Chapter 4

The Economics of University Science and the Role of Foreign Graduate Students and Postdoctoral Scholars

Grant C. Black, Indiana University South Bend

Paula E. Stephan*, Georgia State University and NBER

December 2008

We wish to thank Bill Kerr for graciously matching our database to his ethnic database, Kelly Wilkin for data assistance and Mark Regets for supplying helpful information regarding postdoctoral scholars. Bill Amis, Charles Clotfelter, Harold Shapiro, and Michael Teitelbaum made useful comments on an earlier draft of this paper, as did participants at the conference "American Universities in a Global Market," Woodstock, Vermont, October, 2008.

*Corresponding author: pstephan@gsu.edu

Abstract

We document the role that students and postdoctoral scholars (postdocs) play in university research by analyzing authorship patterns for a six-month period for articles published in *Science* having a last author affiliated with a U.S. university. The paper sample is composed of 133 papers with fewer than ten authors, for which we determine the status of all authors residing in the U.S., and for 159 papers regardless of the number of authors, for which we determine the status of first and last author. We find that 86.5% of papers—nearly seven out of eight (133 paper sample)--have either a current postdoc or a student as one of the authors. We find that 103 of the 138 first authors who are in the U.S. and whose position is known are either a postdoc or a student (74.6%).

We identify the ethnicity of authors, drawing on the ethnic-name database created by William Kerr (2008). We find that 59.1% of postdoc authors are neither English nor European and that 39.6% of the graduate student coauthors have neither English nor European names. At the paper level, we infer that 70 of the 133 papers (53%) have a foreign student or postdoc as a coauthor. We infer that almost 60% of the graduate student first authors are foreign and that non-citizens make up slightly more than 54% of the postdocs who are first authors. We conclude that international graduate students and postdocs are not only important in staffing university labs; they play lead roles in university research.

Section I: Introduction

Universities play an important role in the production of knowledge in the United States, authoring nearly 75% (fractional counts) of scientific and engineering articles written in the country.[1] Within the university, research is often performed with the assistance of graduate students, postdoctoral scholars (postdocs) and staff scientists, many of whom are foreign-born and foreign-educated. Currently, for example, over 45% of graduate students enrolled in science and engineering (S&E) are foreign-born and approximately 60% of postdocs are on temporary visas.

This chapter documents the presence and importance of graduate students and postdocs in U.S. academic science. We are particularly interested in the role of the foreign-born and foreign-trained. We begin by examining the importance of teams in university research and then provide an overview of the way in which university research is financed and structured. Next we summarize trends in the number and proportion of foreign-born graduate students and postdocs studying in the United States. To explore the role that postdocs and graduate students play in the production of knowledge we examine articles published in *Science* during a six-month period in 2007 and 2008 which have a U.S. academic-based scientist as the last author. Through web searches we are able to determine the status (postdoc, graduate student, staff scientist, or faculty) of virtually all U.S. coauthors. We also examine the ethnicity of the coauthors by applying an ethnic-name database and infer nativity from ethnicity. We conclude in section VI, summarizing our results and discussing their implications U.S. universities and for the research enterprise.

Section II: The Importance of Teams

Research is rarely done in isolation, especially research of an experimental rather than a theoretical bent (Mary Frank Fox 1991). Scientists work in teams. One way of seeing how team size and collaboration have changed is to examine trends in co-authorship patterns among papers with one or more authors from a "top 110" U.S. university. James Adams et al. (2005) find that for this group, the mean number of authors per paper increased from 2.8 to 4.2 for the 18-year interval, ending in 1999.[2] The rate of growth was greatest during the period 1991-1996 when use of E-mail and the internet was rapidly accelerating.

The growth in authorship is due to a rise in the number of people working on a project within a given university as well as to an increase in the number of institutions—especially foreign institutions--collaborating on a research project. During the period 1988 to 2003, the number of addresses associated with a U.S.-authored article grew by 37% and the number of foreign addresses more than tripled (National Science Board 2006, Table 5-18). Despite this impressive increase, the growth in co-authorship is fueled more by an increase in the number of authors working at the same university than an increase in collaboration across universities, as evidenced

[1] Universities also play a considerably smaller--though growing--role in invention. In 2005, universities produced 3.7% of all patents awarded to U.S. owners. The underlying count of 2725 represents a 50% increase over the number awarded to universities 10 years earlier. (National Science Board, 2008, Appendix Table 5-40).
[2] The study is restricted to articles in science and engineering having one or more authors from a top-110 U.S. university.

by the fact that during the same period the number of names on an article grew by more than the number of addresses on an article (50% vs. 37%).

Several factors contribute to the increased role that collaboration plays in research. First, the importance of interdisciplinary research and the fact that major breakthroughs often occur in emerging disciplines encourage collaboration. Systems biology, which involves the intersection of biology, engineering and physical sciences, is a case in point.[3] By definition, no one has all the requisite skills required to work in the area; researchers must rely on working with others. Second, and related, researchers arguably are acquiring narrower expertise over time in order to compensate for the educational demands associated with the increase in knowledge (Benjamin Jones 2005). Narrower expertise, in turn, leads to an increased reliance on teamwork for discovery. Third, the rapid spread of connectivity, which began in the early 1980s with the adoption of BITNET by a number of universities and accelerated in the early 1990s with the diffusion of the internet, has decreased the costs of collaboration across institutions. (Ajay Agrawal and Avi Goldfarb 2008; Anne Winkler, Sharon Levin, and Paula Stephan 2008). Another factor that fosters collaboration is the vast amount of data that is becoming available, such as that from the Human Genome Project (and the associated GenBank database). Although that is probably the best known, many other large data bases have recently come on line, such as PubChem, which as of this writing contained over 18,000 recorded substances, and the Worldwide Protein Data Bank (wwPDB), a world-wide depository of information regarding protein structures.[4] The practice of sharing research materials also leads to increases in the number of authors appearing on an article.

Increased complexity of equipment also fosters collaboration.[5] By way of example, in the *Science* database that we have assembled for this chapter, four co-authors are identified on web pages as electron microscopists. Andy Barnett, Richard Ault, and David Kaserman (1988) suggest two other factors that lead persons to seek coauthors. One is the desire to minimize risk by diversifying one's research portfolio through collaboration; the other is the increased opportunity cost of time. An additional factor is quality. The literature on scientific productivity suggests that scientists who collaborate produce "better" science than do individual investigators (Stefan Wuchty, Jones and Brian Uzzi (2007); Frank Andrews (1979) and S. M. Lawani (1986)). Some of the factors encouraging collaboration are new (such as connectivity) but growth in the number of authors on a paper is not. Wuchty, Jones and Uzzi (2007) find that team size has grown in all but one of the 171 S&E fields studied during the past 45 years.

Much university research occurs in a lab setting. How these labs are staffed varies across countries. For example, in Europe research labs are often staffed by permanent staff scientists, although increasingly these positions are held by temporary employees (Stephan 2008). In the United States, while positions such as staff scientists and research associates exist, the majority of scientists working in the university lab are doctoral students and postdocs. Stephan, Grant Black and Tanwin Chang's study (2007) of 415 labs affiliated with a nanotechnology center

[3] Systems biology studies the relationship between the design of biological systems and the tasks they perform.
[4] The Large Hadron Collider (L.H.C.) at CERN will create vast amounts of data. According to Elizabeth Kolbert (2007, p. 74), "If all the L.H.C. data were burned onto disks, the stack would rise at the rate of a mile a month."
[5] At the very extreme are the teams assembled to work at colliders. CERN's four colliders have combined team size of just under 6,000: 2520 for the Compact Muon Detector (C.M.S.), 1800 for the Atlas, 1000 for ALICE and 663 for LHCb. (Dennis Overbye 2007).

finds that the average lab has 12 technical staff, excluding the principal investigator (PI). Fifty percent of these are graduate students; 16% are postdocs and 10% are undergrads.[6] Some labs are quite large. A case in point is the Susan Lindquist lab at MIT which has 36 members (excluding Lindquist herself)—20 postdocs, 7 graduate students, 1 visiting scientist, 1 staff scientist, 3 technicians, and 4 administrators.[7]

This way of staffing labs has been embraced in the U.S. for a variety of reasons. Pedagogically, it is an efficient training model. It is also an inexpensive way to staff laboratories. Moreover, and as faculty are not abashed to note, it provides a source of "new" ideas, especially given the relative young age of doctoral students and postdocs. To quote Trevor Penning, while serving as the Associate Dean for Postdoctoral Research Training at the University of Pennsylvania School of Medicine, "A faculty member is only as good as his or her best postdoc" (Penning 1998). In addition, funding is often more readily available for pre-doctoral and postdoctoral students than for staff scientists. The typical National Institutes of Health (NIH) grant, for example, supports both types of training as do many other forms of grants. At least from the perspective of the National Science Foundation (NSF), it has been a conscious policy to fund students. Rita Colwell, the Director of NSF from 1998-2004, said in an interview with *Science* that "In the 1980s, NSF asked investigators to put graduate students on their research budgets, saying it preferred to fund graduate students rather than technicians" (*Science* 1998). There is also the added advantage that postdocs and graduate students, with their short tenure, provide for more flexibility in the staffing of laboratories than do permanent technicians.

This model for staffing labs has undoubtedly contributed to the U.S.'s eminence as a training center for both native and foreign-born students. It provides not only a hands-on learning experience but also financial support for graduate study and postdoctoral work, something that many other countries cannot provide.

Section III: The Structure and Financing of University Labs and Research Groups

Labs at U.S. universities "belong" to the faculty PI, if not in fact, at least in name, as is readily seen by the common practice of naming the lab for the faculty member. A mere click of the mouse, for example, reveals that all of the 26 faculty at MIT in biochemistry and biophysics use their name in referring to their lab.[8] Sometimes, as in the case of the Nobel laureate Philip Sharp, lab members and former members are referred to using a play on the PI's name—in this case "Sharpies."[9]

It is common practice for labs to maintain web pages, discussing research focus, publications, funding, etc. Most pages provide pictures of people who work in the lab, sometimes in a group

---

[6] Approximately a third of the PIs were affiliated with departments of engineering, a third with departments of chemistry and the remainder with departments of physics.

[7] The Linquist lab is large compared to the labs of her colleagues at MIT in biochemistry and biophysics, which have an average of 6.3 graduate students (median of 7) and average of 5.25 postdocs (median of 5).

[8] Details regarding research and staffing are available for 17 of the 26 via lab web pages. Three other faculty have web pages for their labs that are not fully developed. For the other six one can find reference to the name of their lab when searching the internet.

[9] In a similar manner, graduate students and postdocs working in Alexander Pines' lab at Berkely, are referred to as "pinenuts" and alumni are referred to as "old pinenuts" (http://waugh.cchem.berkeley.edu/people/).

shot; in other instances individual shots are included. While most pictures are of a traditional nature, it is not uncommon for the photos to be on the humorous side or slightly over the edge.[10]

Lab pages also traditionally provide links to "people" or "personnel" which include a list of everyone working in the lab, from undergraduate students to graduate students, postdocs, and staff scientists. Technicians and administrators are also listed. Some pages list alumni of the labs.

Research is expensive. Personnel costs alone for a small-to-medium lab, composed of three GRAs, one postdoc, one technician and the PI are approximately $210,000 including salaries and benefits but excluding the cost of buying out the PI's time for research. Each additional graduate student adds approximately $37,000; each additional postdoc adds approximately $52,000 [11]. Additional expenses include the cost of supplies and equipment. For research in the life sciences, supplies can easily average $18,000 per year per lab member, or add another $108,000 to the costs for a lab of six including the PI (Adelle Pelekanos 2008). This excludes the cost of animals, which can be quite expensive. An off-the-shelf mouse costs $50 (U.S.) in 2006; transgenic mice, bred for the study of a specific disease, cost in the neighborhood of $2000[12] and some carry price tags of $15,000 (David Malakoff 2000). With the large number of mice in use (over 13,000 are already published), the cost of mouse upkeep becomes a significant factor in doing research. U.S. universities, for example, charged from $.05 to $.10 per day per mouse (mouse per diem) in 2000 (Malakoff 2000).[13]

Equipping a lab adds considerably more to expenses. Pelekanos (2008) estimates that start-up equipment for a lab in the life sciences costs about $60,000. But equipment can cost much more than this. A microscope used for research in nanotechnology can cost $750,000. (http://www.unm.edu/~market/cgi-bin/archives/000132.html). A sequencer, such as Applied Biosystems' 3730 model, costs approximately $300,000. One reason research in certain fields is conducted outside the university relates to the extremely high cost of equipment and the indivisible nature of this equipment. At the extreme are costs associated with building and running an accelerator. The 27-kilometer-long Large Hadron Collider (LHC) which has recently come on line at CERN costs approximately $8 billion; the Spallation Neutron Source (SNS) at

[10] The White Lab webpage (Christina White, Department of Chemistry, Universityof Illinois) depicts White seated on a stone throne, engulfed in flames and surrounded by 12 of her graduate students, one of whom is sporting horns. See http://www.scs.uiuc.edu/white/

[11] The graduate student amount includes stipend, fringe benefits and tuition and is based on the amount allowed by NIH for the Ruth Kirstein NRSA fellowship for FY 2007. Many institutions pattern their support for other students on the Kerstein fellowship. The postdoc figure includes stipend and fringe benefits; it is the average paid under NIH guidelines for postdocs with varying experience. The fringe amount comes from Pelekanos (2008), as does the cost estimate for the technician.

[12] As an example, the mouse used to study insulin is available at the Jackson Labs for $1900. (http://jaxmice.jax.org/strain/004937_3.html#price).

[13] This cost of mouse up-keep can rapidly add up. Irving Weissman of Stanford University reports that before Stanford changed its cage rates he was paying between $800,000 and $1 million a year to keep the 10,000 to 15,000 mice in his lab. Costs for keeping immune-deficient mice are far greater (on the order of $.65 per day), given their susceptibility to disease.

Oak Ridge National Laboratory in the U.S. cost $1.41 billion. (Science vol 312, 5 May 2006; p. 675).

In order to get started on an independent research career, faculty usually receive resources from the dean at the time they are hired. Included in these start-up-packages are funds for equipment and stipends to hire graduate students, staff scientists and postdocs. Also, and of crucial importance in the lab sciences, they are assigned space. Ronald Ehrenberg, Michael Rizzo and G.H. Jakubson (2003) have surveyed U.S. universities regarding start-up packages. They find that the average package for an assistant professor in chemistry is $489,000; in biology it is $403,071. At the high end it is $580,000 in chemistry; $437,000 in biology. For senior faculty they report start-up packages of $983,929 in chemistry (high-end is $1,172,222); and of $957,143 in biology (high end is $1,575,000).

Start-up packages are exactly that. After several years, the faculty member becomes responsible for procuring the resources for the lab.[14] Faculty do this primarily through the grants system, writing proposals and, if successful, receiving funds from Federal agencies and private foundations.[15] Faculty also receive support for their labs from industry. One exception to the rule is that faculty sometime host postdocs who have received funding through a fellowship or graduate students supported on training grants (awarded to the department) who work (on a rotation basis) in a faculty lab.[16] Increasingly, faculty are expected not only to cover the research expenses of the lab through grants and contracts, but also to cover a portion of their own salary. Indeed, it is becoming increasingly common for faculty in tenured positions at U.S. medical institutions to be required to procure a portion of their salary from grants.[17]

Grant applications and administration divert scientists from spending time on research. A 2006 survey of U.S. scientists found that scientists spend 42 percent of their research time filling out forms and in meetings; tasks split almost evenly between pre-grant (22%) and post-grant work (20%). The tasks cited as the most burdensome were filling out grant progress reports, hiring personnel and managing laboratory finances (Sam Kean 2006).

Organizationally, PI-labs in the United States are structured as pyramids. At the pinnacle is the faculty principal investigator. Below the PI are the postdocs; below the postdocs are graduate students and undergraduates. Some labs, as we note, also have scientists who have completed postdoctoral training in this or another lab and are hired in such non-tenure track positions as staff scientists and research faculty. The pyramid analogy does not stop here, however. In certain ways the research enterprise itself resembles a pyramid scheme. In order to staff their labs, faculty recruit PhD students into their graduate program with funding and the promise of

[14] Start-up packages have been known to have unintended consequences. A chair of a department recounted to one of us that new hires in the department "hoard" their start-up funds, postponing going up for NIH funding until a tenure decision has been made.
[15] The primary sources of federal funds are The National Institutes of Health (NIH), the Department of Energy (DOE), the Department of Defense (DOD) and, to a lesser extent, the National Science Foundation (NSF).
[16] MIT, for example, distinguishes between postdoctoral associates and postdoctoral fellows. The former are supported through grants that faculty have procured at MIT; the latter have received fellowships or stipends to work with a faculty member at MIT.
[17] A survey of medical schools found that tenure is accompanied with no financial guarantee for 35% of basic science faculty and 38% of clinical faculty (Sarah Bunton and William Mallon 2007).

interesting research careers (Stephan and Levin 2002). Upon receiving their degree it is mandatory for students who aspire to a faculty position to first take an appointment as a postdoc. Postdocs then seek to move on to tenure-track positions in academe. The Sigma Xi study of postdocs, for example, found that 72.7% of the postdocs who were looking for a job were "very interested" in a job at a research university and 23.0% were "somewhat interested" (Geoff Davis 2005). In recent years, however, the transition from postdoc to tenure track has been slowed as the number of tenure-track positions has failed to keep pace with the increase in supply.

Faculty not only staff labs with graduate students and postdocs. They actively recruit and select the students who work in their lab. Unlike admission decisions to PhD programs, however, which generally occur at the department level, decisions regarding staffing are usually made by the faculty member who, in effect, is paying for the student.

Not surprisingly, given the role faculty play in staffing decisions, networks or what may more accurately be described as "affinity effects" appear to play a role in staffing. Esra Tanyildiz (2008) has studied paired labs in 82 departments of engineering, chemistry, physics and biology. In each case she matches a lab directed by a "native" PI (as established by name and undergraduate institution) to a lab directed by a foreign PI, either of Chinese, Korean, Indian or Turkish background. She then studies the graduate student composition of the labs, assigning nationalities to the students based on the common-name methodology used by William Kerr (2008). She finds significant differences in the role that ethnicity plays in staffing. The mean paired difference in the percent of Chinese students in a lab directed by a Chinese PI versus a lab in the same department directed by a "native" U.S. faculty is 37.8%; that for Koreans is 29.0%; that for Indians is 27.1%; that for Turkish is 36.3% (very small sample). When she compares labs directed by natives to non-natives from one of these four groups the mean paired difference is 28.9%. Clearly clustering by ethnicity occurs in labs. Tanyildiz also finds that affinity effects are more common in "bottom"-ranked departments; less common in "top" departments.[18]

Not all university research is organized around labs directed by faculty. In the earth sciences, for example, scientists often do not work in a lab setting. In instances of "big" science (such as experimental high energy physics, cosmology and astrophysics), research is often organized around equipment such as a telescope or an accelerator. Often this equipment is located off site, sometimes at national labs, such as the Stanford Linear Accelerator (SLAC), Fermi Lab or the Lawrence Berkeley National Laboratory; sometimes it is located at international labs, as in the case of CERN.[19]

The absence of a lab on campus does not mean that graduate students and postdocs are absent nor that faculty lack a role in choosing who works with them or their group. In many instances of "big" science it is not uncommon for the group to have a web page named for its research focus—e.g. the Caltech Observational Cosmology Group (with the goal of developing novel instruments)-- which lists the research focus and links to faculty, postdocs, graduate students,

---

[18] Using NRC rankings, she finds that the mean difference is 25.9% in "top" departments; 35.9% in "middle" departments and 53.2% in "bottom" departments. These calculations do not include mean differences between native students in native labs vs. native students in non-native labs.

[19] By way of example, physicists at the California Institute of Technology routinely work at telescopes in New Mexico and Hawaii, and at SLAC. They also are playing key roles in developing the Compact Muon Solenoid (CMS), one of the two large general purpose particle physics detectors that will come on line at CERN in 2008.

visitors and staff working in the group. Individual physicists in the group also maintain a web page, but physicists working in the area don't have labs with their name attached to the lab. But it is not only "big" physics that presents itself as a group. It occurs in other areas as well. For example, the Experimental Condensed Matter Research Group at Cal Tech keeps a group web page, as does the Spin Group and the Infrared Arm Group, to give but several examples. Moreover, it is not just experimentalists who speak of their group. Numerous examples can be found where theoretical physicists talk of their "group" on the web even though members of the group may be working by themselves.

Section IV: Trends in the Production of PhDs and Postdoctoral Students by Visa Status

PhD Awards

In the early 1980s, approximately 12,000 PhDs were awarded annually in the United States in science and engineering. By the late 1990s the number had grown to approximately 20,000; by the mid-2000s it had increased to over 23,000, roughly doubling over the entire period. This substantial increase, however, masks wide differences in enrollment patterns among U.S. citizens and non-citizens shown in Figure 1 for the period from 1980-2006.[20]

We see that the number of U.S. students receiving S&E PhDs grew by only 30 percent during the period. Moreover, virtually all of the growth that occurred was among women students. The number of PhDs awarded to citizen women increased by 170 percent from 1980 to 2006, while the number of U.S. males receiving PhDs in science and engineering changed little during the period.

In contrast, the number of temporary residents receiving PhDs grew considerably, with the increase accounting for more than 67 percent of the growth in PhD production in the United States. Permanent residents played a much smaller role, contributing only another 2.3 percent.[21] Growth of the foreign-born was especially strong during the mid-1980s to mid-1990s and again beginning in 2003. The number of foreign-born declined somewhat during the late 1990s and early 2000s.

Almost half of non-citizen PhDs come from the three countries of China, South Korea and India (Thomas Hoffer et al 2006, Table 12). China's role has become so dominant that Tsinghua University and Peking University recently surpassed the University of California, Berkeley, as the most likely undergraduate institution for those earning a PhD at a U.S. institution, regardless of nativity, between 2004 and 2006.[22]

[20] For these data, science and engineering excludes medical and social sciences, *citizen* means a native or naturalized citizen of the United States, *permanent resident* means a non-citizen immigrant holding a green card indicating permanent residency in the United States, and *temporary resident* means a non-immigrant visa holder planning to remain in the United States temporarily (such as a student or temporary worker).

[21] The exception was the large increase in permanent residents in the early-to-mid 1990s which, along with the accompanying decrease of temporary-resident recipients, reflects the passage of the Chinese Student Protection Act that permitted Chinese nationals temporarily residing in the United States to switch to permanent-resident status.

[22] The calculations are for degrees awarded between 2004 and 2006 (Jeffrey Mervis 2008). The University of California, Berkeley is now in third place, followed by South Korea's Seoul National University, Cornell University and the University of Michigan, Ann Arbor.

The growth in the number of temporary residents receiving S&E PhDs has been dramatic across most fields, as seen from Figure 2. The percent of PhD recipients who were temporary residents at the time the degree was received more than doubled from 1980 to 2006 in the fields of math and computer sciences, the physical sciences, geosciences, and life sciences. These high growth rates dramatically increased the proportion of foreign-born receiving degrees in certain fields. For example, in math and computer sciences, the proportion rose from 19 percent to over 51 percent; in the life sciences, from approximately 12 percent to 27 percent. Growth in the number of degrees awarded to the foreign-born was lower in engineering, where temporary residents have long received a considerable share of degrees. By 2006 almost 60 percent of all PhDs in engineering were awarded to individuals on temporary visas.

The fields of the geosciences and the physical sciences owe most of their growth during the period to the large influx of foreign students. In the former, for example, temporary residents made up over 96 percent of the growth in number of degrees; in the latter, they comprised 92 percent. In terms of magnitude of change in the number of temporary residents receiving PhDs, the greatest growth took place in the fields of engineering and the life sciences. In 1980 the number of engineering PhDs awarded to temporary residents was 861; by 2006 that number had risen to almost 4,300. In the life sciences, almost 620 temporary residents received PhDs in 1980 compared to over 2,400 in 2006. The latter was undoubtedly spurred by increased resources made available for the support of graduate students which resulted from the doubling of the NIH budget in the late 1900s and early 2000s.

Recent Trends in Graduate Student Enrollments

Data concerning the number of PhDs awarded reflect conditions and decisions made six to seven years prior to the award date. Thus, the increases that we have documented were put in motion long before 9/11. Following 9/11 considerable attention was focused on the observed decline in applications and admissions of international graduate students observed and what this would mean for graduate education in the United States. For example, between 2003 and 2004 graduate applications across the board declined by 28%, admissions by 18 percent, and enrollments by 6 percent (National Academies 2005, p. 31).[23] These concerns have been somewhat mitigated by the modest rise in the enrollment of international graduate students experienced recently. For example, according to The Survey of Graduate Students and Postdoctorates in Science and Engineering for 2006, first-time, full-time enrollment for temporary residents in graduate science and engineering programs rose 16.4 percent between 2005 and 2006, compared to a meager 1.7 percent for U.S. citizens and permanent residents (National Science Foundation WebCASPAR). It remains to be seen whether this turnaround will continue. Clearly, enrollment patterns are affected not only by U.S. visa policy but also by opportunities for study outside the U.S., which in recent years have been increasing.

Postdocs

Estimating the population of scholars working in postdoctoral positions in the United States is complex and leads to different measures based on the methodology that is employed. Thus

[23] Comparable figures for engineering are -36.0, -24.0, and -8.0; for the life sciences, -24.0, -19.0 and -10.0; and for the physical sciences, -26.0, -17.0, and +6.0. Data come from the Council of Graduate Schools (National Academies 2005, p. 31). It should be noted that application and admission data “double count” to the extent that students apply and are admitted to multiple programs.

estimates must be read with caution. Complications arise from several factors, including survey sampling frameworks that omit or do not easily identify some postdocs, especially in non-academic sectors, or those with doctorates from foreign institutions; the timing of survey data collection that can miss increasingly migratory S&E PhDs; exclusions and discrepancies surrounding some S&E occupations in certain standard surveys; and institutional difficulties in identifying workers as postdocs and by visa status (National Science Board 2008; Regets 2007). By way of illustration, Mark Regets (2007) offers the anecdotal example of officials at a major research university who expressed confidence in their ability to identify all temporary-visa postdocs at their institution on the assumption that only J-1 visas were used for postdocs. It was later discovered that Labor Condition Applications – the first step in the H1-B visa process – had been filed by the university for several hundred "postdoctoral appointments." There is also the issue of job title. It is not uncommon for individuals who are esentially postdocs to be called by another title, such as research scientist. Classification problems such as this mean that many postdocs go uncounted because of a wide range of measurement issues.[24]

Figure 3 shows the number of postdocs working at academic institutions in science and engineering in the United States from 1985 to 2006, based on the Survey of Graduate Students and Postdocs.[25] We see that in 1985 there were slightly more than 16,000 postdocs at academic institutions. Within a decade, that number had grown to over 25,000, and by 2006 the number of postdocs had surpassed 34,000 – an increase of 110 percent from 1985 to 2006. Growth was steady through the early 1990s and continued to increase in the remainder of the 1990s but at a slower rate. The number of postdocs declined slightly in 2001 but has since increased, particularly in 2002-2003.[26]

Growth in the number of postdocs has been fueled largely by scholars coming from abroad. The number of postdocs with temporary-resident visas (identified as foreign postdocs in Figure 3) almost tripled between 1985 and 2006, rising from 7,032 in 1985 to 20,521 in 2006. While in 1985 temporary residents made up just over 43 percent of all postdocs, by the 2000s they comprised approximately 60 percent of all academic postdoctoral scholars, reaching a peak of 61 percent in 2001. In contrast, the number of postdocs who are U.S. citizens or permanent residents (identified as US postdocs in Figure 3) grew by less than half during the same period. Indeed, the difference is so dramatic that from 1996 to 2006 alone, the number of temporary-resident postdocs grew by over 52 percent – more than the rate for U.S. citizens and permanent residents over the entire 1985-2006 period. The difference is so pronounced that temporary-resident postdocs grew at an annual rate of 5.2 percent compared to only 1.9 percent for native and permanent-resident postdocs during the period. Tightened visa-security measures may have

---

[24] NSF is acutely aware of the many problems involved in measuring postdocs and is in the process of designing a new methodology to measure the number and characteristics of postdoctoral scholars in the U.S.

[25] These data are also based on science and engineering--excluding the medical and social sciences--and account only for postdocs identified by surveys of academic institutions with graduate programs in science and engineering. Although the majority of postdoctoral positions are at academic institutions, postdocs can also be found in other sectors. Using the 2006 Survey of Doctorate Recipients, Hoffer *et al* (2008) estimate that 75 percent of postdocs in science, engineering, and health fields were at educational institutions, 12 percent were in government, 11 percent were at for-profit or non-profit organizations, and 2 percent were at other types of institutions.

[26] The number of postdocs depends not only upon the propensity to take a postdoc but also upon the duration of the postdoc period of training. Stephan and Jennifer Ma (2005) show that not only the propensity to take a postdoc but also the duration of the postdoc training period relates to the state of the academic labor market, suggesting that the postdoc position can become a "holding tank" where people wait for better market conditions.

contributed to the slowdown in temporary-resident postdocs since 2003. In 2001, less than 8 percent of J-1 visa applications were denied; in 2003, almost 16 percent were refused (Regets 2005).[27]

While many postdocs earn their PhD in the U.S. prior to applying for a postdoctoral position, a remarkable number receive their PhD training outside the U.S. and come to the U.S. to take a postdoctoral position. Indeed, Regets (2005) estimates that almost five out of ten academic postdocs in the United States earned a doctorate in another country. Moreover, four out of five postdocs with temporary visas earned their doctorate outside the United States.[28]

Figure 4 shows the distribution of foreign S&E postdocs by field for the period 1985-2006. The dominant role of the life sciences is striking. For example, in 2006, close to six out of every ten postdocs on a temporary visa were in the life sciences. In terms of raw numbers, the figure shows that the life sciences also experienced the greatest growth in the number of postdoctoral positions held by those on temporary visas, going from 3,341 in 1985 to 11,694 in 2006. By way of contrast, the increase in engineering was 2,193; that in the physical sciences was 1,853. The magnitude of the change in the life sciences is likely a result of the increased demand for postdocs in the field occasioned by the doubling of the NIH budget in the late 1990s and early 2000s.The fastest growth of postdocs on temporary visas occurred in the geosciences, where the number increased by a factor of more than six times. In math and computer sciences, the figure grew by over 300 percent. The number of temporary-resident postdocs grew by over 300 percent in math and computer sciences, 250 percent in the life sciences, 240 percent in engineering, and only 74 percent in the physical sciences.

Section V: Authorship Patterns in *Science*

To examine the contributions of postdocs, graduate students and undergraduates to research in academe, we collected data on the authors of articles published in *Science* from November 2, 2007 to May 2, 2008.[29] We focused on papers in the Research Articles and Reports sections of the journal. In many fields of science the last author is the principal investigator; while other rules or variations exist in terms of author order, we apply this common convention to our

[27] Foreign postdocs have traditionally been in the U.S. on either a J or an H visa, with some on F-1's for one year of optional practical training. The Sigma Xi survey (with a non-representative sample) found that 51% of foreign postdocs were on J's; 41% on H's and 3% on F-1s; the remaining 4% were on "other" visas (http://www.sigmaxi.org/postdoc/by_citizenship/). See also Davis (2005). Mark Regets reports (informal correspondence) that there is some evidence that the proportion on H-1B visas has been growing, based on the number of Labor Condition Applications that explicitly contain the search string "postdoc." The number on F-1 visas is expected to grow, because optional practical training time was recently increased from 12 months to 29 months for most S&E advanced degrees.

[28] These estimates are based on a comparison of counts from the NSF Survey of Doctorate Recipients and the NSF Survey of Graduate Students and Postdoctorates in 2001. For example, in 2001, 17,900 academic postdocs with temporary visas were reported through the Survey of Graduate Students and Postdoctorates, while only 3,500 postdocs with temporary visas were reported in the Survey of Earned Doctorates, which only collects data on doctorates earned in the United States. Regets attributes the difference in these counts to postdocs with PhDs earned outside the United States.

[29] We code 22 issues for the six-month period. The four issues not coded are November 23rd and 30th and December 7th and 14th, 2007.

analysis to determine if a paper has a U.S. origin.[30] We further restrict the analysis to papers with a last author affiliated with a U.S. academic institution, given our interest in studying science in academe.

We chose *Science* because of its multidisciplinary nature (the journal devotes 40% of its space to the physical sciences and 60% to the life sciences) and its position as a leading, if not the leading, journal in science. Moreover, and as is to be expected, the journal is highly selective. In 2007 the journal published 817 of the 12,450 articles that it received (6.6%); 461 of these (56.4%) had a first author from the U.S. (Chiara Franzoni, Giuseppe Scellato and Paula Stephan 2008).

For each paper we record the broad field related to the subject of the research, the number of authors, the name of each author, institutional affiliation as listed in the article, and the location (country) of the listed institutions.[31] We collect additional information from internet searches on the authors, including the academic position of an author and whether the author is affiliated with the same lab as the last author. In some instances this information is obtained from the last author's webpage but more commonly it comes from the web page for the last author's lab. Such web pages are particularly useful in identifying postdoctoral students, graduate students, undergraduate students and staff scientists and technicians working in the lab. In cases where information could not be found (most frequently regarding the position of an author and whether the author has an affiliation with the last author's lab) missing values were coded. We believe this approach provides an accurate count of the number of students involved in the research. The count of postdocs is likely to be downward biased, however, since some postdocs, as noted earlier, have job titles that make it difficult to distinguish them from staff scientists. We thus view the postdoc count as a lower bound.

For papers having a last author affiliated with a non-U.S. academic institution, we code only the field, number of authors, and location of the last author. Data on the 51 papers for which the last author is affiliated with a nonacademic institution, such as a private business, non-profit organization, or government agency, were not collected regardless of country of last author.[32] All told, data on 267 academic papers was collected. Of these, 159 had a last author at a U.S. academic institution and 108 at a foreign academic institution. The distribution of papers by last author affiliation is summarized in Table 1.

The median number of authors for U.S. academic papers is 5; the minimum is 1 and the maximum is 71. Web pages could be found either for the last author's lab or for the last author in all but one case.

The last authors come from 69 different U.S. academic institutions. The largest numberof last authors (16) come from either Harvard or Harvard Medical School; 9 come from U.C. Berkeley and 8 from Stanford. Six institutions have scientists publishing five articles during the six-month

[30] Had we instead used the country of the first author to determine origin, the sample would have had 150 papers rather than the159 papers we analyzed.

[31] For publications with ten or more authors (26 of the 159 U.S. papers), only the first and last authors were recorded.

[32] Of these 51 papers, 36 have a U.S. address; 4 have a German address; 3 have a Japanese address. The remaining 8 are authored by individuals in Australia (1), Canada (2), France (2), Iceland (1), The Netherlands (1) and the UK (1).

time period. The institutions are: California Institute of Technology, Johns Hopkins, MIT, University of Michigan-Ann Arbor, University of Washington and Yale. Several lesser known institutions are represented, such as Minnesota State University Mankato, Franklin and Marshall College, the University of Southern Mississippi, and Georgia Southern University.

The distribution of U.S. academic articles by area is given in Table 2. The distribution mirrors *Science's* overall editorial practice of having a 60/40 split between the life and physical sciences. The median number of authors is highest in genetics; it is lowest in chemistry and neurology. The most authors were on a paper in biology.

Authorship Patterns

We first discuss the data for the 133 articles having 9 or fewer authors; we then summarize the data for all U.S. papers regardless of number of authors, focusing on an analysis of first and last author.

The data for articles with 9 or fewer authors is summarized in Table 3. Of the 648 authors, 585 lived in the U.S.[33] We could find information on the position of 550 of these (94.0%). Of these, 123 were postdocs (22.4%); another 108 (19.6%) were graduate students, 8 (1.5%) were undergraduate students and 8 (1.5%) were students or postdocs, specific status not known. An additional four were alumni of the program, having either been a graduate student or a postdoc.[34] The postdoc count is, as we noted above, an undercount in all likelihood given that some postdocs have titles that make it difficult to distinguish them from staff scientists. When the categories are combined, we find that almost one out of two authors (45.6%) was a postdoc, a student or a recent alum of the program.[35]

Of perhaps more interest to our study is the fact that 115 (86.5%) of this class of papers had either a current postdoc or student as one of the authors. Five of the 18 papers that have neither postdocs nor students as coauthors are either singly authored or have only one U.S. author. Two of the eighteen papers were in the field of astronomy; three in earth sciences and two in material sciences. The field least likely to have either a postdoc or a student as a coauthor is astronomy (two for two) followed by material science (with two of the seven papers having neither a postdoc nor a graduate student author) and earth sciences (three of the thirteen had neither a postdoc nor a graduate student author). The fields most likely to have a postdoc or a graduate student as a coauthor are biochemistry, genetics, nano-related, and chemistry and chemistry-related. Indeed, all of the 42 papers published in these four areas (with less than 10 authors) had one or more graduate students or postdocs as co-authors. Fields not far behind are biology (27 of 28 papers) and physics (11 of 12).

[33] In several cases the individual is listed with two affiliations; one is in the U.S.; the other is outside the U.S. In this case we count the individual as being in the U.S.

[34] This is an undercount of alums given that not all web pages list alumni of the program and in some instances faculty do not keep webpages.

[35] A third of the postdocs are the only postdoc author on the paper; another third share authorship with one other postdoc; and another third share authorship with more than one other postdoc. Two papers have 5 postdocs as authors; 12 papers that three postdoc authors.

All but 27 of the papers with less than 10 authors have one or more authors working in the same lab as the senior U.S. author.[36] These patterns differ by field. The earth science papers are the least likely to have another individual working in a lab with the senior author (6 out of 13 earth science papers have no overlap in address). By way of contrast, 90% or more of the articles in biochemistry, genetics, nano-related areas, neurology and physics have at least one co-author working in the same lab as the senior author.

Only eleven of the 115 papers with a postdoc or graduate student as a coauthor have no authors that are in the same lab as the senior U.S. author. But it does not follow that all of the postdoc and student authors work in the lab of the last author. In a number of instances they work outside this lab, either with someone else at the same university or with someone in another university.

First and last authorship patterns are summarized in Table 4 for all U.S. academic articles appearing during the six-month period. The role of postdocs and students is especially striking when one looks at first-author position, a position of particular importance since in most fields the first author does the "heavy lifting," contributing the most to the article.[37] Fully 73.7 % of the 137 first authors who are U.S. and whose position is known are either a postdoc or a student. Seven of the last authors are either a postdoc or a student. Four of these papers are in the area of earth science, further confirmation that the earth sciences are organized somewhat differently than the other fields we are looking at. Two of the papers that have postdocs as last author are in biochemistry. One paper in physics has an undergraduate student, Jacob Simones from the Minnesota State University Mankato, as the last author. The article has 10 other authors, including his undergraduate advisor. Simones appears to have done related work during the summer of 2006 as a research experience for undergraduates (REU) at Minnesota State University funded by NSF.[38]

Ethnicity of U.S. Authors

Ideally, we would like to know the citizenship status or birth origin of the students and postdoc co-authors. Short of fielding a survey this is not possible, because most postdocs and students do not put CV's on the web. Instead, we follow the approach used by Bill Kerr, drawing on the same ethnic-name database that he used to identify the ethnicity of U.S. inventors (Kerr 2008).

Specifically, ethnicity is identified using data that Kerr obtained from the Melissa Data Corporation.[39] The Melissa data is particularly strong at identifying Asian ethnicities, especially

[36] Five of these 27 papers have only one U.S. author. In some instances the PI does not have a lab. We include these instances in this count.

[37] Authorship patterns vary by discipline. In the life sciences the last author is generally the PI and the one who supplied the resources. The first author is the one who contributed the greatest amount to the research. This pattern is also true in chemistry and can also be the pattern in physics. In some disciplines, such as the earth and environmental sciences, authorship order is arranged entirely in terms of contribution. Authors are rarely listed in alphabetical order on scientific papers. For example, only 26 of the 159 papers we identified listed authors alphabetically; 19 of these papers had only two authors, implying that there was a 5 percent chance of their being alphabetical regardless of practice.

[38] See http://www.physics.umn.edu/outreach/reu/REU2006Proceed.pdf for papers by the REU interns.

[39] We are grateful to Bill Kerr not only for providing us access to the database but also for doing the actual match.

Chinese, Indian/Hindi, Japanese, Korean, Russian, and Vietnamese names. In addition to the Asian ethnicities, we are able to distinguish four other ethnicities: Russian, English, European and Hispanic.[40] The approach exploits the idea that authors with "the surnames Chang or Wang are likely of Chinese ethnicity, those with surnames Rodriguez or Martinez of Hispanic ethnicity and so on" (Kerr 2007).

The methodology uses both first and last names and thus minimizes ambiguity in assigning names with multiple ethnicities, such as Lee and Park. Using ethnic names to identify citizenship status of graduate students and postdocs clearly has some limitations. If Asian and Hispanic names are classified as being foreign, the technique will overcount the foreign representation, given the number of U.S. citizens with Asian and Hispanic names. On the other hand, if English and European names are used to classify individuals as "native," the native count will be overstated, given the number of European, English and Canadian students and postdocs working in the United States.

Some indication of the degree of bias is given by examining the ethnicity of PhD recipients in the United States and the country of origin of PhD recipients who are non-citizen (either permanent or temporary resident). For example, in 2006, 1,164 PhDs in S&E were awarded to U.S. citizens who self-identify as being "Asian" (Jaquelina Falkenheim, 2007, Table 2). Concurrently, 7,918 PhDs were awarded to non-U.S. citizens (permanent and temporary visas) from the Asian countries of China, India, Korea, Japan, and Thailand (Falenkeim, 2007, Table 4). Assuming that citizens who self-identify as "Asian" have Asian last names leads to the conclusion that 13% of all PhD degrees awarded in the U.S. to individuals with Asian names went to citizen graduate students; 87% went to foreign graduate students. We cannot make a similar calculation for postdocs, given that neither the ethnicity of postdocs nor the source country of postdocs is ascertained. But we have reason to believe that the 87% is an undercount, given that not only among U.S. Ph.D. recipients is the postdoc-taking rate for non-citizen Asians high (Stephan and Ma 2005) but, in addition, a large percent of postdocs receive their PhDs outside the U.S. Many of these, we assume, are Asian.

We estimate that approximately 1,132 PhDs in S&E were awarded to non-U.S. citizens from English and European countries in 2005.[41] Using "white" as synonymous with "English" and "European" and noting that the number of S&E degrees awarded to "white" citizens in 2005 was 12,514 (Hoffer et al 2006, Table 8), we "guestimate" that 8% of the English and European PhD names belong to non-citizens. In a similar way we "guestimate" that 40% of Hispanics receiving

---

[40] In some instances, the matching procedure attributes a name to several ethnicities, providing the probability of ethnicity associated with each match. In these instances we coded the ethnicity that had a greater than 50 percent probability. By way of contrast, Kerr (2008), who has a significantly larger database and addresses different questions, summed probabilities associated with an ethnicity rather than assuming a specific ethnicity in cases that he refers to as "ties."

[41] NSF provides data on the top 30 countries of origin of non-U.S. citizens earning doctorates regardless of field (Hoffer et al 2006, Table 12). We classify three of these countries as English: Australia, Great Britain, and Canada. The total number of PhD recipients from these countries is 800. We classify three as "European:" Germany, Italy, and France; the number of recipients from the three is 581. We estimate that 82% of all doctorate degrees awarded to non-citizens in the U.S. are in S&E (Hoffer et al 2006, Table 11). From this, we estimate that 1,132 PhDs were awarded in S&E to individuals who have European or English names and are non-U.S. citizens.

degrees are non-citizens.[42] In light of our counts, taken together these "biases"come close to cancelling each other out and we believe that we have fairly reasonable overall counts for non-citizen PhD students by "keying" on ethnicity of name if we classify English and European as "native" and all others as foreign. We believe this undercounts the total number of non-citizens among postdoctorates, given the large number of individuals who come with PhD in hand to take a postdoc postion as well as the large number of non-citizen Ph.D. recipients who stay in the U.S. for postdoctoral training.

It is more difficult to ascertain the magnitude of the bias for positions such as faculty and staff scientist. For our purposes, however, we will use the same convention as that noted above.

The ethnicity of U.S. authors on papers with less than 10 authors is presented in Table 5 by position. We identified no Vietnamese authors and hence this category is not included in the table. "Other" refers to ethnicities not contained in the Melissa data.[43]

We find that 57.2 % of authors with a U.S. address (and writing with a last author at a U.S. institution of higher education) are identified as having English names and 6.4% have European names. We find that 4.3% have Hispanic names, 16.6% have Chinese names and 4.3% have Indian/Hindi names. Koreans, Japanese, Russians, and "other" make up the remaining 11.4%.

Of particular interest to our study is that 71 of the 120 postdoc authors are neither English nor European (59.1%). This is remarkably close to the 60 percent that NSF estimates for 2006.[44] We find that 42 of the 108 graduate student co-authors have neither English nor European names (39.6%). This is slightly lower than the percent of U.S. PhDs awarded in science and engineering to non-citizen PhDs in 2006 (Falkenheim, 2007, Table 2) but consistent with the finding of John Bound and Sarah Turner (this volume) that higher-ranked institutions (from which most of these authors are drawn) have a lower proportion of foreign-graduate students than do lower-ranked institutions. We note that a large percent of the faculty authors are English or European (79.2%); the next most likely ethnic group to be a faculty author is Chinese (8.8%). We also classify authors according to whether they are a staff scientist or a technician. We find that fully 60% of authors in such positions have English or European names; 13.6% have Chinese names.

Focusing on articles, we find that 70 of the 133 papers (53%) with fewer than 10 U.S. authors have a foreign student or postdoc as a coauthor. This represents approximately 60% of the 115 papers that have either a student or a postdoc author. We infer that it is the norm, not the exception, to have an international student or postdoc as a coauthor in papers published in *Science*.

---

[42] We classify four countries in the "top 30 countries" list as "Hispanic." (Hoffer et al Table 12 2006). These are Mexico, Colombia, Argentina and Spain. Collectively, 618 PhDs were awarded to individuals from these countries. We estimate that 82% of these are awarded in S&E (507), using data from Table 11, Hoffer et al 2006. 744 degrees were awarded in S&E to citizens who self-identify as Hispanic (Hoffer et al Table 8). From these two figures we "guestimate" that 41% of the degrees awarded to Hispanics are to non-citizens.

[43] The database used for the ethnicity match contained several edits that were not present in the database used in creating Tables 1-4. Thus, while the counts in the ethnicity tables are very close to those in the earlier tables, they do not always correspond perfectly.

[44] Note that NSF calculations classify "permanent residents" with U.S. citizens in determining citizenship status of postdocs.

Table 6 shows position and ethnicity for U.S. first authors from our sample of all papers. We find that 55% are either of English or European ethnicity, the remaining 45% are "foreign"--17.8% are Chinese, 7.8% are Indian/Hindi, 4.7% are Hispanic, and 14.3% are drawn from other ethnicities. The heavy representation of graduate students and postdocs in the first-author position has already been noted. But what we learn from this table is the important role of "foreign" graduate students and postdocs. To wit, using our convention, we find that almost 59% of the graduate student first authors are foreign—a figure significantly higher than the percent of non-citizen PhD recipients in science and engineering and higher than the percent of "foreign" graduate students among graduate student coauthors in general (Table 6). Non-citizens also make up slightly more than 54% of the first-author postdocs. Clearly international graduate students and postdocs are important not only in staffing labs; they play lead roles in research. It is also interesting to note that faculty play a relatively minor role as first author, while staff scientists and technicians play a relatively important role (other category.)

The position and ethnicity for last authors is given in Table 7. It is of less interest to our study, given the small role that graduate students and postdocs play as "last authors." Briefly, and using the same convention, we note that 78% of last authors are "native"; 22% are foreign. Almost 50% of the "foreign" last authors are Chinese.

Our findings regarding nativity are summarized in Table 8. Slightly more than 44% of first authors are foreign; almost sixty percent of postdoc authors are foreign. Last authors are very likely to be native (over 73%) and six out of ten graduate student authors are native.

Finally, in Table 9, we examine "affinity effects" by comparing the ethnicity of the last author to the ethnicity of coauthors working in the U.S. for all papers with less than 10 authors. Proceeding in such a manner, we find that 73.8% of the coauthors of English last authors are English. If non-last authors were distributed randomly across articles, we would expect it to be 54.5%, based on the distribution in our database of authors. In a similar manner, we find that 53.8% of the coauthors of Chinese last authors are Chinese—a figure that is strikingly higher than the 18.6% that we would expect. Affinity effects also appear to be present for Hispanics but the cell sizes are very small. We find no evidence of affinity effects for European last-authors.

Section VI. Conclusion

Summary of Findings

Universities play an important role in the production of knowledge in the United States, authoring nearly 75% of scientific and engineering articles written within the country. Within the university, research is often performed with the assistance of graduate students, postdoctoral scholars and staff scientists, many of whom are foreign-born and, in the case of graduate students and postdocs, are studying in the U.S. on temporary visas.

Here we document the important role played by students and postdocs in university research by analyzing authorship patterns for a six-month period for articles published in *Science* having a last author affiliated with a U.S. university. We choose *Science* because of its multidisciplinary nature and its position as a leading, if not the leading, journal in science. The fast turn-around time (decisions are generally made in less than a month and publication rapidly follows) also means that we are able to do web research regarding the status of authors.

We analyze authorship patterns for two sets of papers: (1) papers having fewer than ten authors, in which case we determine the status of all authors residing in the U.S. and (2) all papers regardless of the number of authors, in which case we determine the status of the first and the last author. The first dataset contains 133 articles; the second data set contains 159 papers. We determine the status of each author with a U.S. affiliation through web-based research, starting with the last author's web page, which often contains a link to the lab and the group working in the lab. We find the web to be a powerful tool: of the 585 U.S. authors we can determine the status of 550. We believe we are the first to use such a methodology to investigate the role that students and postdocs play in research.[45]

Our analysis demonstrates the important role that students and postdocs play in university research. We find that 45.6% of all authors, or almost one out of two, is a postdoc, student or a recent alum of the program. By category, 22.4% are postdocs, 19.6% are graduate students, 1.5% are undergraduate students, another 1.5% are student or postdoc, status not known, and a handful are alums of the program. What is even more indicative of the important role that students and postdocs play in university research is our finding that 86.5% of papers—nearly seven out of eight (133-paper sample) -- have either a current postdoc or student as one of the authors.

The role of postdocs and students is especially striking when one looks at first-author position on all U.S. papers, regardless of the number of authors. To wit, we find that 101 of the 137 first authors who are in the U.S. and whose position is known are either a postdoc or a student (74.6%);, 7 of the last authors are either a postdoc or a student.

We identify the ethnicity of authors, drawing on the ethnic-name database that Kerr (2008) used to identify ethnicity of U.S. inventors. The methodology is particularly strong at identifying Asian ethnicities. This approach clearly has some limitations. If Asian and Hispanic names are classified as being foreign, the technique overcounts the foreign representation, given the number of U.S. citizens with Asian and Hispanic names. On the other hand, if English and European names are used to classify individuals as "native," the native count will be overstated, given the number of European, English and Canadian students and postdocs working in the United States. We draw upon the distribution of PhDs awarded in 2006 to investigate the degree of this bias. We conclude that approximately 87% of the Asians we identify are non-citizens; 8% of the English and Europeans we identify are non-citizens; and 40% of the Hispanics are non-citizens. In light of our counts, these "biases" approximately cancel each other out and we believe that we get fairly reasonable overall counts for non-citizen PhD students and postdocs by "keying" on ethnicity of name and defining "English" and "European" as native.

Using this approach, we find that 59.2% of postdoc authors are neither English nor European, a figure that is remarkably close to the 60 percent that NSF estimates. We find that 39.6% of the graduate-student coauthors have neither English nor European names. This is slightly lower than the percent of PhDs awarded in science and engineering to non-citizens in 2006. At the paper level, we find that 70 of the 133 papers (53%) with fewer than 10 U.S. authors have a foreign

[45] G. Vogel (1999) examines authorship patterns for two issues of *Science* In 1999.

student or postdoc as a coauthor. This represents approximately 60% of the 115 papers that have either a student or a postdoc author. Clearly, it is the norm, not the exception, to have an international student or postdoc as a coauthor in papers published in *Science.*

Using the same convention we find that almost 60% of the graduate-student first authors are foreign and that non-citizens make up slightly more than 59% of the postdocs who are first authors. We conclude that international graduate students and postdocs are important not only in staffing university labs; they play lead roles in university research.

Discussion

It has long been known that the foreign-born play an important role in U.S. science and engineering. The basis for much of this understanding has been the role the foreign-born play as faculty or when working in industry. The results of the present study suggest that the foreign-born play an important role in doing research, much of which is of a basic nature, while they are graduate students and postdocs. The finding is not surprising but prior to this study no one has set about to investigate the degree to which the foreign-born contribute in this way.

The contributions of the foreign-born graduate students and postdoctoral scholars to U.S. science, of course, do not end when their training is completed. Many choose to stay in the United States. Finn, for example, finds that approximately 70 percent of PhD recipients on temporary visas in science and engineering were in the U.S. two years after receiving their PhD degree; the five-year stay rate was only slightly lower (Michael Finn 2005, Table 3). The rate is highest for Chinese, who have a five-year stay rate of 90%, followed by Indians with a five-year stay rate of 86%. (Finn 2005, Table7.) No one has made comparable estimates for postdocs, but the assumption is that a number who come to train stay on after their training is completed. The ethnicity of faculty authors in this study is suggestive of this; approximately one in five had neither English nor European names. The group making up the highest percent of non-native faculty was of Chinese ethnicity.

This is not to say that scientists and engineers contribute to U.S. science only when they stay. Many who return end up coauthoring papers with colleagues in the U.S. We see some examples of this in our data. The work of Adams et al finds that the international co-authorship patterns of faculty at U.S. universities are influenced by the number of foreign students trained in their department who return to their home country (2005). Moreover, co-authorship is not the only way by which scientists in one country benefit from the work and expertise of others. Published science is a public good; regardless of whether they stay or leave, these researchers will continue to contribute to the creation of knowledge.

That foreign-born graduate students and postdoctoral fellows play an important role appears indisputable from this research. But it does not follow that their places would be left empty if they were not to come. Considerable debate has focused on the degree to which foreign-born students displace U.S. students. The question is difficult to answer but there is reasonable agreement regarding several facts. First, natives, especially native males, when choosing a career are responsive to alternative opportunities. In the last twenty or so years many of these opportunities—for example the law and business—have proved relatively more attractive, requiring shorter training times and offering higher salaries. Second, if the incentive structure were to change, the number of U.S. citizens entering S&E would arguably change as well. By way of example, Richard Freeman (2005) finds the size of the applicant pool for NSF Graduate Research Fellowships to be responsive to the relative value of the stipend and concludes "that the supply of highly skilled applicants is sufficiently responsive to the value of awards that increases in the value of stipends could attract some potentially outstanding science and engineering students who would otherwise choose other careers." Third, and by way of contrast, foreign-born have had fewer alternatives available that offer the option of support while in school and employment at a favorable relative wage. Fourth, the alternatives open to the foreign-born are changing. Programs outside the U.S. are becoming more and more competitive. Since the late 1980s the number of S&E PhD degrees awarded in Europe has surpassed the number in the U.S. In the late 1990s, the number of degrees awarded in Asian countries surpassed the number awarded in the U.S. In China alone the number accelerated from virtually zero in 1985 to approximately 13,500 by 2004 (National Science Board 2008, Appendix Tables 2-42 and 2-43). At the same time, programs in the U.S. are at risk of becoming less attractive to foreign-born students and postdoctoral scholars. This is not only because funds for graduate and postdoctoral support are diminishing as agencies such as NIH experience real decrease in funding levels, but also because of problems faced by foreign nationals in the U.S. since 9/11. A case in point is the special vetting required for foreign nationals to work on research supported by federal agencies and considered "sensitive but unclassified."[46]

Nor does it follow that the demand for graduate students and postdocs to work at universities will necessarily persist at its current level. The technology of discovery is changing. By way of example, in 1990 the best-equipped lab could sequence 1000 base pairs a day. By January 2000 the 20 labs involved in mapping the human genome were collectively sequencing 1000 base pairs a second, twenty-four/ seven. The cost per finished base pair fell from $10.00 in 1990 to under $.05 in 2003 (Francis S. Collins, Michael Morgan, Aristides Patrinos 2003) and was roughly $.01 in 2007 (www.biodesign.asu./edu/news/232/). As the technology of discovery changes, the need for skilled lab workers—many of whom are graduate students and postdocs--may decline. Moreover, as equipment becomes increasingly sophisticated and more expensive, research procedures may increasingly be outsourced to non-university facilities. Mail-in crystallography, where crystals are sent to large non-university labs for analysis, is but one example. There is also the question of whether the Federal government will continue to provide resources for graduate research assistants and postdocs at the level it has in the past. The financial crisis of 2008 and the many demands it has placed on the Federal government may lead us to look back on the recent past, despite its many funding issues, as a time when university research flourished.

The heavy reliance on graduate students and post doctoral scholars in the performance of university research has contributed to the U.S. eminence as a training center for both native and foreign-born students. It provides not only hands-on learning but also financial support for graduate study and postdoctoral work, something that many other countries cannot provide. Factors that reduce either the demand for or supply of graduate students and postdocs have the potential of threatening the U.S.'s eminence as a training center and producer of research.

[46] This may change in the near future. In June of 2008 DOD Under Secretary John Young wrote a directive stating that "classification is the only appropriate mechanism" for restricting participation by foreign nationals or for restricting publication (Bhattacharjee 2008, p. 325).

References

Adams, James. D., Grant C. Black, Roger Clemons, and Paula Stephan. 2005. Scientific team and institutional collaborations: evidence from U.S. universities, 1981-1999. *Research Policy 34*, 259-285.

Andrews, Frank M., ed. 1979. *Scientific Productivity: The Effectiveness of Research Groups in Six Countries.* Cambridge: Cambridge U. Press.

Agrawal, Ajay & Goldfarb, A. 2008. Restructuring Research: Communication Costs and the Democratization of University Innovation. *American Economic Review* (forthcoming).

Barnett, Andy, Richard W. Ault and David Kaserman. 1988. The Rising Incidence of Co-authorship in Economics: Further Evidence. *Review of Economics and Statistics* 70:539-543.

Bhattacharjee, Yudhijit. 2008. New Policy Tries to Ease Security Restrictions. *Science*, 321:325.

Bunton, Sarah A. and William T. Mallon. 2007. The Continued Evolution of Faculty Appointment and Tenure Policies at U.S. Medical Schools. *Academic Medicine*, 82:3:281-289.

Collins, Francis, Michael Morgan and Aristides Patrinos. 2003. The Human Genome Project. *Lessons from Large-Scale Biology, Science* 300 (5617):286-290.

Davis, Geoff. 2005. Doctors without Orders. *American Scientist* 93 (3):supplement.

Ehrenberg, Ronald G., Michael J. Rizzo and G. H. Jakubson. 2003. Who Bears the Growing Cost of Science at Universities? National Bureau of Economic Research Working Paper 9627.

Falkenheim, Jaquelina. 2007. "U.S. Doctoral Awards in Science and Engineering Continue Upward Trend in 2006. InfoBrief, Science Resources Statistics, NSF.

Finn, Michael G. 2005. "Stay Rates of Foreign Doctorate Recipients from U.S. Universities, 2003." Oak Ridge Institute for Science and Education.

Fox, Mary Frank. 1991. Gender, Environmental Milieu, and Productivity in Science. In *The Outer Circle: Women in the Scientific Community*, edited by H. Zuckerman, J. R. Cole and J. T. Bruer. New York: W. W. Norton.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. 2008. Changing Incentives to Publish and the Consequences for Submission Patterns. Paper prepared for presentation at the DIME-BRICK Workshop "The Economics and Policy of Academic Research," July 2008 Collegio Carlo Alberto, Moncalieri (Torino), Italy.

Freeman, Richard B. 2005. "Supporting 'The Best and the Brightest' in Science and Engineering: NSF Graduate Research Fellowships," NBER Working Paper no. 11623.

Hoffer, Thomas, Karen Grigorian, and Eric Hedbert. 2008. "Postdoc Participation of Science, Engineering, and Health Doctorate Recipients," InfoBrief, National Science Foundation, March.

Hoffer, T.B., V. Welch, Jr., K. Webber, K. Williams, B. Lisek, M Hess, D. Loew, and I. Guzman-Barron. 2006. Doctorate Recipients from United States Universities: Summary Report 2005. Chicago: National Opinion Research Center.

Jones, Benjamin. 2005. The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? NBER Working Paper 11360.

Kean, Sam. 2006. Scientists Spend Nearly Half Their Time on Administrative Tasks, Survey Finds. *Chronicle of Higher Education*, A23.

Kerr, William R. 2007. The Ethnic Composition of U.S. Inventors: Evidence Building from Ethnic Names in U.S. Patents. Paper prepared and presented at NBER workshop on Career Patterns of Foreign Born Scientists and Engineers Trained and or Working in the U.S., November 7, 2007.

Kerr, William R. 2008. Ethnic Scientific Communities and International Technology Diffusion, *Review of Economics and Statistics*, 90(3):518-530.

Kolbert, Elizabeth. 2007. Crash Course: The World's Largest Particle Accelerator. *The New Yorker*, May 14, 68-78.

Lawani, S. M. 1986. Some Bibliometric Correlates of Quality in Scientific Research. *Scientometrics* 9:13 25.

Malakoff, David. 2000. The Rise of the Mouse, Biomedicine's Model Mammal. *Science* 288:248-253.

Mervis, Jeffrey. 2008. Top Ph.D. Feeder Schools Are Now Chinese. *Science*, 321:185.

National Academies. 2005. *Policy Implications of International Graduate Students Postdoctoral Scholars in the United States.* Committee on Policy Implications of International Graduate Students and Postdoctoral Scholars in the United States, Committee on Science, Engineering and Public Policy. Washington, DC: The National Academies Press.

National Science Board. 2008. Science and Engineering Indicators, Arlington, VA.

National Science Board. 2006. Science and Engineering Indicators, Arlington, VA.

National Science Foundation. WebCASPAR <http:/webcaspar.nsf.gov/>.

Overbye, Dennis. 2007. A Giant Takes on Physics' Biggest Questions. *New York Times* May 15, D1, D4-D5.

Pelekanos, Adelle. 2008. Money Management for Scientists: Lab Budgets and Funding Issues for Young Pis. Science Alliance, New York Academy of Sciences, June 16th.

Penning, Trevor. 1998. The Postdoctoral Experience: An Associate Dean's Perspective." *The Scientist* September 12.

Regets, Mark. 2007. "An Overview of Foreign-born S&E Workers in the United States and their Measurement," unpublished manuscript prepared for presentation at the NBER Conference on Career Patterns of Foreign-born Scientists and Engineers, November 7, 2007.

Regets, Mark. 2005. "Foreign Students in the United States," presentation at the Dialogue Meeting on Migration Governance: European and North American Perspectives, Brussels, Belgium, June 27, 2005.

Science, 1998. The Biocomplex World of Rita Colwell. 281:1944-1947.

Stephan, Paula E. 2008. Job Market Effects on Scientific Productivity. In Scientific Competition, pp. 11-29, edited by Max Albert, Dieter Schmidtchen, and Stefan Voigt, Conferences on New Political Economy 25, Mohr Siebeck: Tubingen, Germany.

Stephan, Paula E., Grant Black and Tanwin Chaing. 2007. The Small Size of the Small-Scale Market: The Early-Stage Labor Market for Highly Skilled Nanotechnology Workers. *Research Policy*, 36(6):887-892.

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

Stephan, Paula E. and Sharon G. Levin. 2002. "The Importance of Implicit Contracts in Collaborative Research," (Paula Stephan and Sharon Levin) in *The New Economics of Science*, pp. 412-430. Edited by Philip Mirowski and Esther-Mirjam Sent, University of Chicago Press.

Tanyildiz, Zeynep Esra. 2008. The Effects of Networks on Institution Selection by Foreign Doctoral Students in the U.S. Unpublished dissertation, Georgia State University.

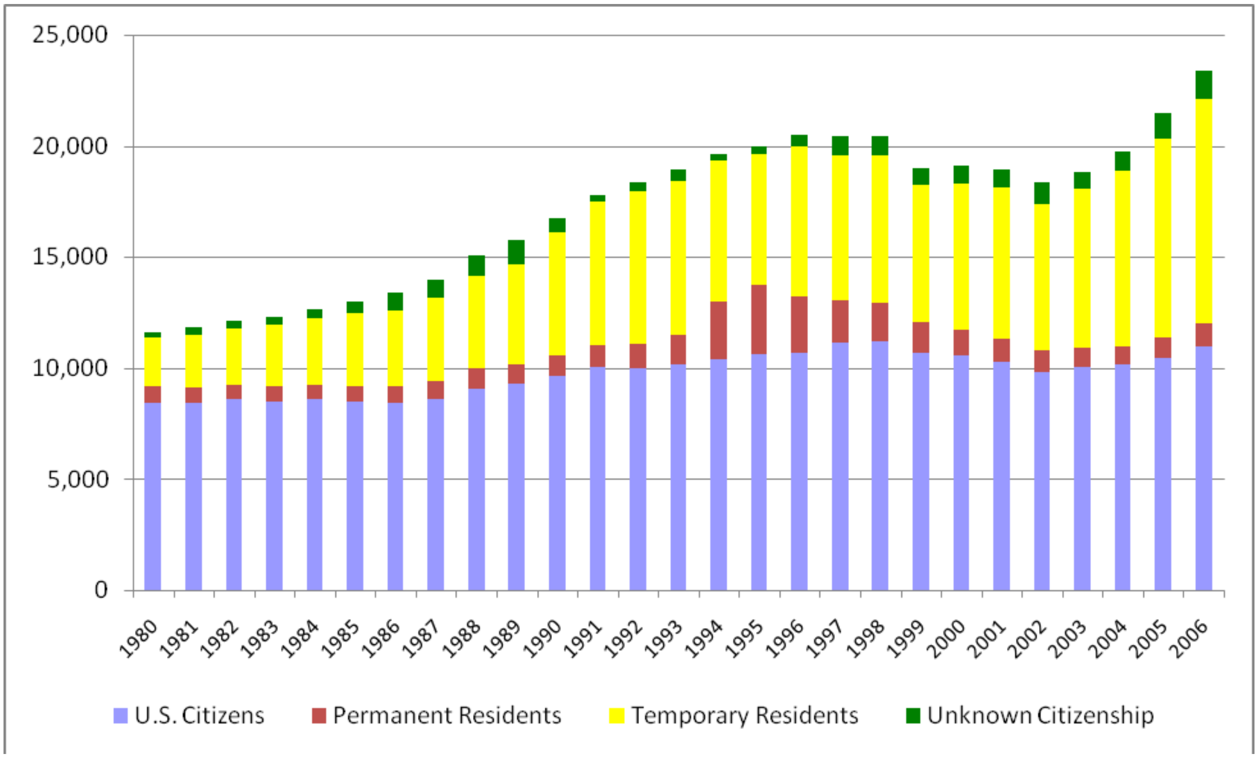
Vogel, G. 1999. Working Conditions: A Day in the Life of a Topflight Lab." Science, 285,1531-2.

Winkler, Anne, Sharon Levin and Paula Stephan. 2008. The Diffusion of IT in Higher Education: Publishing Productivity of Academic Life Scientists.

Wuchty, Stefan, Benjamin Jones and Brian Uzzi. 2007. The Increasing Dominance of Teams in the Production of Knowledge. *Science* 316:1036-1030.

Figure 1

S&E PhDs Awarded by Citizenship Status, 1980-2006

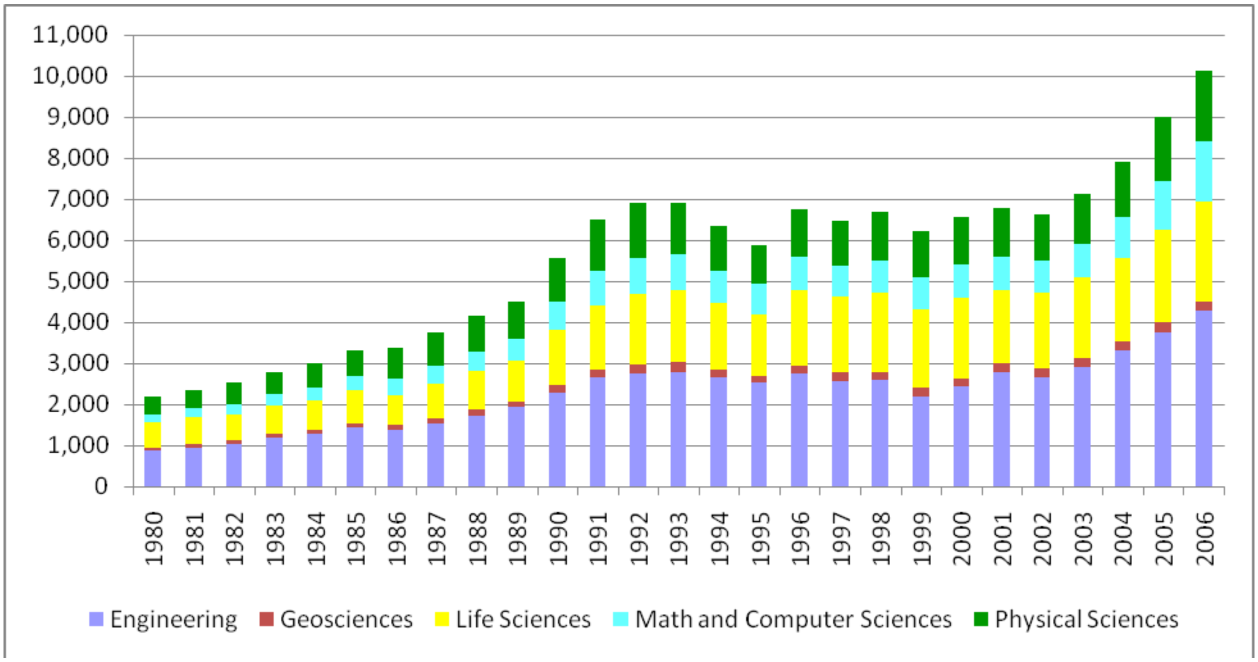


Source: National Science Foundation, WebCASPAR database[47]

[47] Data for Figures 1-4 come from WebCASPAR. WebCASPAR is an online integrated database of data from U.S. academic institutions emphasizing science and engineering. WebCASPAR includes data sources from the National Science Foundation and the National Center for Education Statistics. The National Science Foundation oversees the WebCASPAR database. WebCASPAR data used in this study originally come from NSF's Survey of Earned Doctorates and Survey of Graduate students and Postdoctorates in Science and Engineering (also known as the Graduate Student Survey, or GSS). Data used in Figures 1-4 were selected from WebCASPAR based on status as a Ph.D. recipient, graduate student or postdoc; citizenship status; S&E field; and year.

Figure 2

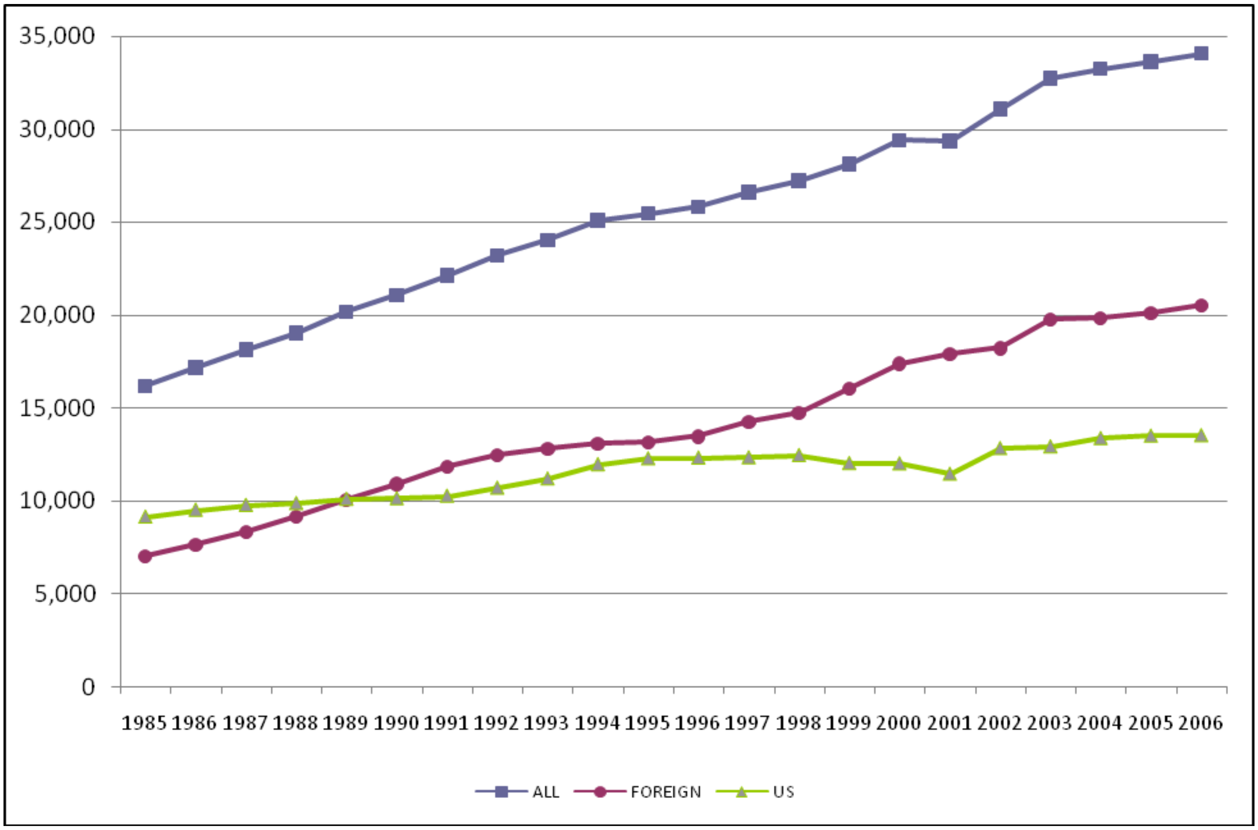
Number of S&E PhDs Awarded to Temporary Residents by Field, 1980-2006



Source: National Science Foundation, WebCASPAR database

Figure 3

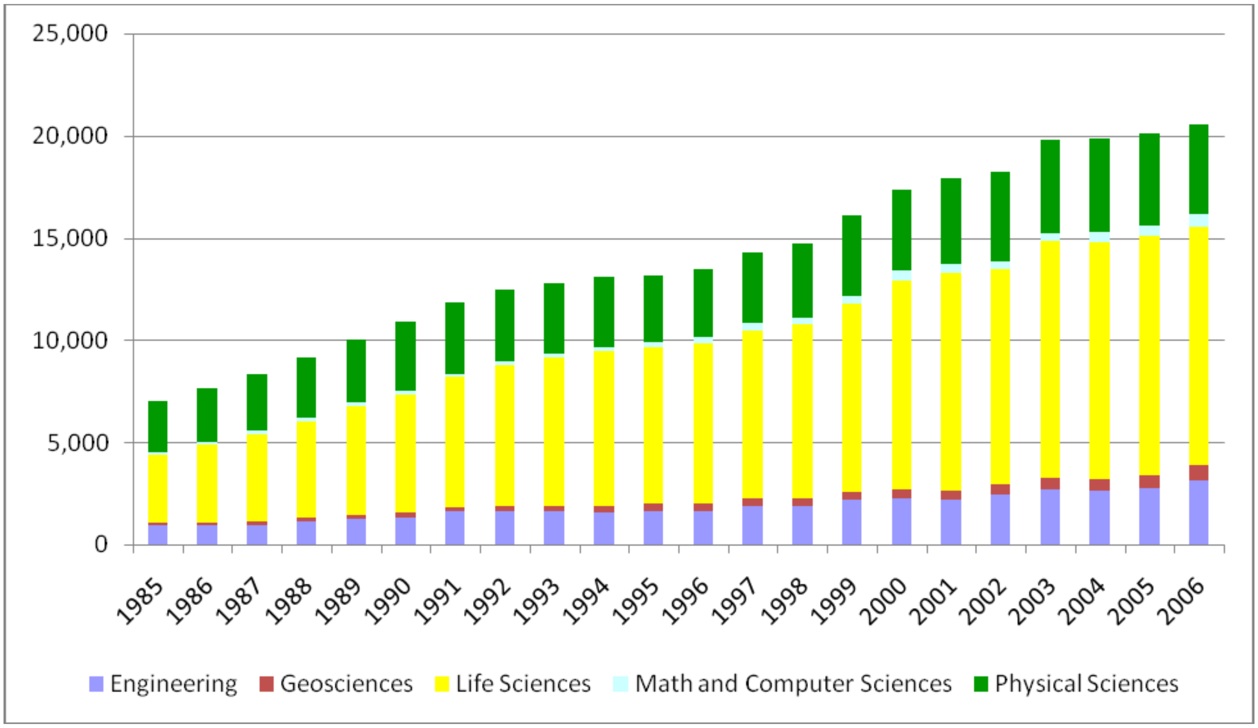
Number of S&E Postdocs Working in Academe, 1985-2006



Source: National Science Foundation, WebCASPAR database

Figure 4

Number of Foreign S&E Postdocs by Field, 1985-2006



Source: National Science Foundation, WebCASPAR database

Table 1: Distribution of *Science* Papers by Last Author Affiliation

Number of issues coded	Number of articles in issues	Number for whom last author has a non-academic affiliation	Number for whom last author has an academic affiliation	Number for whom last author has a U.S. academic affiliation
22	318	51	267	159

Source: Authors

Table 2: *Science* Articles by Field

Area	Number of articles	Median number of authors	Minimum number of authors	Maximum number of authors
Biochemistry	21	5	3	15
Biology	34	6	1	71
Chemistry and related	9	4	2	9
Earth sciences	16	5	1	22
Genetics	16	7	3	42
Material science	8	5	3	10
Nano-related	6	5.5	4	15
Neurology	12	4	3	14
Physics	17	5	2	14
Other	20	5	2	11

Source: Authors

Table 3: Descriptive Data for Articles with Less than 10 Authors (133)

Total number of authors	648
Total number of authors in U.S.	585
Total number of U.S. authors for whom position is known	550
Total number postdocs	123
Total number of graduate students	108
Total number of undergraduate students	8
Student (grad or undergrad) or postdoc; status/ unknown	8
Was affiliated with lab	4
Number of papers with one or more author who is a postdoc, grad student or undergraduate student	115

Source: Authors

Table 4: First and Last Authorship Patterns

	All U.S. articles	First author (restricted to counts for articles having more than one author)	Last author
Number of U.S. papers	159	157	159
Number of authors in U.S.	303	142	159
Total number of U.S. authors for whom position is known	296	137	157
Total number of postdoc authors	59	57	2
Total number of graduate student authors	45	41	4
Total number of undergraduate student authors	1	0	1
Student/postdoc author but exact status unknown	3	3	0

Source: Authors

Table 5: Ethnicity of Authors of Papers with Less than 10 Authors

Position	European	English	Chinese	Indian	Japanese	Hispanic	Russian	Korean	Other	Row Total (Percent of Total)
					(Percent of Row Total)					
Postdoc	7	42	35	6	7	10	4	3	6	120
	(5.8%)	(35.0%)	(29.2%)	(5.0%)	(5.8%)	(8.3%)	(3.3%)	(2.5%)	(5.0%)	(20.8%)
Graduate student	4	60	20	5	1	2	3	3	8	106
	(3.8%)	(5.7%)	(18.9%)	(4.7%)	(0.9%)	(1.9%)	(2.8%)	(2.8%)	(7.5%)	(18.3%)
Undergraduate student	0	5	1	1	0	0	0	0	1	8
	(0%)	(62.5%)	(12.5%)	(12.5%)	(0%)	(0%)	(0%)	(0%)	(12.5%)	(1.4%)
Student or postdoc, status not identified	0	4	1	0	0	1	0	0	0	6
	(0%)	(66.7 %)	(16.7%)	(0%)	(0%)	(16.7%)	(0%)	(0%)	(0%)	(1.0%)
Faculty	19	152	19	8	2	7	1	2	6	216
	(8.8%)	(70.4%)	(8.8%)	(3.7%)	(0.9%)	(3.2%)	(0.5%)	(0.9%)	(2.8%)	(37.4%)
Staff scientist/techni-cian	5	35	9	0	3	4	5	1	4	66
	(7.6%)	(53.0%)	(13.6%)	(0%)	(4.5%)	(6.1%)	(7.6%)	(1.5%)	(6.1%)	(11.4%)
Other	0	10	5	1	0	0	0	0	0	16
	(0%)	(62.5%)	(31.3%)	(6.3%)	(0%)	(0%)	(0%)	(0%)	(0%)	(2.8%)
Not known	2	23	6	3	1	1	2	0	2	40
	(5.0%)	(57.5%)	(15.0%)	(7.5%)	(2.5%)	(2.5%)	(5.0%)	(0%)	(5.0%)	(6.9%)
Total	37	331	96	24	14	25	15	9	27	578
	(6.4%)	(57.3%)	(16.6%)	(4.2%)	(2.4%)	(4.3%)	(2.6%)	(1.6%)	(4.7%)	(100%)

Source: Authors

Table 6: Position and Ethnicity for U.S. First Authors

Position	European	English	Chinese	Indian	Japanese	Hispanic	Russian	Korean	Other	Total
Postdoc	3	23	13	5	0	5	3	0	4	57*
Graduate student	2	15	8	6	1	1	1	2	5	41
Undergraduate student	0	0	0	0	0	0	0	0	0	0
Student or postdoc, status not identified	0	4	0	0	0	0	0	0	0	4
Faculty	2	12	0	0	0	0	0	0	0	14
Other (including not known)	1	16	4	0	2	0	2	0	0	25
Total	8	70	25	11	3	6	6	2	9	140/141*

Source: Authors

*Postdoc total includes one individual whose ethnicity is not classified.

Table 7: Position and Ethnicity for U.S. Last Authors

Position	European	English	Chinese	Indian	Japanese	Hispanic	Russian	Korean	Other	Total
Postdoc	0	2	0	0	0	0	0	0	0	2
Graduate student	0	4	0	0	0	0	0	0	0	4
Undergraduate student	0	1	0	0	0	0	0	0	0	1
Faculty	7	92	13	9	2	9	2	2	3	139
Other (including not known)	1	10	1	0	0	0	0	0	1	13
Total	8	109	14	9	2	9	2	2	4	159

Source: Authors

Table 8: Authorship Patterns by Nativity (Percent)

Position	Native	Foreign
First authors	55.7	44.3
Last authors	73.6	26.4
Postdoc authors	40.8	59.2
Graduate students	60.4	39.6

Source: Authors

Table 9: Affinity Effects in Authorship Patterns

Ethnicity of last author	Expected percent of coauthors with same ethnicity	Actual percent of coauthors with same ethnicity	Number of papers
English	54.5	73.8	88
Chinese	18.6	53.8	13
Indian	3.4	5.5	9
European	6.7	0.0	7
Hispanic	4.3	23.3	6

Source: Authors