

Supplementary Materials for
Labor advantages drive the greater productivity of faculty at elite universities

Sam Zhang *et al.*

Corresponding author: Sam Zhang, sam.zhang@colorado.edu; Aaron Clauset, aaron.clauset@colorado.edu

Sci. Adv. **8**, eabq7056 (2022)
DOI: 10.1126/sciadv.abq7056

This PDF file includes:

Data Preprocessing
Tables S1 to S14
Figs. S1 to S11
References

Supplementary Materials

DATA PREPROCESSING

Linking employment records with Web of Science

For each person in our employment dataset, we first queried their publications in Web of Science (WoS) through the WoS API. We intentionally created a permissive query that would accept a high false positive rate in exchange for a low false negative rate, since our filtering algorithms could subsequently reduce the false positives but not false negatives. Furthermore, department names tended to be highly idiosyncratic, and difficult to query exactly. Thus the query consisted of the last name of the faculty member, the first three letters of the first name followed by a wildcard for the rest of the first name, as well as a standardized name for their institution in the address field that replaced certain conjunctions and punctuation with boolean keywords.

Web of Science returns data on the institution (“org”) and departmental unit (“suborg”) along with a street address for the address field of authors on papers, and authors are grouped into blocks that share the same address. Multiple spellings of each org are provided by WoS, with one marked as the preferred entry. We chose the preferred entry, although it sometimes contained less precise information than some of the alternatives, for example omitting the individual campus for a state university system. To take advantage of the standardization of the preferred org but to avoid losing geographic data on campus, we created our own organization identifier that concatenated the org with the city. We found this to be specific enough to match each institution of our employment records directly, and we performed that institutional linkage through a combination of heuristics and manual checking.

We merged the queried publications back to the employment records using the first three letters of the first name, last name, and the mapping that we developed above between the institutions of the two datasets. We also expanded our result set to include anyone else in the results whose WoS distinct author ID also appeared in the matched set, with the same affiliation as the employment record. Unlike the preferred org names, suborgs displayed a long tail of unique values, with the same department spelled several different ways. We kept the 10,000 most common suborgs, which included both highly commonly reused names such as “Dept Biol”, but also enough of the long tail. We include all document types indexed by Web of Science aside from letters, corrections, and retractions. Thus we include journal articles (82.7%), proceedings papers (11.5%), reviews of all kinds (3.9%), editorial material (1.8%) and biographical and news items (less than 0.1%). The remaining document types in Web of Science are associated with fewer than 0.01% of the

total publications, and are primarily associated with the humanities, such as poetry, creative fiction, and reviews of films, records, and music.

Next, we computed the most common Web of Science subject associated with papers published by people with affiliations with that suborg name, as long as over 30% of papers with that suborg label had that subject label. Each Web of Science author record was associated with a distinct author ID (“daisng ID”). We filtered to daisng IDs who published at least one paper with a matching subject label to their suborg. This resulted in 1,829,326 papers, each with at least one of 134,776 unique tenure-track faculty authors linked to our employment dataset. We further have 654,866 WoS distinct author IDs for non-tenure track collaborators on these papers. Viewing this as a bipartite network where half of the nodes are papers and the other half are authors, with an edge between a paper and an author if that author coauthored the paper, we have 4,932,471 total edges. We use the employment records, which contain the year that each person received their degree, to filter the data to publications after the year of each person’s degree. This left 1,789,333 publications and 133,747 unique tenure-track faculty before joining with the NSF Survey of Graduate Students and Postdoctorates in Science and Engineering.

Web of Science has greater coverage within some disciplines compared with others, and such coverage issues are difficult to avoid for large-scale bibliometric analyses such as these. For instance, Web of Science has uneven coverage of Computer Science conference proceedings, where the bulk of Computer Science publications take place. Since we control for discipline in most of our analyses, uniformly omitting publications within disciplines would not present an issue for our causal identification. We present a more detailed discussion of possible issues with confounding with Web of Science in the “Potential threats to causal identification” section below.

Linking employment records with labor availability data

We measure institutional-discipline and departmental counts of funded and unfunded labor through the NSF Survey of Graduate Students and Postdoctorates in Science and Engineering [28], which is an annual census administered to US academic institutions that grant research-based master’s and doctoral degrees in science, engineering, and certain health fields. The census is administered in the fall of the survey year, and has been continuously collected since 1966. Eligible institutions are determined through the federal Integrated Postsecondary Education Data System (IPEDS), and institutional coordinators collect the data from units within their institutions to report to the NSF. In 2020, 96.8%

of units provided complete or partial data, and 94.9% of institutions responded [28]. Missing data are imputed by the NSF.

As funded research labor, we counted all postdoctoral researchers and graduate students on a research assistantship, fellowship, or traineeship. As unfunded labor, we counted the remaining graduate students, which included teaching assistants and self-funded students. Prior to 2017, the data did not distinguish between master’s and doctoral students. In 2017, two versions of the data were available, one which separated master’s and doctoral students, one which did not. We consistently used the version of the data that did not distinguish between master’s and doctoral students, since our data spanned 2008–2017. We matched the data at the annual level for all of our employment records, and we averaged the statistics across time at the same time that we did so for other departmental and individual characteristics like counts of tenure-track faculty and productivities.

The data are provided with institution identifiers and codes identifying the field of study. We manually match each institution identifier with their corresponding identifier in the employment data, where possible, as well as matching the field codes with the disciplinary codes in the employment data. There could be many units within an institution with the same disciplinary label, and thus we perform a robustness test by analyzing the data at two levels: first, by only including the pairs of units where the NSF unit appears only once, since the NSF unit is at a higher level of aggregation ($N = 739$ units), which we call the strict linkage, and second, by accepting all matches and aggregating units up to the institution-discipline level ($N = 1800$ institution-discipline pairs), which we call the non-strict linkage. For the departmental regression models in the main paper, we use the strict linkage (Table S1), but we show that the results are robust to the non-strict linkage in Table S2. We present each regression model using both the strict and non-strict linkage, and the descriptive statistics using the non-strict linkage. Since there are few mid-career changes of institution, we use the non-strict linkage to perform the matching experiment to maximize statistical power. We also use the more non-strict linkage for the aggregated statistics that we compute for each prestige decile.

The main quantity of interest is the ratio of either funded or unfunded researchers to faculty in a department. We would like to take the logarithm of this ratio, but since there are departments with no funded graduate or postdoctoral researchers, this logarithm is improper without some smoothing. We accomplish this by including the tenure-track faculty themselves as funded researchers, so the funded faculty to labor ratio consists of the ratio of funded graduate and postdoctoral researchers plus the number of tenure-track faculty to the number of tenure-track faculty.

Inferring research group members from data

Research group sizes are difficult to measure accurately, even from seemingly ideal data sources like surveys and CVs. For example, for faculty with large research groups, students who pass through their lab for a rotation or who switch advisors halfway through their degrees may end up being omitted from that faculty’s memory, in the case of a single retrospective survey, and be omitted from that faculty’s CV. Moreover, there is no standard convention for including students and their names in faculty CVs, and when they are included, often only completion dates are included, if dates are included at all. Approximating the group size is useful as a control variable in our analysis, especially in the matching analysis, where we would like matched sets of people to have similar pre-move group sizes. Moreover, identifying potential group members at the individual level allows us to quantify their average productivities.

Further, since our concern is with the ways that students contribute to productivity, we seek to further restrict our attention to the active or productive research group of a faculty. For example, a graduate student who graduates with a masters degree without any publications does not contribute to the productivity of any faculty, and is thus not a part of anyone’s active research group.

This criteria, combined with the above difficulties on measuring group size, lead us to infer group members from publication data. Our heuristic is that a research group member is someone who is not tenure-track faculty, but who coauthors papers with a given faculty member using the same departmental affiliation. We use the Web of Science distinct author identification system ID to find all of the papers that any given group member coauthors in collaboration with a tenure-track faculty in their department. This includes funded graduate students from the same department, as well as postdocs, but also any unfunded graduate students on teaching assistantships or undergraduate students, as well as non-tenure-track faculty and research staff who coauthor papers with tenure-track faculty in their departments.

We compute the year span that a group member appears in the data, by subtracting the year of the last publication that they appear in a department from the year of the first publication, plus one. Since we are interested in the role of graduate and postdoctoral labor, rather than research staff or non-tenure-track faculty, we filter the non-faculty data to exclude people whose year span is 8 years or greater, and anyone whose first publication in a department is before 2007. We also exclude papers that group members author without any tenure-track faculty coauthors in their department.

Group members at more elite institutions tend to exhibit a slightly higher average year span than group members at less prestigious institutions, with each increase in prestige decile associated with an average increase of 0.043 years of year span (t-test, $p < 0.001$; Fig. S1A).

	(p/c)	(gp/c)	(gs/c)	(p/nc)	(gp/nc)	(gs/nc)
Is private	-0.06 (0.10)	-0.05 (0.09)	-0.07 (0.07)	-0.10* (0.05)	-0.50*** (0.11)	-0.19*** (0.03)
Num. faculty	0.08 (0.06)	0.12* (0.06)	0.10** (0.04)	-0.06 (0.04)	-0.16 (0.10)	-0.04 (0.04)
Prestige	0.06 (0.06)	0.04 (0.05)	0.07** (0.02)	0.07 (0.05)	0.05 (0.06)	-0.03 (0.05)
Unfunded labor ratio	-0.05 (0.07)	-0.01 (0.07)	-0.00 (0.05)	-0.02 (0.03)	-0.11 (0.07)	-0.04 (0.03)
Funded labor ratio	0.15*** (0.03)	0.17*** (0.03)	0.13*** (0.02)	0.02 (0.04)	0.06 (0.09)	0.09* (0.04)
Num. obs.	457	457	457	282	282	282
Deviance	263.57	194.68	572.19	84.49	50.28	100.33
Pseudo R ²	0.03	0.04	0.06	-0.02	-0.03	-0.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S1. **Poisson regression in departments using strict matches with the NSF data.** Results are separated by dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms), using only one-to-one matches between employment data and NSF GSS ($N = 739$). The funded labor covariate is the base 2 logarithm of the ratio of funded researchers (including tenure-track faculty) to the tenure-track faculty. The unfunded labor covariate is the base 2 logarithm of the ratio of unfunded researchers to the tenure-track faculty. All continuous variables are standardized to have zero mean and unit variance.

	(p/c)	(gp/c)	(gs/c)	(p/nc)	(gp/nc)	(gs/nc)
Is private	-0.05*** (0.01)	-0.06* (0.02)	-0.04 (0.02)	-0.03 (0.03)	-0.21 (0.15)	-0.12*** (0.03)
Num. faculty	0.07*** (0.02)	0.08*** (0.02)	0.09*** (0.01)	0.04 (0.08)	-0.02 (0.16)	0.11 (0.10)
Prestige	0.14*** (0.03)	0.14*** (0.03)	0.09*** (0.02)	0.07* (0.03)	0.00 (0.04)	0.02 (0.03)
Unfunded labor ratio	-0.06*** (0.02)	-0.06** (0.02)	-0.07*** (0.02)	-0.01 (0.03)	-0.02 (0.04)	-0.01 (0.03)
Funded labor ratio	0.12*** (0.02)	0.14*** (0.02)	0.13*** (0.01)	0.06 (0.04)	0.12* (0.05)	0.08*** (0.02)
Num. obs.	1293	1293	1293	507	507	507
Deviance	560.48	421.93	1141.05	124.00	84.78	183.88
Pseudo R ²	0.06	0.08	0.11	0.00	0.01	0.01

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S2. **Poisson regression in departments using non-strict matches with the NSF data.** Results are separated by dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms), allowing many-to-many matches between employment data and NSF GSS ($N = 1800$). Models are trained using the strict linkage data. The funded labor covariate is the base 2 logarithm of the ratio of funded researchers (including tenure-track faculty) to the tenure-track faculty. The unfunded labor covariate is the base 2 logarithm of the ratio of unfunded researchers to the tenure-track faculty. All continuous variables are standardized to have zero mean and unit variance.

This has implications for how we compute the average productivity of non-faculty in the data. To compute the average productivity for each non-faculty group member, we divide the total number of their publications by some range of time. If we use the empirical year span as the range of time, then non-faculty group members at less

elite institutions will have a smaller average denominator in their productivity, artificially increasing their productivity. For this reason, instead of using the year span directly, we use the minimum of the year span and some range of time, namely five years in Fig. 1A. To check that our results are insensitive to the choice of the minimum

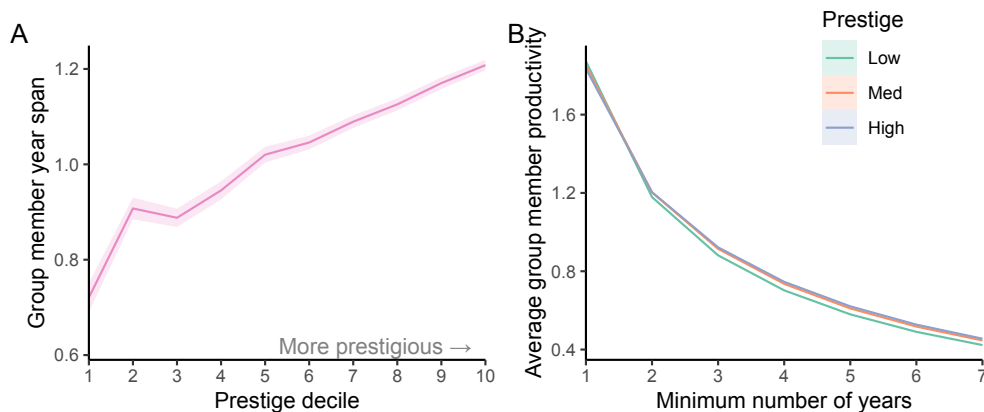


FIG. S1. **Group member productivity with different assumed career lengths.** (A) Non-faculty group members at more prestigious institutions tend to have a higher productive range in the data. Displayed are means with 95% confidence interval at each prestige decile. (B) When non-faculty have only one or two publications in the data in a short length of time, then we need to choose what we mean by their “average” productivity. In particular, we can impose a minimum length of time that we consider a group member to be present, filling in years outside of that range as zeros. As we increase the minimum span of time, the average productivity declines, but nonlinearly, since some, but not all, group members already appear for at least that length of time. We find that increasing the minimum span helps distinguish between the productivity of more prestigious and less prestigious group members. Displayed are means with 95% confidence intervals around the means for each minimum year span.

year span, we plot the average productivities of different prestige strata by the minimum year span (Fig. S1B).

The task of inferring group size from publication data requires assumptions about how the publication process works and how coauthorship norms work, and implies that non-publishing members of a group will be “invisible” to us because they do not appear in publication data. We find that using a window of three years allows us to count 94% of productive group members, and a window of four years allows us to count 97.6% of productive group members (Fig. S1B). Thus, using these cutoffs does, at least empirically, account for the vast majority of productive group members, but will, of course, miss any non-publishing group members.

For the matching, we derive a measure for the group size for each faculty, based on these potential group members. To reduce the extent to which our measure of group size is confounded by productivity, we window our measure of group size over several years; that is, we count the number of unique same-department non-faculty collaborators over a sliding window of k years. For example, certain research group members may not publish every year with faculty, but have gaps in their publication records. To select a useful window, we consider the tradeoffs between window length and coverage to pick the smallest acceptable window (Fig. S2). A window size of 3 only omits about 6% of all groups, which we view as an acceptable range of error for this analysis. To perform the actual construction, we slide this window of 3 or 4 years over the publication data, counting the number of unique group member collaborators of each faculty member in our dataset.

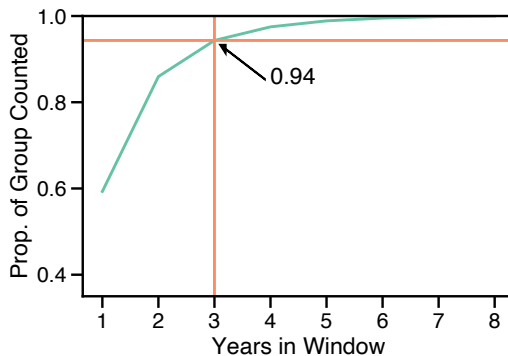


FIG. S2. **Choosing a window size for estimating group sizes.** Certain research group members may only publish with faculty every couple of years, and we would prefer to consider those people to be active group members during their inactivity, if they would later publish with the faculty again within a maximum length of time, which we set to be 7 years. Choosing a window size of 3 years includes 94% of all active group members, while moving to four years increases the coverage to 97.6%.

Inferring genders from names

We inferred a gender (man, woman, or unknown) for each faculty member using first and last names. First, we checked complete names against two offline dictionaries: a hand-annotated list of faculty employed at Business, Computer Science, and History departments [7], and the open-source python package gender-guesser [47]. These dictionaries assigned each full name as either fe-

male, male, or unable to classify. Second, where the dictionaries disagreed or where either dictionary was unable to assign a gender to the name, we queried Ethnea [48] and used the gender they assigned the name. Using this approach we were able to assign genders to 88.3% of faculty. Faculty whose names could not be associated with a gender were excluded from the individual-level regressions but still included in other analyses. We recognize that gender is nonbinary, although this procedure assigns binary (woman/man) labels to faculty. This is a compromise between the technical limitations of name-based gender inference and the importance of studying gender inequality in science, and it is not intended to reinforce the gender binary.

Counting cumulative productivity and cumulative group size

We generate cumulative counts of productivity and group size in order to plot the relationship between the two across prestige strata (Fig. 2C). To start, we need to pick a reference point for starting the count t_0 , which we take as the latter of either their degree year or the first year that someone appears in Web of Science with the same affiliation as the record we have in the employment data. Then, to compute cumulative (group) size for a given year $t > t_0$, we add up the number of publications they have starting from the first year they appear in the data. To count cumulative group size, we maintain a record of all of the past group members (same-department non-faculty collaborators) in a hash set data structure that have been seen up to year t , and we count the number of unique elements in that set. We assume our publication data is complete in the sense that in years where we have no data for faculty in Web of Science, we assume they publish zero papers, and introduce zero new group members.

Disciplines with and without collaboration norms

The labor advantage mechanism depends on graduate, postdoctoral, and other non-faculty labor contributing to faculty productivity. Norms differ across academic disciplines on how closely faculty collaborate with research group members on projects, and even among disciplines where such collaborations do occur, on whether faculty tend to share credit with students in the form of coauthorships. We have relied on sociological evidence to separate disciplines into these two categories in Table S3, with broadly social sciences, humanities, and mathematics as the disciplines without research group collaboration norms, and the natural and medical sciences and engineering as the disciplines with collaboration norms [18–24].

The concept of an academic discipline has been continuously redefined by practitioners [49], and our use of the

Disciplines with collaboration norms	Disciplines without collaboration norms
Biological Sciences	Economics
Engineering	Mathematical Sciences
Medical Sciences	Language, Literature, Culture
Psychological Sciences	Political Science
Physical Sciences	History
Chemical Sciences	Anthropology
Computational Science	Sociology
Health	Education
Business	Philosophy
Earth Sciences	Arts
Agriculture	Theology and Religion
Architecture, Design, Planning	Geography
	Linguistics

TABLE S3. **Disciplines by collaboration norms.** Disciplines are separated into those where research groups tend to collaborate with principal investigators on projects, contributing to the PI’s productivity (left) and those where those expectations aren’t widely held (right).

word “discipline” may be closer to what others may call a “disciplinary grouping”, where we combine, for example, computer and information sciences into “computational sciences”. To test the sensitivity of the results to our choice of aggregation, we perform a separate bipartition at the more detailed taxonomic level of a “field”, which we use as a term to be more granular than “discipline” [50, 51]. The results were robust to this finer partitioning scheme, and the main results in the paper were given in terms of the broader partitions. In the field bipartition, all of engineering and natural sciences remained as having collaboration norms.

- We identified as fields with collaboration norms: General Psychology, Neuroscience, Computer Science, Statistics, Biostatistics, Information Science, Finance, Animal Science, Food Science, Agronomy, Soil Science, Nursing, Veterinary Medical Sciences, Physiology, Pharmacy, Epidemiology, Communication Disorders and Sciences, Nutrition Sciences, and Pharmacology.
- The remaining fields without collaboration norms were then Mathematics, Political Science, General Economics, Sociology, Anthropology, Geography, and Agricultural Economics.

MODELING DETAILS

Matching

To perform our matching analysis, we first construct a dataset of mid-career moves. First, we expand our consideration to include 25 disciplines, rather than only the 17 in the NSF GSS, to observe as many mid-career moves as possible. However, this eliminates our ability to use available departmental labor as a matching variable. Thus, we use prestige as a proxy for labor availability. Since our dataset consists of a longitudinal census of faculty between the years 2011 and 2017, we can look for people who changed institutions during that time. We avoid looking at cases where someone only changed departments within the same institution, since that may not be a sufficiently large change in terms of working conditions and available resources. We require that faculty only move once during the period of observation, that they be present in at least four years of employment data, and that they stay within the same academic discipline before and after their move. This leaves 5,709 mid-career moves. However, not all of these people have publication records available in WoS for the years before and after their move. Once we impose the requirement that people have publications within the four years before their move, as well as two years after their move, we are left with 2,316 people.

The ideal experiment to assess the impact of available labor on productivity would be to randomly allocate funded graduate and postdoctoral labor to departments. Because such an experiment is impractical, we consider an observational study based on faculty mobility instead: we use mid-career moves in our dataset as a quasi-random allocation of faculty into high and low labor availability environments. However, we only have departmental labor data for the disciplines that are included in the NSF survey, and we have few enough mid-career moves that we prefer to use as much data as possible. To that end, we rely on prestige as a proxy for labor availability in the matching (see Supplementary Material), and the treatment condition consists of faculty who move upwards in the prestige hierarchy, and the control condition are faculty who move to a less prestigious institution. The outcome variable is the productivity with group members, which indirectly measures the labor advantage effect, since group members themselves are not more productive at elite institutions. In the matching, we control for academic discipline, rank (assistant, associate, or full professor), pre-move prestige, pre-move productivity, and pre-move group size. Then, we look at with and without group productivity two years after the move. Picking two years allows us to avoid the noise associated with the move itself, while leaving enough people in to make a precise estimate.

We first attempted 1:1 nearest neighbor propensity score matching, but found inadequate balance. Instead, we used full matching on the propensity score, which

yielded adequate balance. We find similar distributions of propensity scores between the matched control and treated units (Fig. S3). Prior to matching, we had substantial imbalance in the pre-move prestige between cohorts, and slight imbalance in total and group productivity. The difference in the propensity score (“distance”) between groups before matching was also substantial. After matching, imbalance in all covariates and the propensity score is reduced to an acceptable threshold of below 0.1 mean difference. (Fig. S11).

To estimate the treatment effect of prestige on research group size, we used a linear regression on the number of papers published with department non-faculty collaborators as the outcome, a binary indicator variable whether the place that a faculty moved to was higher prestige as the treatment, and including covariates from before the move such as productivity, productivity with group members, prestige, research discipline, unique same-department non-faculty collaborators on papers, gender, and title. We include the full matching weights in the estimation. We estimate standard errors from a cluster-robust variance using the matching stratum membership as the clustering variable, and perform a t-test on whether the coefficient of the treatment variable is statistically significantly different from zero.

We perform two additional robustness tests. First, we considered the impact of midcareer moves on group productivity rather than group size, to confirm that the effect of moving to an environment with more available funded labor drives increased group productivity. We find this to be the case (Fig. S4; t-test, $p < 0.001$). Second, we also considered the impact on productivity three years, instead of two, after the move. Since this requires an additional year of data, our sample size decreased from 2,316 to 805, with 417 (51.8%) moves up the prestige hierarchy. We followed the exact same matching procedure as before, and we found adequate balance due to the full propensity score adjustment. Similar to the two year case, we found an effect on group productivity of moving up rather than down the prestige hierarchy, but the signal became weaker, with upward movers publishing 0.498 more papers on average three years after moving than downward movers (t-test, $p = 0.00062$). On the other hand, an effect on individual productivity appeared, with individual productivity for upward movers exceeding downward movers by 0.471 on average on the third year after the move (t-test, $p < 0.001$). Further investigation would be necessary to illuminate whether this is due to noise, or whether it reflects faculty choosing to prioritize the ramping up of a new research group after moving before working on individual publications.

We found that faculty tended to have smaller groups after they move than before they move in general, reflecting a general decrease or “shock” to group size immediately after making a mid-career move. This suggested that matching faculty who make mid-career moves to control faculty who do not move would not be useful, since we would likely find decreases in both conditions.

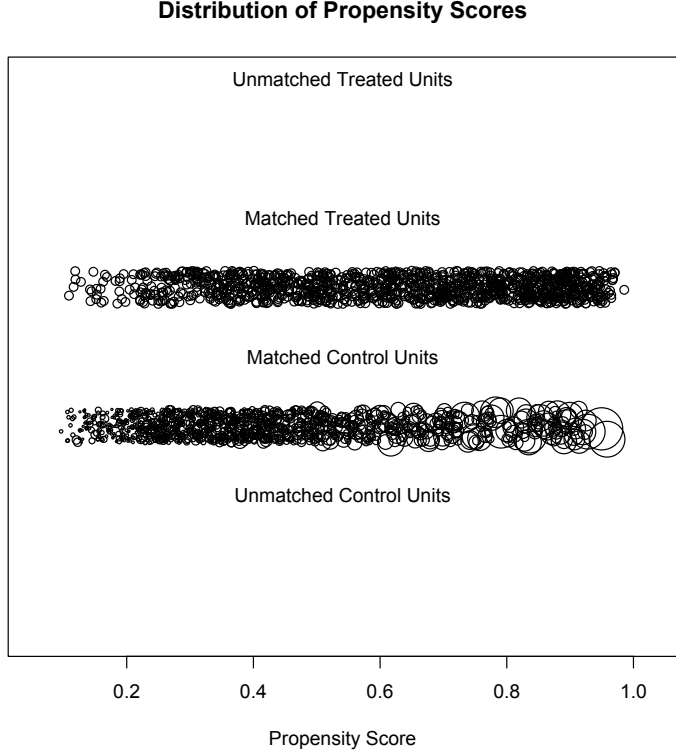


FIG. S3. **Distribution of propensity scores across treatment and control conditions.** The propensity score of matched treated and control units are displayed with vertical jitter along the horizontal axis. The size of the circle corresponds to the weight of the point, where points with higher weights are displayed with larger circles. Since we use full propensity score matching, no treatment or control units are unmatched. The treatment and control groups exhibit a similar distribution of propensity scores after the full propensity score matching.

Regression specifications

An ideal dataset for analyzing the impact of labor on productivity would be one that reported the research group size of each faculty member for each year. However, such a dataset would be difficult to define and construct, since the exact dates that students or postdocs begin working with a faculty member are not necessarily clear: an informal advising relationship may progress into a formal one after a trial period, for example through a series of laboratory rotations during the beginning of a graduate program, or from the student attending the class of a faculty and being advised on a class project. From a data gathering perspective, faculty often only report the graduation year of their PhD students on their CVs, reflecting and compounding on the difficulty of pinning down the start of these formal relationships. On the other hand, aggregated departmental data on labor availability can be reliably sourced from our census of tenure-track faculty and NSF survey data for certain disciplines. We thus perform an analysis at the aggregated department level, where we use a regression to consider the relative impact of departmental variables such as the ratio of funded (and unfunded) researchers to faculty within

the department on departmental productivity.

We summarize how each of the departmental variables are measured:

- Funded labor availability is the base-2 logarithm of the ratio of funded researchers (including faculty) to faculty in an institution-discipline.
- Unfunded labor availability is the base-2 logarithm of the ratio of unfunded graduate students to faculty in an institution-discipline, and 72 within-discipline institutions (9.74%) that had no unfunded graduate students in any years were omitted.
- The log department size is the base-2 logarithm of the number of faculty in an institution-discipline.
- The group size for each professor at year t was measured by counting unique same-address non-faculty coauthors on their papers for a 3 year period ending in year t . This window size captures 94% of productive same-address group members, which minimizes the extent to which our measure of productive group size is confounded by productivity (see

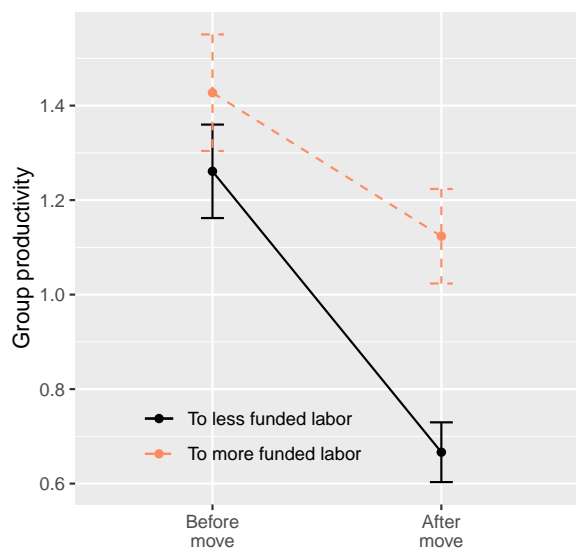


FIG. S4. **Matching on midcareers using group productivity as the dependent variable.** For matched-pairs of faculty, mean group productivity in the 3 years before and after moving to a location with more (dashed orange) or less (solid black) available funded labor than their pre-move location. Error bars indicate one standard error. Faculty who move to work environments with greater labor availability exhibit higher post-move group productivity than their matched peers who move to environments with less labor availability ($p < 0.001$).

Supporting Information). Then for each department, we use the average group size of faculty in that department.

- Prestige is computed within disciplines for each year of our data, and averaged across all years.
- We include an indicator variable from the Department of Education College Scorecard on whether each institution is private, where it is set to 0 if it is public, and 1 if it is private.

All variables were transformed in the following way, with steps performed sequentially: averaged over faculty within institution-discipline, then averaged across years, then transformed using the logarithm, then standardized and rescaled to have mean 0 and variance 1.

To test the robustness of our regression specification, we run alternate models using individual-level data. Associating individuals with additional metadata relevant to their productivity, such as their last known rank, their gender inferred from their name, and the number of years since their degree, we average their productivity to create a dataset for predicting individual average productivity from both individual and institutional characteristics. This produces three nested levels: individual, institution-discipline, and discipline. A simple model with few assumptions is a Poisson regression where disci-

plines are controlled using fixed effects, and then we adjust the standard errors to account for clustering at the institution-discipline level (Tables S4,S5; Fig. S7). Another natural choice, though with more assumptions, is to model the hierarchy explicitly through a generalized hierarchical linear model, where institution-disciplines and disciplines are the two levels (Tables S6,S7; Fig. S8). In both models, we find that funded labor is statistically and practically significant in explaining the dependent variables, although only in disciplines with collaboration norms. We find similar results for the role of funded labor availability in models trained on both the strict and non-strict data, except that the regressions using non-strict data show a greater role for prestige. This could be because the strict matching provides more accurate measurement of funded labor availability, while the non-strict matching introduces more measurement error.

In addition to the variables present in the departmental analysis, our individual and hierarchical models include the following individual-level variables:

- “Is man” is a binary indicator variable from the result of name-gender disambiguation on individual faculty, which is set to 1 if the faculty’s name is confidently associated with men, while 0 if confidently associated with being women. We drop the 11.8% of faculty whose names were not in either category.
- “Years since degree” is the average number of years since the individual received their PhD degree in the data.
- “Is assoc. prof” and “Is full prof” are binary indicator variables representing the highest title held by the individual in the time spanned by the data, where both values are set to 0 if the individual is an assistant professor during the entire time span.

In all of our analyses, we average over time to smooth over potentially complicated and long-term dynamics between prestige and the availability of funded labor.

Potential threats to causal identification

We encode our assumptions into a causal diagram (Fig. S5), which shows three main threats toward causal identification: the variables that are unobserved to us but available to the hiring committee, additional pathways from the prestige of the environment to productivity that do not operate through funded labor, and unobserved missingness within Web of Science or our data linkage process.

In the descriptive decomposition of productivity into individual and group productivity, and disciplines into those with group collaboration norms and those without (Fig. 1), we present a specific set of empirical observations that theories about the relationship of prestige and productivity must explain. In particular, even if

	(gs/c)	(gs/c)	(gs/c)	(gs/nc)	(gp/c)	(gp/nc)	(p/c)	(p/nc)
Is full prof.	0.62*** (0.04)	0.63*** (0.04)	0.62*** (0.04)	0.38*** (0.05)	0.76*** (0.04)	0.82*** (0.11)	0.66*** (0.04)	0.32*** (0.06)
Is assoc. prof.	0.26*** (0.03)	0.25*** (0.03)	0.26*** (0.03)	0.18*** (0.04)	0.27*** (0.04)	0.38*** (0.10)	0.18*** (0.03)	-0.03 (0.06)
Years since degree	-0.24*** (0.02)	-0.25*** (0.02)	-0.24*** (0.02)	-0.14*** (0.02)	-0.34*** (0.02)	-0.30*** (0.04)	-0.30*** (0.02)	-0.23*** (0.02)
Log dept. size	0.07*** (0.02)	0.01 (0.02)	0.07*** (0.02)	-0.01 (0.03)	0.07** (0.03)	-0.03 (0.05)	0.06** (0.02)	-0.00 (0.02)
Unfunded labor availability	-0.01 (0.02)	0.03 (0.02)	-0.01 (0.02)	-0.08 (0.04)	0.03 (0.03)	-0.11 (0.08)	0.03 (0.02)	0.01 (0.04)
Prestige		0.11*** (0.02)	0.02 (0.03)	-0.09** (0.03)	0.05* (0.03)	-0.11* (0.05)	0.09*** (0.02)	0.02 (0.03)
Funded labor availability	0.18*** (0.03)		0.17*** (0.03)	0.14*** (0.04)	0.20*** (0.04)	0.19* (0.08)	0.15*** (0.03)	0.08* (0.04)
Is man	0.08** (0.03)	0.09*** (0.03)	0.08** (0.03)	0.07 (0.04)	0.14*** (0.03)	0.12 (0.08)	0.14*** (0.03)	0.14*** (0.04)
Num. obs.	10360	10360	10360	4262	10360	4262	10360	4262
Num. groups: Area	9	9	9	6	9	6	9	6
Deviance	49032.04	50021.68	49020.41	4353.80	15527.83	2492.11	19320.49	4076.48
Pseudo R ²	0.08	0.06	0.08	0.02	0.09	0.02	0.07	0.02

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S4. **Poisson regression on individual faculty using strict linkages.** As an alternate specification to the institution-discipline regressions, here we present the regression coefficients from a regression on individual faculty using the strict linkage data. We use a Poisson regression at the individual faculty level controlling for discipline using fixed effects and clustering standard errors within departments, we report the coefficients of standardized (zero mean and unit variance) individual, departmental, and institutional covariates in predicting departmental productivity, group productivity, and group sizes, in disciplines with and without collaboration norms, with 95% confidence intervals. The model names correspond to the dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms). We also include versions of the p/c model without the prestige and labor variables. All continuous variables are standardized to have zero mean and unit variance.

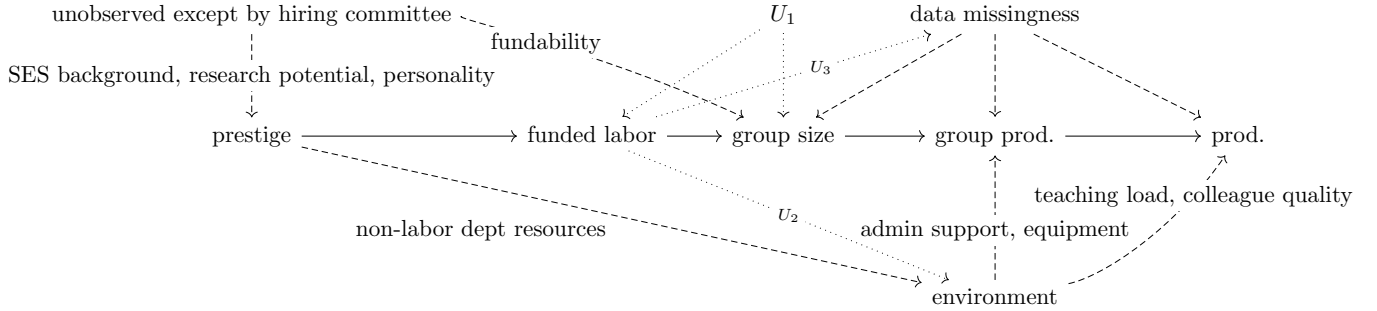


FIG. S5. **Detailed causal diagram with confounds.** The solid arrows display the main causal pathways of interest (as in Fig. 3). The dashed arrows denote possible unobserved confounds, which we annotated with plausible reasons for why they may exist. The main threats to causal identification are labeled U_1 , U_2 and U_3 with dotted arrows.

we assume (implausibly) that selection committees have perfect foresight into new hires' future funding, group sizes, and productivity, there remains the observation that the relationship between average individual productivity and prestige appears very similar across the two categories of disciplines. In conjunction with the obser-

vation that group members themselves tend to exhibit equal average productivity across prestige (Fig. 1A), the theoretical conclusion that we draw is that individuals are constrained by fairly general limitations on their productivity, while expanding their group size allows them to exceed those limitations (exacerbating prestige inequal-

	(gs/c)	(gs/c)	(gs/c)	(gs/nc)	(gp/c)	(gp/nc)	(p/c)	(p/nc)
Is full prof.	0.64*** (0.02)	0.63*** (0.02)	0.63*** (0.02)	0.43*** (0.05)	0.76*** (0.02)	0.82*** (0.09)	0.69*** (0.02)	0.40*** (0.05)
Is assoc. prof.	0.25*** (0.02)	0.25*** (0.02)	0.25*** (0.01)	0.22*** (0.04)	0.24*** (0.02)	0.41*** (0.07)	0.18*** (0.02)	0.05 (0.04)
Years since degree	-0.23*** (0.01)	-0.23*** (0.01)	-0.23*** (0.01)	-0.16*** (0.02)	-0.33*** (0.01)	-0.28*** (0.03)	-0.31*** (0.01)	-0.25*** (0.02)
Log dept. size	0.07*** (0.01)	0.02 (0.01)	0.04** (0.01)	0.20 (0.11)	0.03* (0.01)	0.14 (0.13)	0.04** (0.01)	0.06 (0.06)
Unfunded labor availability	-0.07*** (0.02)	-0.01 (0.02)	-0.05** (0.02)	-0.00 (0.03)	-0.03 (0.02)	0.01 (0.04)	-0.03 (0.02)	0.00 (0.02)
Prestige		0.13*** (0.01)	0.10*** (0.01)	0.03 (0.02)	0.15*** (0.01)	0.03 (0.04)	0.16*** (0.01)	0.10*** (0.02)
Funded labor availability	0.14*** (0.01)		0.08*** (0.01)	0.02 (0.02)	0.07*** (0.02)	0.03 (0.04)	0.06*** (0.02)	0.01 (0.02)
Is man	0.12*** (0.01)	0.13*** (0.01)	0.13*** (0.01)	0.02 (0.04)	0.18*** (0.02)	0.05 (0.07)	0.18*** (0.01)	0.11*** (0.03)
Num. obs.	50991	50991	50991	8731	50991	8731	50991	8731
Num. groups: Area	11	11	11	6	11	6	11	6
Deviance	222187.94	220928.68	220136.12	11924.51	73852.69	5910.75	93909.50	9485.68
Pseudo R ²	0.07	0.07	0.08	0.04	0.08	0.05	0.07	0.03

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S5. **Poisson regression on individual faculty using non-strict linkages.** As an alternate specification to the institution-discipline regressions, here we present the regression coefficients from a regression on individual faculty using the non-strict linkage data. We use a Poisson regression at the individual faculty level controlling for discipline using fixed effects and clustering standard errors within departments, we report the coefficients of standardized (zero mean and unit variance) individual, departmental, and institutional covariates in predicting departmental productivity, group productivity, and group sizes, in disciplines with and without collaboration norms, with 95% confidence intervals. The model names correspond to the dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms). We also include versions of the p/c model without the prestige and labor variables. All continuous variables are standardized to have zero mean and unit variance.

ities).

Variables observed by the hiring committee could potentially confound the matching, whereby faculty are selected to work at more elite institutions due to the hiring committee’s insight into the faculty candidate’s active grants or their ability to perceive the future success of faculty candidates at acquiring grants or otherwise recruiting group members. However, such unobserved variables only act as confounds insofar as they are uncorrelated with the observed variables.

By reaching qualitatively similar conclusions through both regression and matching, our claim that increased available funded labor drives larger faculty group sizes is made more robust, since its validity depends on fulfilling only one set of assumptions (for either the regression or the matching). Our regression analysis is less sensitive to the possibility of confounding by the hiring committee, since we close the backdoor path from funded labor to prestige to the hiring committee to group size by conditioning on prestige. The types of confounders that would threaten the identification in our regression model would need to impact departmental funded labor and group size

independently of prestige (labeled U_1 in Fig S5).

In other words, for our regression result to be confounded by hiring committee, it is not sufficient for hiring committees to place the most future-productive people into the most prestigious institutions. Instead, the selection committee must place the most future-group-productive people into work environments with more available funded labor even after controlling for prestige. This is a more stringent assumption, but is not impossible to imagine. For instance, researchers in disciplines with group collaboration norms who want to construct large research groups may choose between working at two institutions with similar prestige, and decide to join the one with more funded researchers per faculty (or some latent quality associated with greater available funded labor). However, importantly, this selection effect must occur orthogonally to prestige, and it must face the challenge of our matching analysis in turn: these two researchers who exhibit vastly different propensities for constructing research group sizes (perhaps due to their subfields) must have led similar careers up to that point in terms of their group sizes and productivity. Thus our theory is threat-

	(gs/c)	(gs/c)	(gs/c)	(gs/nc)	(gp/c)	(gp/nc)	(p/c)	(p/nc)
Is full prof.	0.56*** (0.01)	0.56*** (0.01)	0.56*** (0.01)	0.35*** (0.05)	0.70*** (0.03)	0.81*** (0.11)	0.60*** (0.02)	0.31*** (0.05)
Is assoc. prof.	0.23*** (0.01)	0.23*** (0.01)	0.23*** (0.01)	0.16*** (0.04)	0.24*** (0.03)	0.37*** (0.10)	0.16*** (0.02)	-0.03 (0.05)
Years since degree	-0.23*** (0.01)	-0.23*** (0.01)	-0.23*** (0.01)	-0.13*** (0.02)	-0.32*** (0.01)	-0.29*** (0.04)	-0.27*** (0.01)	-0.22*** (0.02)
Log dept. size	-0.11*** (0.01)	-0.18*** (0.01)	-0.11*** (0.01)	-0.02 (0.02)	-0.02 (0.02)	-0.05 (0.05)	-0.04* (0.02)	-0.03 (0.03)
Unfunded labor availability	0.11*** (0.01)	0.24*** (0.01)	0.10*** (0.01)	-0.01 (0.03)	0.14*** (0.02)	-0.00 (0.07)	0.15*** (0.02)	0.08* (0.04)
Prestige		0.11*** (0.02)	-0.02 (0.02)	-0.08** (0.03)	-0.01 (0.03)	-0.12 (0.06)	0.03 (0.02)	0.01 (0.03)
Funded labor availability	0.33*** (0.02)		0.33*** (0.02)	0.16*** (0.04)	0.36*** (0.03)	0.21** (0.08)	0.32*** (0.02)	0.11** (0.04)
Is man	0.07*** (0.01)	0.08*** (0.01)	0.07*** (0.01)	0.08** (0.03)	0.16*** (0.02)	0.16* (0.07)	0.16*** (0.02)	0.14*** (0.04)
# groups: Dept.:Disc.	686	686	686	299	686	299	686	299
# obs.	10360	10360	10360	4262	10360	4262	10360	4262

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S6. **Hierarchical Poisson regression on individual faculty using strict linkages.** As an alternate specification to the institution-discipline regressions, here we present the regression coefficients from a hierarchical Poisson regression on individual faculty using the strict linkage data. The model names correspond to the dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms). We also include versions of the p/c model without the prestige and labor variables. All continuous variables are standardized to have zero mean and unit variance.

ened only through the combination of two separate selection effects: (a) researchers must be selected for working at locations with greater available labor (and likely prestige) in ways predicted by the hiring committee that have not been visible in their publication record or group size prior to the move, and (b) researchers with greater preference or potential for large-group high-productivity research must select for institutions with greater funded labor availability, orthogonally to prestige. Again, this is not unimaginable, if for instance, one of the matched faculty performs a midcareer move in order to switch subfields, where the new subfield has different expectations around research group sizes and productivity. However, we find this situation specific enough that we do not believe it confounds our analyses, even if we cannot rule out the possibility without further research.

The backdoor path through the environment variable is also closed by conditioning on prestige, unless there is a path from departmental funded labor to environmental resources independent of prestige (U_2 in the diagram). Such a confound may look something like the following: perhaps having more funded researchers per faculty in the department (independent of both prestige and the number of tenure-track faculty, which are controlled for) allows the group members to overcome research obstacles faster by discussing with one another, and thus improves their productivity. However, the threat of such a path is

limited by our observation in Fig. 1A that group members have similar average productivity across prestige, which we know is highly correlated with funded labor availability.

Certainly Web of Science has imperfect coverage, and our data linkage has introduced missingness among publication records (and thus group size and group productivity counts), which we represent as dashed lines from data missingness to the group size, group productivity, and productivity variables. However, these gaps are only a threat to causal identification if it lies on the causal path of interest in a way that cannot be solved by controlling for prestige (U_3). In our matching analysis, for missingness to become a serious issue, faculty who otherwise publish in similar ways prior to mid-career moves would need to alter the types of publication venues in which they submit their manuscripts such that moving to a place with greater or less availability funded labor is associated with a greater probability of submitting a manuscript to a publication venue that is not indexed by Web of Science. For the regression, since we condition on prestige directly, Web of Science coverage only poses a problem if it selectively indexes journals published at institutions with more or less labor availability even after controlling for discipline and prestige. We do not find either of these scenarios particularly plausible, since the most well-known patterns of missingness in Web of

	(gs/c)	(gs/c)	(gs/c)	(gs/nc)	(gp/c)	(gp/nc)	(p/c)	(p/nc)
Is full prof.	0.65*** (0.01)	0.65*** (0.01)	0.65*** (0.01)	0.44*** (0.03)	0.59*** (0.01)	0.87*** (0.07)	0.69*** (0.01)	0.41*** (0.04)
Is assoc. prof.	0.28*** (0.01)	0.28*** (0.01)	0.28*** (0.01)	0.22*** (0.03)	0.21*** (0.01)	0.43*** (0.07)	0.20*** (0.01)	0.06 (0.03)
Years since degree	-0.23*** (0.00)	-0.23*** (0.00)	-0.23*** (0.00)	-0.15*** (0.01)	-0.24*** (0.00)	-0.28*** (0.03)	-0.31*** (0.00)	-0.24*** (0.01)
Log dept. size	-0.01 (0.01)	-0.04*** (0.01)	-0.03** (0.01)	-0.02 (0.06)	0.03* (0.01)	0.03 (0.12)	0.03** (0.01)	-0.01 (0.06)
Unfunded labor availability	-0.02** (0.01)	-0.01 (0.01)	-0.02** (0.01)	-0.03 (0.02)	0.01 (0.01)	-0.02 (0.04)	0.01 (0.01)	-0.01 (0.02)
Prestige		0.14*** (0.01)	0.13*** (0.01)	0.09*** (0.02)	0.14*** (0.01)	0.10* (0.04)	0.16*** (0.01)	0.15*** (0.02)
Funded labor availability	0.07*** (0.01)		0.05*** (0.01)	-0.01 (0.02)	0.09*** (0.01)	0.00 (0.04)	0.06*** (0.01)	-0.03 (0.02)
Is man	0.11*** (0.00)	0.11*** (0.00)	0.11*** (0.00)	0.08*** (0.02)		0.12* (0.05)	0.18*** (0.01)	0.13*** (0.03)
# groups: Dept.:Disc.	3750	3750	3750	635	3865	635	3750	635
# obs.	50991	50991	50991	8731	67707	8731	50991	8731

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S7. **Hierarchical Poisson regression on individual faculty.** As an alternate specification to the institution-discipline regressions, here we present the regression coefficients from a hierarchical Poisson regression on individual faculty using the non-strict linkage data. The model names correspond to the dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms). We also include versions of the p/c model without the prestige and labor variables. All continuous variables are standardized to have zero mean and unit variance.

Science occur across disciplines, rather than within disciplines. Missingness across disciplines may weaken our descriptive analyses, or introduce noise into our inferential studies, but do not fundamentally threaten the causal inference.

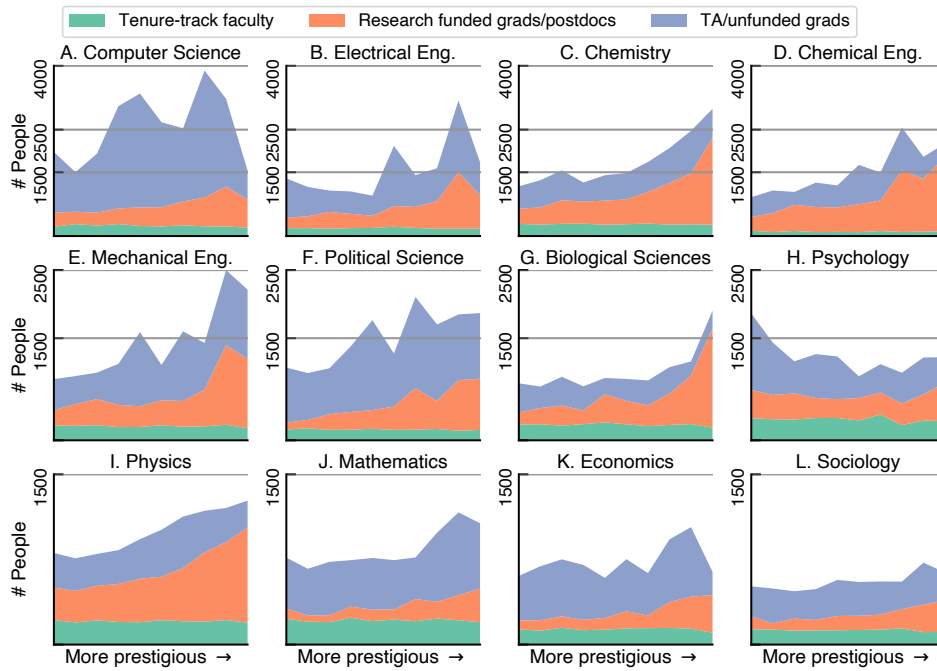


FIG. S6. **Labor advantages exist in absolute terms across research fields.** We group departments within each field into prestige deciles with the goal of keeping the counts of tenure-track faculty in each decile roughly the same, to highlight the ratio of funded and unfunded labor to faculty. Across most fields, the ratio of funded graduate students and postdoctoral researchers (orange) to faculty increases steadily with prestige, while graduate teaching assistants and self-funded graduate students (purple) remains relatively constant with prestige.

	(first/c)	(last/c)	(first/nc)	(last/nc)
Is private	-0.05 (0.15)	-0.16* (0.07)	0.07 (0.05)	-0.00 (0.07)
Num. faculty	0.01 (0.07)	0.07 (0.05)	-0.05 (0.05)	-0.03 (0.05)
Prestige	0.02 (0.07)	0.17*** (0.04)	0.04 (0.05)	0.09*** (0.03)
Unfunded labor ratio	-0.09 (0.07)	-0.01 (0.06)	0.01 (0.03)	-0.03 (0.02)
Funded labor ratio	0.01 (0.04)	0.10** (0.03)	-0.03 (0.04)	0.00 (0.04)
Num. obs.	457	457	282	282
Deviance	47.56	110.54	41.23	36.25
Pseudo R ²	-0.02	0.05	-0.04	-0.04

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S8. **Poisson regression in departments for first and last author publications.** Results are separated by dependent variable (first=annual papers authored as first author, last=annual papers authored as last author) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms). The funded labor covariate is the base 2 logarithm of the ratio of funded researchers (including tenure-track faculty) to the tenure-track faculty. The unfunded labor covariate is the base 2 logarithm of the ratio of unfunded researchers to the tenure-track faculty. All continuous variables are standardized to have zero mean and unit variance.

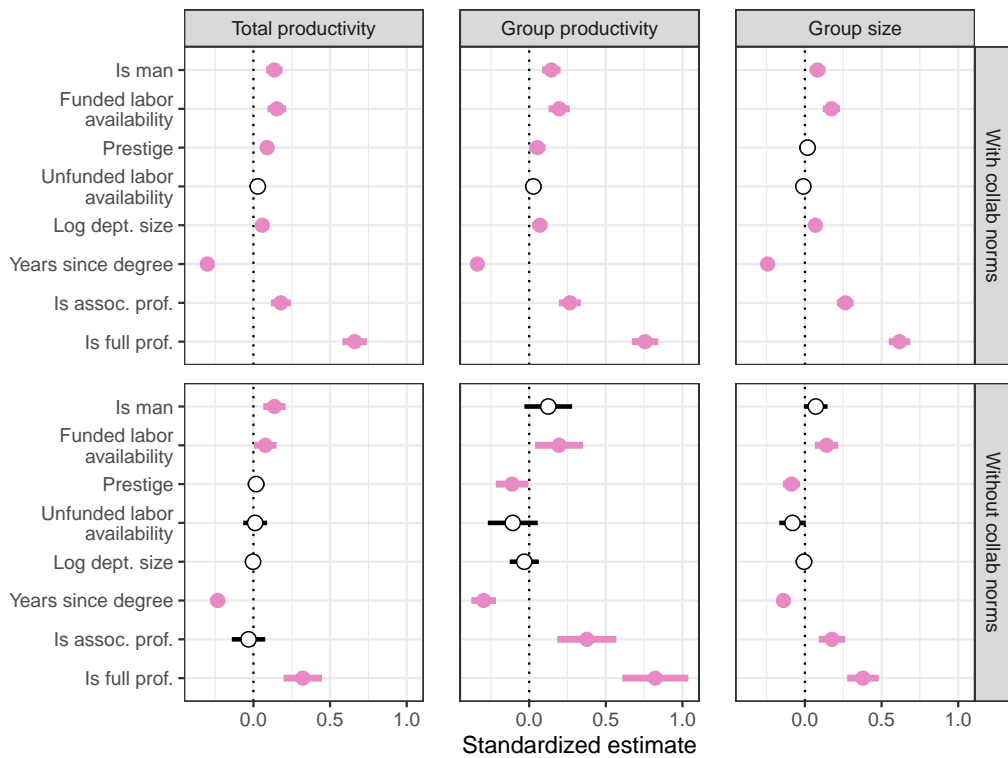


FIG. S7. **Coefficients of individual-level regression.** Performing a Poisson regression at the individual faculty level controlling for discipline using fixed effects and clustering standard errors within departments, we report the coefficients of standardized (zero mean and unit variance) individual, departmental, and institutional covariates in predicting departmental productivity, group productivity, and group sizes, in disciplines with and without collaboration norms, with 95% confidence intervals using the strict linkage data. Statistically significant coefficients at $p < 0.05$ are depicted in pink with a filled-in circle. The availability of funded labor has a significant impact on all dependent variables, even after controlling for prestige, especially in disciplines with collaboration norms.

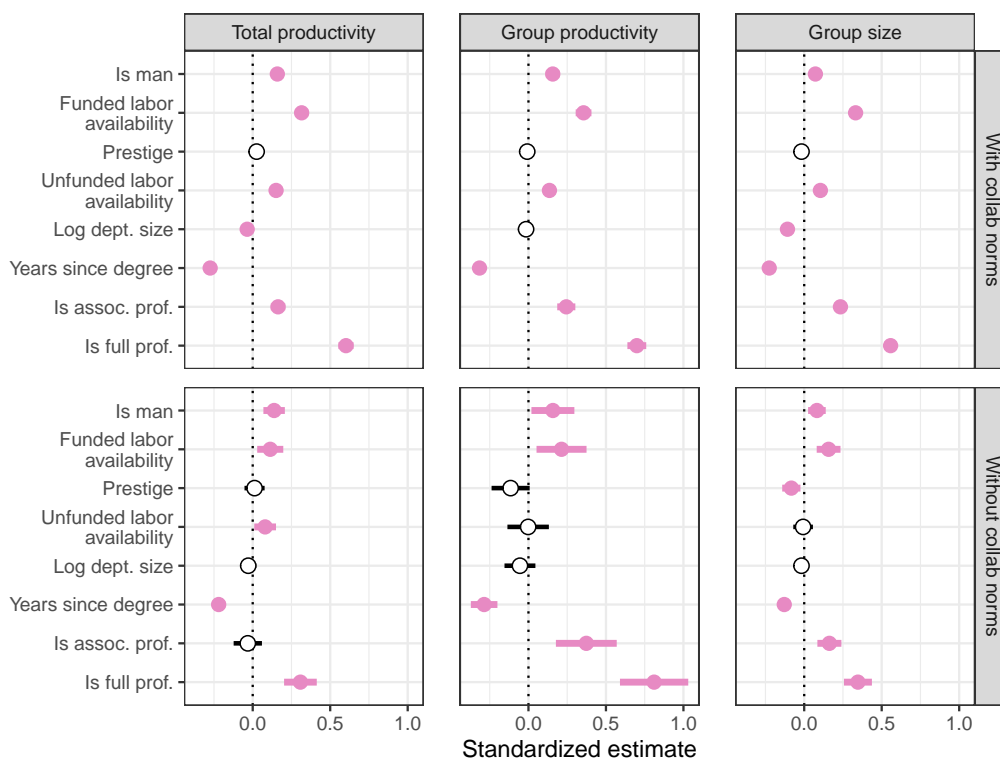


FIG. S8. **Coefficients of hierarchical regression.** Performing a hierarchical regression using individual faculty as our unit, we report the coefficients of standardized (zero mean and unit variance) individual, departmental, and institutional covariates in predicting departmental productivity, group productivity, and group sizes, in disciplines with and without collaboration norms, with 95% confidence intervals using the strict linkage data. Statistically significant coefficients at $p < 0.05$ are depicted in pink with a filled-in circle. The availability of funded labor has a significant impact on all dependent variables, even after controlling for prestige, especially in disciplines with collaboration norms.

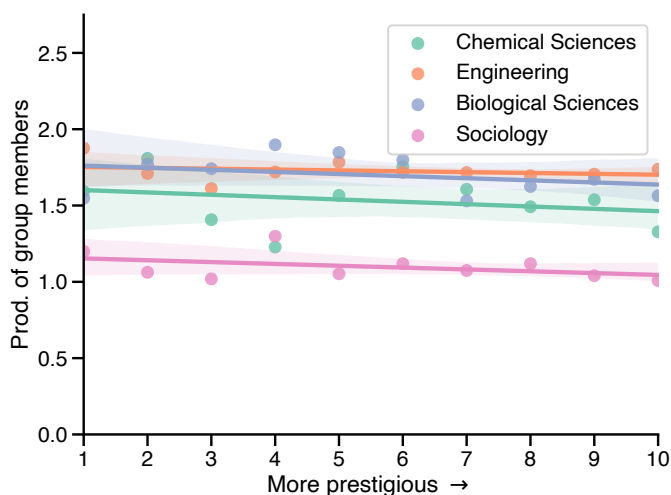


FIG. S9. **Non-faculty are roughly equally productive across prestige, across disciplines.** Disaggregating non-faculty into four disciplines, using no minimum year span, we find that non-faculty collaborators of faculty do not publish more papers in collaboration with faculty in their department at more elite institutions. Displayed are average productivities for each prestige decile, and the 95% confidence interval for the ordinary linear regression through the ten points for each discipline.

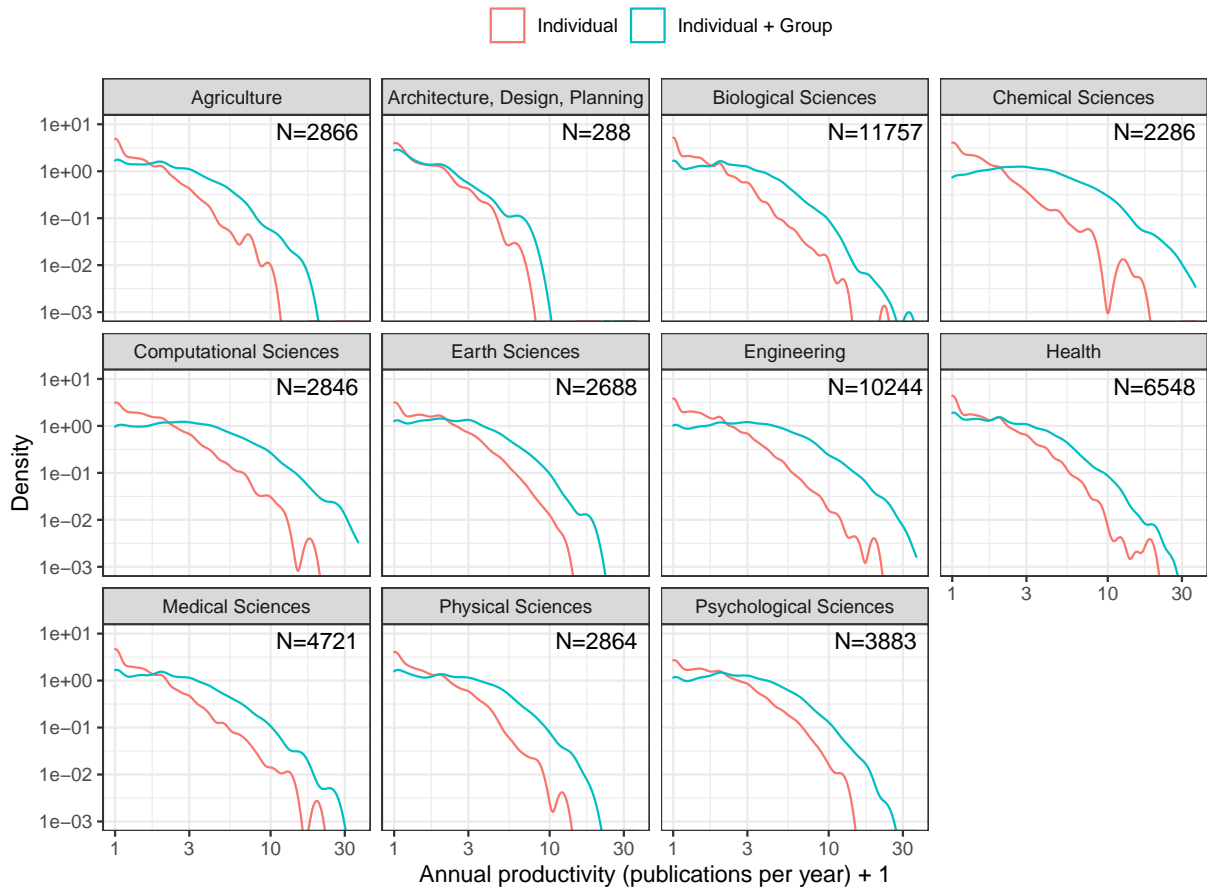


FIG. S10. **Distribution of individual vs. overall productivity by discipline.** For each discipline with research group collaboration norms, we plot the density of the individual and overall (individual + group) productivity on a log-log scale. Individual productivity drops off steeply, with very few faculty exhibiting greater than 10 individual papers per year. When faculty are highly productive, the bulk of their productivity consists of papers coauthored with group members.

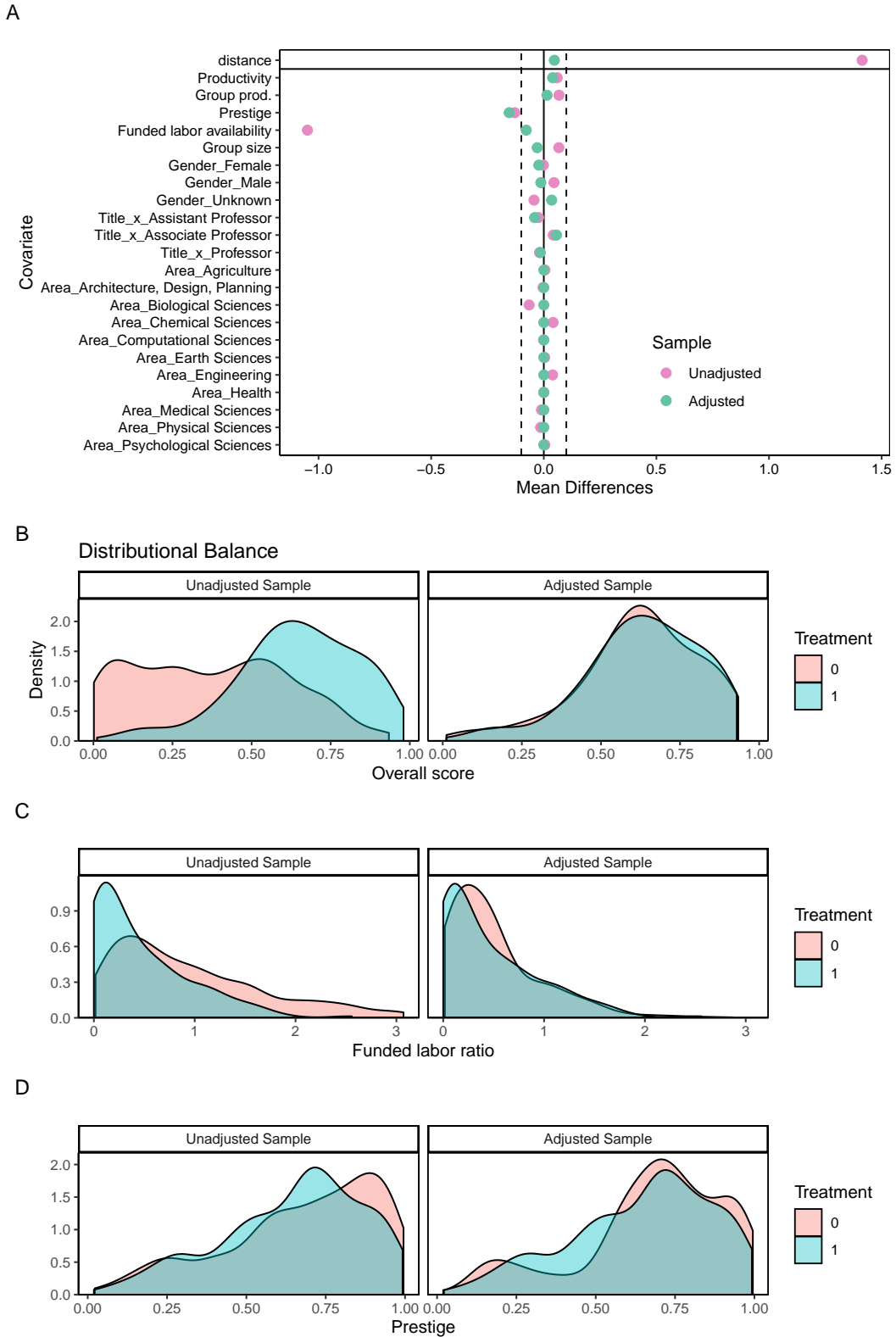


FIG. S11. **Balance of covariates before and after matching.** (A) Diagnostic test of covariate balance before and after matching, as well as overall propensity score difference between groups (“difference” in top row). Prior to matching, we had substantial imbalances in both covariates and overall distance (pink), that were largely corrected to an acceptable level of below 0.1 mean difference after adjustment (green). The exception is that prestige remained imbalanced after matching, which we diagnose further below. (B) Overall distributional balance of propensity scores is corrected by the matching. (C) The most imbalanced pre-adjustment variable, the logarithm of the ratio of funded labor to faculty in departments, is adequately balanced by the matching. (D) The only variable that remained imbalanced after adjustments was prestige. In particular, upward movers are more likely to come from less prestigious institutions, and downward movers are more likely to come from more prestigious institutions, due to the inability of those at the bottom of the prestige hierarchy to move downward and the inability of those at the top of the prestige hierarchy to move upward.

	(gs/c)	(gs/c)	(gs/c)	(gs/nc)	(gp/c)	(gp/nc)	(p/c)	(p/nc)
Is full prof.	0.52*** (0.02)	0.52*** (0.02)	0.52*** (0.02)	0.21*** (0.02)	0.35*** (0.02)	0.14*** (0.02)	0.39*** (0.02)	0.17*** (0.02)
Is assoc. prof.	0.24*** (0.02)	0.24*** (0.02)	0.24*** (0.02)	0.10*** (0.02)	0.14*** (0.02)	0.07*** (0.01)	0.12*** (0.02)	0.01 (0.02)
Years since degree	-0.22*** (0.01)	-0.22*** (0.01)	-0.22*** (0.01)	-0.08*** (0.01)	-0.17*** (0.01)	-0.05*** (0.01)	-0.19*** (0.01)	-0.11*** (0.01)
Log dept. size	0.02 (0.01)	-0.03* (0.01)	0.02 (0.01)	-0.00 (0.01)	0.01 (0.01)	-0.00 (0.01)	0.01 (0.01)	0.00 (0.01)
Unfunded labor availability	0.05** (0.02)	0.12*** (0.02)	0.05** (0.02)	0.01 (0.02)	0.05*** (0.01)	0.01 (0.01)	0.07*** (0.01)	0.04** (0.02)
Prestige		0.10*** (0.01)	0.00 (0.02)	-0.05*** (0.02)	0.01 (0.01)	-0.03** (0.01)	0.02 (0.01)	-0.00 (0.01)
Funded labor availability	0.22*** (0.02)		0.22*** (0.02)	0.09*** (0.02)	0.14*** (0.01)	0.04*** (0.01)	0.16*** (0.01)	0.06*** (0.02)
Is man	0.04* (0.02)	0.05** (0.02)	0.04* (0.02)	0.03 (0.01)	0.04*** (0.01)	0.02 (0.01)	0.06*** (0.01)	0.06*** (0.02)
# groups: Discipline	9	9	9	6	9	6	9	6
# groups: Dept.:Disc.	686	686	686	299	686	299	686	299
# obs.	10360	10360	10360	4262	10360	4262	10360	4262
% variance between Discipline	2.3%	4.5%	2.3%	1.8%	2.3%	3.8%	1%	0.82%
% variance between Dept. within Disc.	16%	16%	16%	14%	13%	12%	15%	9.9%

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S9. **Hierarchical linear regression on individual faculty using strict linkage.** As an alternate specification to the institution-discipline regressions, here we present the regression coefficients from a hierarchical linear regression on individual faculty using log outcomes with plus-one smoothing using the strict linkage between employment records and NSF survey data. We use two hierarchical levels: the discipline and the department, to quantify the proportion of variance in the model explained by discipline and department. The model names correspond to the dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms). We also include versions of the p/c model without the prestige and labor variables. Here we display the percent of the total variation due to between-discipline variance and between-department within-discipline variance, with the residual due to within-department faculty variance. All continuous variables are standardized to have zero mean and unit variance.

	(gs/c)	(gs/c)	(gs/c)	(gs/nc)	(gp/c)	(gp/nc)	(p/c)	(p/nc)
Is full prof.	0.58*** (0.01)	0.58*** (0.01)	0.58*** (0.01)	0.26*** (0.02)	0.37*** (0.01)	0.16*** (0.01)	0.43*** (0.01)	0.21*** (0.02)
Is assoc. prof.	0.27*** (0.01)	0.27*** (0.01)	0.27*** (0.01)	0.12*** (0.01)	0.15*** (0.01)	0.08*** (0.01)	0.14*** (0.01)	0.05** (0.01)
Years since degree	-0.22*** (0.00)	-0.23*** (0.00)	-0.22*** (0.00)	-0.09*** (0.01)	-0.17*** (0.00)	-0.05*** (0.00)	-0.20*** (0.00)	-0.13*** (0.01)
Log dept. size	0.04*** (0.01)	0.01 (0.01)	0.02** (0.01)	0.01 (0.03)	0.01* (0.00)	-0.00 (0.02)	0.02** (0.01)	0.00 (0.02)
Unfunded labor availability	-0.02* (0.01)	0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)
Prestige		0.10*** (0.01)	0.09*** (0.01)	0.02* (0.01)	0.07*** (0.00)	0.01 (0.01)	0.09*** (0.00)	0.05*** (0.01)
Funded labor availability	0.08*** (0.01)		0.04*** (0.01)	0.01 (0.01)	0.03*** (0.01)	0.00 (0.01)	0.03*** (0.01)	0.00 (0.01)
Is man	0.07*** (0.01)	0.08*** (0.01)	0.08*** (0.01)	0.03** (0.01)	0.06*** (0.01)	0.01 (0.01)	0.08*** (0.01)	0.06*** (0.01)
# groups: Discipline	11	11	11	6	11	6	11	6
# groups: Dept:Disc	3750	3750	3750	635	3750	635	3750	635
# obs.	50991	50991	50991	8731	50991	8731	50991	8731
% variance between Discipline	10%	12%	11%	4.1%	9.6%	5.7%	6.5%	2.3%
% variance between Dept. within Disc.	15%	15%	15%	18%	12%	12%	12%	7.4%

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

TABLE S10. **Hierarchical linear regression on individual faculty using non-strict linkage.** As an alternate specification to the institution-discipline regressions, here we present the regression coefficients from a hierarchical linear regression on individual faculty using log outcomes with plus-one smoothing. We use two hierarchical levels: the discipline and the department, to quantify the proportion of variance in the model explained by discipline and department. The model names correspond to the dependent variable (p=productivity, gp=group productivity, gs=group size) and the partition of the disciplines (c=disciplines with collaboration norms, nc=disciplines with no collaboration norms). We also include versions of the p/c model without the prestige and labor variables. Here we display the percent of the total variation due to between-discipline variance and between-department within-discipline variance, with the residual due to within-department faculty variance. All continuous variables are standardized to have zero mean and unit variance.

Outcome	Subset	Mean	SD	q_5	q_{25}	q_{50}	q_{75}	q_{95}
Group prod.	All	1.06	1.64	0.00	0.00	0.50	1.38	4.00
Group prod.	Collab norm	1.18	1.70	0.00	0.14	0.62	1.50	4.25
Group prod.	No collab norm	0.30	0.66	0.00	0.00	0.00	0.33	1.33
Group size	All	5.01	6.05	0.14	1.25	3.17	6.75	15.29
Group size	Collab norm	5.49	6.22	0.17	1.67	3.86	7.33	16.00
Group size	No collab norm	1.74	2.09	0.00	0.80	1.00	2.00	5.00
Indiv. prod.	All	0.73	1.06	0.00	0.00	0.40	1.00	2.60
Indiv. prod.	Collab norm	0.72	1.07	0.00	0.00	0.38	1.00	2.62
Indiv. prod.	No collab norm	0.80	1.01	0.00	0.14	0.50	1.00	2.56
Productivity	All	1.80	2.25	0.00	0.38	1.00	2.38	6.00
Productivity	Collab norm	1.90	2.33	0.00	0.43	1.17	2.56	6.12
Productivity	No collab norm	1.10	1.35	0.00	0.29	0.71	1.43	3.50

TABLE S11. **Descriptive statistics for outcome variables.** We show descriptive statistics across all disciplines, only disciplines with research group collaboration norms, or only disciplines without research group collaboration norms. Shown are means, standard deviations, and the 5th, 25th, 50th (median), 75th, and 95th percentiles of the person-average data for each subset. That is, values are first averaged over years to individuals, and then summary statistics are taken over the individual data within each subset. Group size is collected over a three year window.

Discipline	Productivity					Group prod.					Individual prod.					Group size				
	Mean	SD	q ₂₅	q ₅₀	q ₇₅	Mean	SD	q ₂₅	q ₅₀	q ₇₅	Mean	SD	q ₂₅	q ₅₀	q ₇₅	Mean	SD	q ₂₅	q ₅₀	q ₇₅
Engineering	2.60	3.02	0.62	1.71	8.00	1.79	2.33	0.29	1.00	6.16	0.82	1.17	0.11	0.44	2.89	6.21	6.17	2.00	4.50	17.89
Biological Sci.	1.62	1.82	0.40	1.00	5.00	0.99	1.24	0.14	0.62	3.25	0.63	0.94	0.00	0.33	2.25	5.69	5.70	2.00	4.14	16.00
Physical Sci.	1.59	1.82	0.29	1.00	5.00	0.94	1.31	0.00	0.50	3.28	0.65	0.91	0.00	0.33	2.44	6.35	12.05	1.67	3.67	18.14
Computational Sci.	2.50	3.05	0.56	1.57	8.12	1.66	2.31	0.20	1.00	6.00	0.84	1.21	0.11	0.44	3.00	6.10	6.34	2.00	4.38	17.00
Health	1.49	1.84	0.29	1.00	4.86	0.75	1.11	0.00	0.33	2.75	0.75	1.09	0.00	0.38	2.85	3.80	4.28	1.00	2.50	11.60
Agriculture	1.39	1.66	0.29	0.89	4.50	0.84	1.17	0.00	0.43	3.00	0.55	0.84	0.00	0.25	2.00	4.27	4.14	1.38	3.00	12.00
Chemical Sci.	2.64	3.06	0.67	1.75	8.31	2.03	2.50	0.38	1.22	6.66	0.61	1.03	0.00	0.33	2.25	8.38	7.74	3.00	6.31	22.49
Psychological Sci.	1.92	2.07	0.50	1.29	5.88	0.98	1.24	0.12	0.57	3.33	0.95	1.20	0.14	0.57	3.20	4.39	4.03	1.50	3.33	11.99
Medical Sci.	1.62	2.04	0.33	1.00	5.40	0.99	1.32	0.00	0.50	3.50	0.64	1.08	0.00	0.29	2.43	6.31	6.51	2.00	4.50	18.60
Earth Sci.	1.70	1.86	0.43	1.14	5.29	0.87	1.16	0.00	0.50	3.00	0.84	1.07	0.00	0.50	2.86	3.91	3.54	1.40	3.00	10.48
Arch. & Planning	0.71	1.00	0.00	0.33	2.74	0.22	0.62	0.00	0.00	1.00	0.49	0.72	0.00	0.17	2.00	1.64	2.15	0.50	1.00	5.00
Political Science	0.89	1.00	0.29	0.62	2.82	0.14	0.36	0.00	0.00	0.62	0.75	0.91	0.14	0.50	2.38	1.20	1.17	0.71	1.00	2.80
Sociology	1.12	1.28	0.33	0.80	3.38	0.32	0.70	0.00	0.12	1.33	0.80	0.98	0.14	0.56	2.49	1.77	1.86	0.88	1.20	5.00
Mathematical Sci.	1.39	1.63	0.33	1.00	4.50	0.45	0.82	0.00	0.14	1.88	0.94	1.17	0.17	0.57	3.00	2.22	2.67	1.00	1.38	7.00
Economics	0.85	0.95	0.25	0.62	2.55	0.21	0.48	0.00	0.00	0.99	0.63	0.71	0.14	0.44	2.00	1.46	1.71	0.75	1.00	3.60
Geography	1.22	2.12	0.14	0.69	3.80	0.48	0.91	0.00	0.14	1.87	0.74	1.53	0.00	0.18	2.56	2.38	2.81	0.86	1.50	7.95
Anthropology	0.91	1.07	0.29	0.62	2.96	0.20	0.50	0.00	0.00	1.00	0.72	0.85	0.17	0.50	2.29	1.43	1.46	0.71	1.00	3.83

TABLE S12. **Descriptive statistics for outcome variables by discipline.** Shown are means, standard deviations, and the 25th, 50th (median), and 75th percentiles of the person-average data for each discipline. That is, values are first averaged over years to individuals, and then summary statistics are taken over the individual data within each discipline. Group size is collected over a three year window.

Discipline	# Insts.	# tenure-track (TT)					Funded per TT					Unfunded per TT				
		Mean	SD	q ₂₅	q ₅₀	q ₇₅	Mean	SD	q ₂₅	q ₅₀	q ₇₅	Mean	SD	q ₂₅	q ₅₀	q ₇₅
Engineering	159	79.83	53.17	42	68	188	1.48	2.40	0.29	0.69	5.92	2.25	5.33	0.44	1.01	6.19
Biological Sci.	189	68.25	57.32	25	51	189	0.99	2.32	0.14	0.38	3.90	0.78	1.05	0.13	0.36	2.71
Physical Sci.	125	30.16	20.19	15	24	67	2.55	3.19	0.47	1.64	7.37	1.64	1.73	0.61	0.99	5.70
Computational Sci.	124	34.27	22.36	18	28	83	2.36	4.21	0.39	0.93	8.66	5.39	10.42	0.90	1.94	19.33
Health	147	66.63	55.82	24	54	179	0.53	1.29	0.03	0.13	2.79	1.20	1.93	0.20	0.51	4.55
Agriculture	50	71.20	59.76	24	57	174	2.43	2.14	0.85	1.78	6.75	1.11	1.20	0.33	0.73	3.59
Chemical Sci.	117	39.56	26.69	17	34	83	2.32	3.09	0.63	1.17	8.60	1.80	1.82	0.73	1.25	5.00
Psychological Sci.	150	29.21	17.68	17	25	65	1.09	1.56	0.20	0.58	3.63	1.44	1.68	0.46	0.89	4.57
Medical Sci.	102	65.34	46.93	31	56	160	0.99	1.97	0.11	0.30	4.04	0.47	1.22	0.00	0.06	2.18
Earth Sci.	98	36.01	34.15	12	26	115	1.53	3.41	0.22	0.60	4.19	1.13	1.99	0.23	0.67	2.93
Arch. & Planning	32	40.00	23.01	25	32	82	0.31	0.52	0.01	0.10	1.33	2.43	5.82	0.33	0.72	5.69
Political Science	92	24.15	13.73	15	21	52	2.04	4.65	0.16	0.62	8.54	4.51	7.74	1.08	2.47	13.07
Sociology	83	23.55	15.10	14	21	45	1.09	1.73	0.13	0.34	4.05	1.97	2.70	0.52	1.17	5.28
Mathematical Sci.	131	30.43	21.67	14	24	64	1.10	1.92	0.10	0.44	4.37	2.97	4.03	0.81	1.69	10.33
Economics	98	24.33	21.74	11	17	59	1.60	3.74	0.13	0.41	7.30	4.40	7.71	1.08	1.85	13.04
Geography	38	17.68	11.49	10	14	38	1.12	1.50	0.31	0.72	3.37	2.50	3.82	0.86	1.43	5.21
Anthropology	65	21.46	14.31	11	18	49	1.11	1.60	0.19	0.53	3.50	2.25	1.90	0.87	1.58	6.02

TABLE S13. **Descriptive statistics for departmental variables by discipline.** Shown are means, standard deviations, and the 25th, 50th (median), and 75th percentiles of the data, as well as the number of institutions represented for each discipline. For institutions with multiple departments in the same discipline, we report the average values of the departments in that institution-discipline pairing, rounded for the number of tenure-track faculty column. The “Funded per TT” refers to the ratio of funded graduate and postdoctoral researchers in the department to tenure-track faculty, and “Unfunded per TT” refers to the ratio of unfunded graduate students in the department to tenure-track faculty.

			Mean absolute error (MAE)			
norms	agg.	DV	baseline	prestige	labor	both
collab	dept	prod.	0.80	0.76	0.75	0.74
collab	dept	gp.	0.57	0.54	0.53	0.53
collab	dept	gs.	2.08	1.97	1.93	1.91
collab	indiv.	prod.	1.52	1.50	1.49	1.49
collab	indiv.	gp.	1.10	1.09	1.08	1.08
collab	indiv.	gs.	4.16	4.12	4.08	4.08
non-collab	dept	prod.	0.40	0.40	0.40	0.40
non-collab	dept	gp.	0.17	0.18	0.17	0.17
non-collab	dept	gs.	0.56	0.57	0.56	0.56
non-collab	indiv.	prod.	0.77	0.77	0.77	0.77
non-collab	indiv.	gp.	0.31	0.31	0.31	0.31
non-collab	indiv.	gs.	0.97	0.97	0.97	0.96

TABLE S14. **Comparison of predictive performance of models** Using 10-fold cross-validation, we compare the mean average error (MAE) of the predicted dependent variable of the discipline-clustered Poisson regressions on the strict linkage data including either prestige, funded labor availability, neither, or both. We repeat this process for disciplines with vs. without collaboration norms, departmental vs. individual regression, and the different dependent variables productivity (prod), group productivity (gp), and group size (gs).

REFERENCES AND NOTES

1. J. S. Long, Productivity and academic position in the scientific career. *Am. Sociol. Rev.* **43**, 889 (1978).
2. S. F. Way, A. C. Morgan, D. B. Larremore, A. Clauset, Productivity, prominence, and the effects of academic environment. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10729–10733 (2019).
3. S. Cole, J. R. Cole, Visibility and the structural bases of awareness of scientific research. *Am. Sociol. Rev.* **33**, 397–413 (1968).
4. I. Aaltojärvi, I. Arminen, O. Auranen, H.-M. Pasanen, Scientific productivity, web visibility and citation patterns in sixteen Nordic sociology departments. *Acta Sociol.* **51**, 5–22 (2008).
5. A. C. Morgan, D. J. Economou, S. F. Way, A. Clauset, Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Sci.* **7**, 40 (2018).
6. N. Kaplan, J. R. Cole, S. Cole, Social stratification in science. *Technol. Cult.* **15**, 662 (1974).
7. A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Sci. Adv.* **1**, e1400005 (2015).
8. R. K. Merton, The Matthew effect in science. *Science* **159**, 56–63 (1968).
9. H. Dundar, D. R. Lewis, Determinants of research productivity in higher education. *Res. High. Educ.* **39**, 607–631 (1998).
10. P. E. Graves, J. R. Marchand, R. Thompson, Economics departmental rankings: Research incentives, constraints, and efficiency. *Am. Econ. Rev.* **72**, 1131–1141 (1982).
11. R. G. Ehrenberg, Studying ourselves: The academic labor market: Presidential address to the Society of Labor Economists, Baltimore, May 3, 2002. *J. Labor Econ.* **21**, 267–287 (2003).
12. P. Stephan, *How Economics Shapes Science* (Harvard Univ. Press, 2012).

13. B. Bozeman, E. Corley, Scientists' collaboration strategies: Implications for scientific and technical human capital. *Res. Policy* **33**, 599–616 (2004).
14. D. Beaver, Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics* **52**, 365–377 (2001).
15. S. Lee, B. Bozeman, The impact of research collaboration on scientific productivity. *Soc. Stud. Sci.* **35**, 673–702 (2005).
16. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *On Being A Scientist: A Guide to Responsible Conduct in Research: Third Edition* (The National Academies Press, 2009).
17. B. Alberts, Promoting scientific standards, *Science* **327**, 12 (2010).
18. M. A. Maher, B. C. Timmerman, D. F. Feldon, D. Strickland, Factors affecting the occurrence of faculty-doctoral student coauthorship. *J. High. Educ.* **84**, 121–143 (2013).
19. B. Kamler, Rethinking doctoral publication practices: Writing from and beyond the thesis. *Stud. High. Educ.* **33**, 283–294 (2008).
20. V. Larivière, On the shoulders of students? The contribution of PhD students to the advancement of knowledge. *Scientometrics* **90**, 463–481 (2012).
21. W. Shrum, J. Genuth, I. Chompalov, *Structures of Scientific Collaboration* (The MIT Press, 2007).
22. W. O. Hagstrom, *The Scientific Community* (Southern Illinois Univ. Press, 1965).
23. L. L. Hargens, *Patterns of Scientific Research* (American Sociological Association, 1975).
24. S. Kyvik, J.-C. Smeby, Teaching and research. The relationship between the supervision of graduate students and faculty research performance. *High Educ.* **28**, 227–239 (1994).

25. A. Ebadi, A. Schiffauerova, How to boost scientific production? A statistical analysis of research funding and other influencing factors. *Scientometrics* **106**, 1093–1116 (2016).
26. D. Shaw, L. Vaughan, Publication and citation patterns among LIS faculty: Profiling a “typical professor”. *Libr. Inf. Sci. Res.* **30**, 47–55 (2008).
27. D. R. Baker, M. V. K. Wilson, An evaluation of the scholarly productivity of doctoral graduates. *J. Soc. Work Educ.* **28**, 204–213 (1992).
28. National Center for Science and Engineering Statistics (NCSES), National Science Foundation survey of graduate students and postdoctorates in science and engineering (2008–2017); www.nsf.gov/statistics/srvygradpostdoc/ [accessed 10 December 2019].
29. U.S. Department of Education, College scorecard; <https://collegescorecard.ed.gov> [accessed 10 December 2019].
30. S. F. Way, A. C. Morgan, A. Clauset, D. B. Larremore, The misleading narrative of the canonical faculty productivity trajectory. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9216–E9223 (2017).
31. K. H. Wapman, S. Zhang, A. Clauset, D. B. Larremore, Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature* **610**, 120–127 (2022).
32. J. Surowiecki, *The Wisdom of Crowds* (Anchor, 2005).
33. S.-K. Han, Tribal regimes in academia: A comparative analysis of market structure across disciplines. *Soc. Networks* **25**, 251–280 (2003).
34. V. Burris, The academic caste system: Prestige hierarchies in PhD exchange networks. *Am. Sociol. Rev.* **69**, 239–264 (2004).
35. C. De Bacco, D. B. Larremore, C. Moore, A physical model for efficient ranking in networks. *Sci. Adv.* **4**, eaar8260 (2018).

36. W. Shockley, On the statistics of individual variations of productivity in research laboratories. *Proc. IRE* **45**, 279–290 (1957).
37. R. Sattari, J. Bae, E. Berkes, B. A. Weinberg, The ripple effects of funding on researchers and output. *Sci. Adv.* **8**, eabb7348 (2022).
38. R. L. Geiger, *Research and Relevant Knowledge: American Research Universities Since World War II* (Oxford Univ. Press, 1993).
39. J. R. Cole, *The Great American University: Its Rise to Preeminence, Its Indispensable National Role, Why It Must be Protected* (Public Affairs, 2010).
40. Y. Xie, “Undemocracy”: Inequalities in science. *Science* **344**, 809–810 (2014).
41. W. B. Rouse, J. V. Lombardi, D. D. Craig, Modeling research universities: Predicting probable futures of public vs. private and large vs. small research universities. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12582–12589 (2018).
42. B. Hofstra, V. V. Kulkarni, S. Munoz-Najar Galvez, B. He, D. Jurafsky, D. A. McFarland, The diversity-innovation paradox in science. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9284–9291 (2020).
43. S. Page, *The Difference* (Princeton Univ. Press, 2008).
44. D. Kozlowski, V. Larivière, C. R. Sugimoto, T. Monroe-White, Intersectional inequalities in science. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113067119 (2022).
45. N. Laberge, K. H. Wapman, A. C. Morgan, S. Zhang, D. B. Larremore, A. Clauset, Subfield prestige and gender inequality in computing. arXiv:2201.00254 [cs.CY] (2022); <https://arxiv.org/abs/2201.00254>.
46. L. Wu, D. Wang, J. A. Evans, Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
47. F. Elmas, gender-guesser (2016); <https://pypi.org/project/gender-guesser>.

48. V. I. Torvik, S. Agarwal, Ethnea: An instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database (2016); <https://hdl.handle.net/2142/88927>.
49. C. R. Sugimoto, S. Weingart, The kaleidoscope of disciplinarity. *J. Doc.* **71**, 775–794 (2015).
50. J. Fry, Scholarly research and information practices: A domain analytic approach. *Inf. Process. Manag.* **42**, 299–316 (2006).
51. N. Carayol, T. U. N. Thi, Why do academic scientists engage in interdisciplinary research? *Res. Eval.* **14**, 70–79 (2005).