	(0.365)	(0.188)	(0.298)
β_2: Research Patentability Stock	0.368	0.883**	0.969*	0.296*	2.276*
	(0.354)	(0.200)	(0.425)	(0.142)	(0.908)
Female	-0.391*	-0.754*	-0.876†	-0.455†	-1.127†
	(0.187)	(0.370)	(0.475)	(0.241)	(0.584)
MD/PhD	0.342	0.738†	-0.407	0.635†	0.044
	(0.582)	(0.387)	(0.422)	(0.357)	(0.478)
Log(University NIH $)	-0.175*	-0.064	-0.346**	-0.123*	-0.459**
	(0.072)	(0.127)	(0.112)	(0.057)	(0.049)
Log(University Patents)	0.082	-0.047	0.221†	0.086	0.421*
	(0.156)	(0.292)	(0.118)	(0.092)	(0.182)
β_1/β_2 [Parametric, from Table 7]	0.773	0.963	0.586	0.856	0.563†
	(0.352)	(0.610)	(0.298)	(0.565)	(0.299)
β_1/β_2 [Semi Non-Parametric]	0.686	0.788†	0.589*	0.898	0.330†
	(0.480)	(0.443)	(0.287)	(0.743)	(0.179)
ρ				0.172	-0.113
Order of Hermite Polynomial	3	3	3	3;4	3;4
Log-pseudolikelihood	-266	-200	-60	-1,639	-495
Observations (Scholars)	489	368	121	368	121
Nb. of Postdoc Mentors	333	262	102	333	333
Potential Dyads				9,678	3,097

Note: Research patentability, university patents and university NIH funding are measured as of 2007—the end of our observation period. Models (2) and (4) include eleven cohort indicator variables. Models (3) and (5) include four cohort indicator variables. All models also include on the right hand side all the covariates in the matching equations (cf. Models (1) and (2) in Table 5, which reports estimates for the first stage matching equations). Though these covariates are included, we do not report the corresponding estimates since they are not of substantive interest in the second stage outcome equations. We also do not report the coefficients for the bases of the Hermite polynomial extension.

Robust standard errors, clustered at the postdoc mentor level reported in parentheses below each coefficient estimate.
†significant at 10%; *significant at 5%; **significant at 1%.

Table A3: Exploring the Impact of the Timing of Patenting Onset for the Postdoc Mentor

	(1)	(2)	(3)	(4)
Subsample	All Scholars		Excl. Mentors who started patenting prior or during	
IPT Weights	No	Yes	No	Yes
Postdoc Mentor started patenting before (or during) the Scholar's affiliation with his/her lab	0.672**	0.971**		
	(0.212)	(0.268)		
Postdoc Mentor starts patenting after the Scholar's departure	0.267	0.210	0.220	0.145
	(0.230)	(0.304)	(0.236)	(0.335)
Research Patentability Flow, no lag	4.173**	4.883†	6.406*	1.629
	(1.303)	(2.709)	(3.228)	(3.315)
Research Patentability Stock (lagged 1-year)	1.010*	0.220	0.119	0.044
	(0.474)	(0.686)	(0.582)	(0.661)
Female	-0.721**	-0.926**	-0.956**	-1.053*
	(0.242)	(0.293)	(0.366)	(0.469)
MD/PhD	0.460*	0.611*	1.014**	1.290**
	(0.221)	(0.297)	(0.258)	(0.331)
Log(University NIH $)	-0.220*	-0.111	-0.278†	-0.110
	(0.097)	(0.138)	(0.165)	(0.212)
Log(University Patents)	0.092	0.068	0.142	0.052
	(0.073)	(0.102)	(0.121)	(0.178)
Constant	-2.078	-4.061	-0.917	-2.911
	(1.897)	(2.620)	(2.949)	(3.579)
Log-pseudolikelihood	5250	5250	3,003	3,003
Observations (Scholars)	-676	-39,781	-343	-18,007
Nb. of Postdoc Mentors	489	489	264	264

Note: Discrete-time hazard rate models, with estimates are displayed as raw coefficients. All models include Scholar-cohort indicator variables, and a full suite of calendar-year indicator variables (not reported). Research patentability is a measure of the underlying patentability of a Scholar's research, derived from the publication and patent records of 9,000 life scientists (see Appendix I).

Robust standard errors, clustered at the postdoc mentor level reported in parentheses below each coefficient estimate.
† significant at 10%; * significant at 5%; ** significant at 1%.