

## Research



**Cite this article:** Petersen AM. 2018  
Multiscale impact of researcher mobility.  
*J. R. Soc. Interface* **15**: 20180580.  
<http://dx.doi.org/10.1098/rsif.2018.0580>

Received: 29 July 2018

Accepted: 3 September 2018

### Subject Category:

Life Sciences—Mathematics interface

### Subject Areas:

biogeography, computational biology,  
biocomplexity

### Keywords:

human mobility, propensity score matching,  
network science, science of science,  
researcher careers, diversity

### Author for correspondence:

Alexander M. Petersen  
e-mail: [petersen.xander@gmail.com](mailto:petersen.xander@gmail.com)

Electronic supplementary material is available  
online at <https://dx.doi.org/10.6084/m9.figshare.c.4226732>.

Alexander M. Petersen

Ernest and Julio Gallo Management Program, Management of Complex Systems Department,  
School of Engineering, University of California, Merced, CA 95343, USA

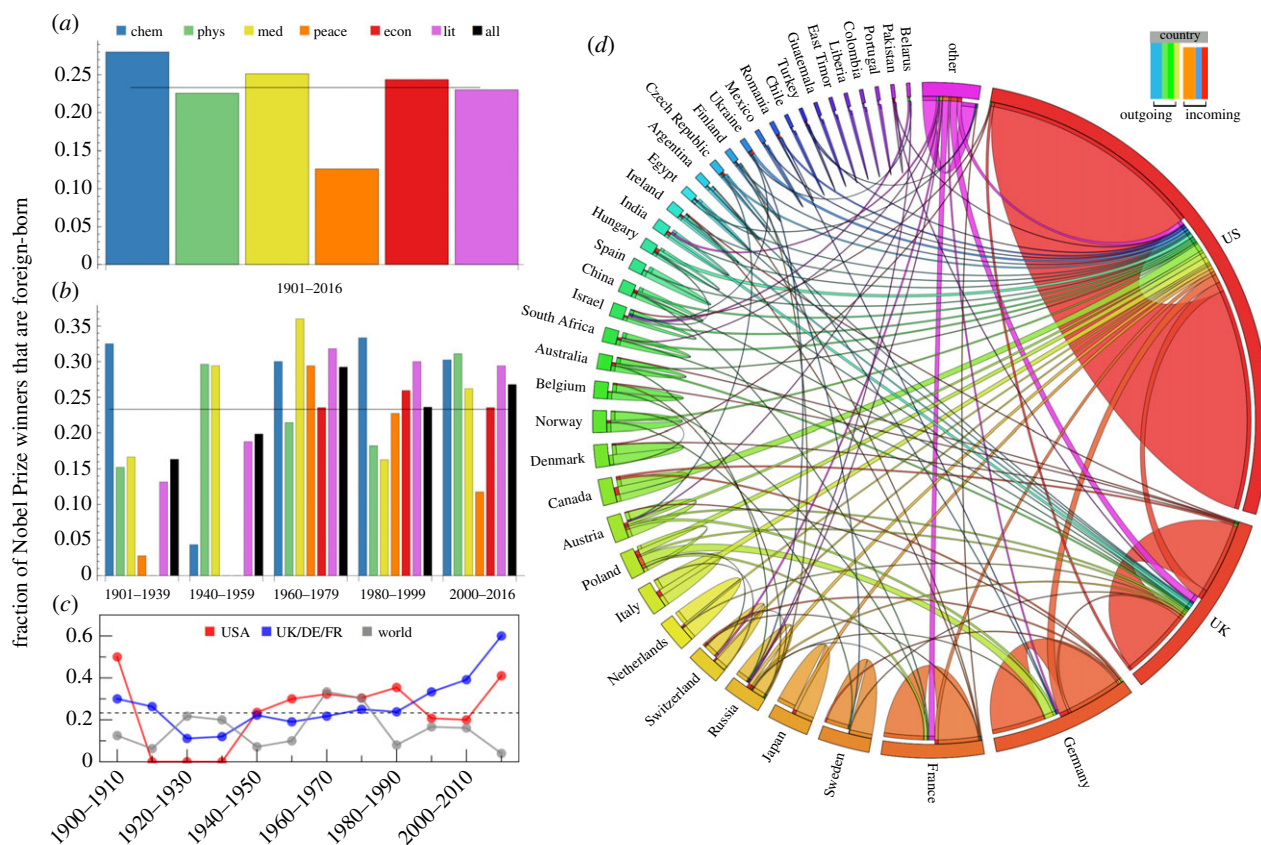
AMP, 0000-0002-0955-3483

International mobility facilitates the exchange of scientific, institutional and cultural knowledge. Yet whether globalization and advances in virtual communication technologies have altered the impact of researcher mobility is a relevant and open question that we address by analysing a broad international set of 26 170 physicists from 1980 to 2009, focusing on the 10-year period centred around each mobility event to assess the impact of mobility on research outcomes. We account for secular globalization trends by splitting the analysis into three periods, measuring for each period the effect of mobility on researchers' citation impact, research topic diversity, collaboration networks and geographical coordination. In order to identify causal effects we leverage statistical matching methods that pair mobile researchers with non-mobile researchers that are similar in research profile attributes prior the mobility event. We find that mobile researchers gain up to a 17% increase in citations relative to their non-mobile counterparts, which can be explained by the simultaneous increase in their diversity of co-authors, topics and geographical coordination in the period immediately following migration. Nevertheless, we also observe that researcher's completely curtail prior collaborations with their source country in 11% of the cross-border mobility events. As such, these individual-level perturbations fuel multiscale churning in scientific networks, e.g. rewiring the connectivity of individuals and ideas and affecting international integration. Together these results provide additional clarity on the complex relationship between human capital mobility and the dynamics of social capital investment, with implications for immigration and national innovation system policy.

## 1. Introduction

The dispersion of knowledge across institutional and national borders is fundamental to scientific progress. Historically, knowledge exchange has been mediated via the physical mobility of people and printed publications, however, it is progressively mediated by a combination of physical and virtual networks. While researcher relocation remains a prominent conduit for knowledge transfer, researchers can now choose from a variety of virtual alternatives to explore new research environments and collaborations. And while the professional prospects and call of adventure associated with relocation may be alluring to some, there are nevertheless risky trade-offs associated with physical relocation that require careful assessment of local versus non-local socio-economic, family, work and funding opportunities [1–4]. This common dilemma factors into the cost of human and social capital investment in science, which is rather substantial and continues to grow with the globalization of the scientific endeavour [1,5].

To stay or leave? We provide guidance on this pivotal question by quantifying the impact of researcher mobility from four perspectives—scientific impact, research topic diversity, collaboration and international integration—together providing insights into the socially mediated network of networks connecting individuals and knowledge production [6]. As such, our results contribute to the literature on how knowledge flows [7,8] and how careers grow [9] following pivotal events—e.g. winning the Nobel Prize [10], initiating a dedicated



**Figure 1.** International mobility of Nobel laureates. (a) Fraction of Nobel Prize winners who received their award for work done in a country different than their birth country—‘foreign-born’; 23% of Nobel Prize winners from 1901 to 2016 belong to the foreign-born category, indicated by the horizontal line. (b) Fraction by award and time period. (c) Fraction by region of achievement over 10-year intervals. (d) The Nobelist birth-discovery network: countries are represented along the arc, with arc-length proportional to the number of laureates born in a given country. Link width represents the number of individuals born in country  $c_b$  that performed award research in country  $c_a$ . The link direction is denoted by the gap, which differentiates incoming and outgoing links: e.g. the majority of the international links associated with the US terminate with a gap, indicating foreign-born. The majority of links are intra-country (e.g. 29% of all laureates have the US as their country of birth and their country of achievement), but the international links represent the nearly 1 in 4 Nobelists who migrated internationally. Data from *Wolfram Alpha* [35]. (Online version in colour.)

partnership [11], obtaining tenure [12] and pursuing cross-disciplinary migration [13]—which in turn mediate shifts in collaboration [14–16] and research topic exploration [17–20].

In particular, we focus on the domain of physics, a broad research field characterized by relatively high levels of geographical mobility, which contributes to growing efforts to measure the impact of high-skilled migration in both industry and academia [1–5,7,8,14–16,21–33]. We employ a data-driven approach, starting with author disambiguation of 355 808 research articles published in American Physical Society (APS) journals over the period 1980–2009, resulting in a final dataset of 26 170 prominent researcher profiles that are broadly distributed both in terms of geography, productivity and scientific impact.

We leverage the size and breadth of this researcher dataset to address common reverse causality and selection bias problems that render identifying causal mobility effects rather challenging [34]. For example, many studies to date focus on select subsets of migrant researchers—e.g. elite scientists, winners of prestigious international fellowships or participants in select national exchange programmes—in which measured shifts in performance may endogenously depend on the selection of migrant researchers being analysed. Instead, following a data-driven approach we are able to estimate the mobility effect within a diverse population, accounting for variation in time period, geographical region, and social and cognitive aspects of research activity.

Our results demonstrate the impact of researcher mobility across scales—from individual careers, to collaboration network churning, to international integration. As motivation, the Nobel Prize offers a prime example of multiscale impact, highlighting how countries with attractive academic and industrial R&D environments that support objective and impartial immigration policy are positioned to benefit greatly from the immigration of even just a single elite scientist capable of prosecuting ground-breaking and transformative research.

To provide insights into the global immigration patterns of elite scientists, we analyse the international pathways of all Nobel Prize winners through 2016 (figure 1). We found that roughly 1 in 4 are foreign-born—having performed their prize-winning research in a country different than their birth country. This finding is rather stable across the different award categories, with the exception of the Peace Prize, which has lower frequencies since there is a propensity for human rights activists to lead movements in their native countries. The UK, Germany and France, the three countries in Europe with the highest R&D spending, have seen a recent surge since the 1990s in the fraction of Nobelists from their country who are foreign-born. The rest of the world, not including the US, shows the opposite trend since the 1960s. There is also significant disparity in the flow counts from source to destination country, as figure 1d illustrates how research hubs like the USA, the UK and Germany attract a disproportionate share of foreign-born Nobelists.

While winning the Nobel Prize is likely to remain out of reach for most scientists, the bulk of scientists will nevertheless encounter an opportunity and be faced with the decision of whether to relocate. Based on our data, we estimate the rate of international mobility among all researchers in our dataset, independent of productivity level, to be roughly 2 in 5 (40%); and including US inter-state cross-border mobility this rate is closer to 3 in 5 (58%).

In what follows, we focus on several thousand relatively productive physicists across three decades, implementing statistical matching techniques that leverage the breadth of our dataset to compare researchers who migrate to similar researchers who do not. As such, we demonstrate the value of cross-border mobility on research outcomes, reflecting the pervasive and persistent value of physical mobility as a conduit for the exchange of intellectual, professional and cultural experience.

## 2. Results

### 2.1. Researcher data

We analysed the APS *Physical Review* datasets ‘Citing article pairs’ and ‘Article metadata’ [36], from which we extracted researcher profiles using a network-based author name disambiguation method [37]; see electronic supplementary material, appendix A.1 for additional details. Application of this algorithm identified 208 734 distinct researcher profiles over the 30-year period 1980–2009. Each researcher profile, indexed by  $i$ , is characterized by the year of researcher’s first APS publication,  $y_i^0$ , and total number of APS publications  $N_i$ . We concentrate on the 26 170 researchers with  $N_i \geq 10$  publications who also meet additional career longevity and productivity criteria. Electronic supplementary material, figure S1(A–C) provides a statistical summary of the publication and researcher profile data.

Despite the large sample size, there are several limitations common to large-scale data-driven approaches, which derive from missing contextual information at the researcher and geographical level. First, we lack data for physics and non-physics research published by these researchers outside of the set of APS journals. Yet because of the reputation and prominence of the APS journal family, we are able to make the assumption that researchers characteristic of physics-related domains are likely to publish in one or more of these journals rather frequently over the course of his/her career. Thus, as long as their publications in the APS journals are characteristic of their contemporaneous research output in other journals (albeit excluding the exceptionally novel and infrequent research published in high-impact multidisciplinary letters journals) then we do not expect this to significantly bias our estimates. In other words, our reported estimates measure the impact of mobility on typical research outcomes.

Second, and relevant to this previous point, in order to facilitate the precision of the statistical matching method used here that leverages the large size of our researcher sample, we only include relatively prolific APS researchers. This is achieved by thresholding on publication rates during the 10-year window around each mobility event, which also ensures that the publication measures are aggregated from sufficient sample size to be a reasonably robust measure of researcher attributes in the immediate (5-year) period before and after the event. However, this also means that we likely exclude many early stage (doctoral) and postdoctoral

researchers who commonly migrate for brief periods, but do not have significant publication output to enter our study before they exit the academic career path. For example, according to US National Science Foundation Statistics, by 2008 the rate of US PhD recipients (postdoctoral fellows) in science and engineering who were foreign-born had grown to roughly 1 in 2 (3 in 5) [5]. Thus, our results do not capture the mobility premium for this highly mobile portion of the academic workforce, which deserves additional attention given shifts in the early and mid-career labour market [12,38–41] and the uncertainty associated with academic career trajectories [24,29,42–44].

And third, in addition to missing bibliometric information, we also lack other important information such as gender, which is an important factor relevant to career longevity and productivity in science [42,45–48], and possibly also migration decisions. In a broader sense, we lack contextual information on what *push and pull* factors underly researchers’ decisions to migrate. For example, a 2011 survey analysis based on more than 15 000 researchers’ responses reports that destination countries with competitiveness-oriented national science policy (i.e. oriented around ‘prestige’ and ‘research excellence’) is a major attractive factor [31]. A different study based on more than 10 000 biomedical researchers, which incorporates social and familial factors in addition to the traditional professional factors, also shows that researchers are attracted by the prospects associated with a competitive professional and peer environment; however, more accomplished researchers are less likely to move if they recently obtained NIH funding or if their children are in high school [3]. Countries with relatively high R&D expenditure levels tend to have higher rates of elite physicist immigration [2]. And researcher mobility between two countries is positively correlated with international student mobility in the opposite direction, indicative of the formation of cross-generational brain circulation channels [49]. Complementing these findings is an analysis of high-skilled migration within Europe over the period 1997–2014, showing that countries with higher government expenditure on education also make significantly more attractive destination countries [4].

### 2.2. Researcher mobility framework

The aim of our study is to measure shifts in researcher profiles before versus after a cross-border mobility event occurring in year  $t_{i,T}^*$ . Electronic supplementary material, figure S2 illustrates our framework for measuring the reconfiguration of research attributes before and after such a mobility event; see electronic supplementary material, A.1 for further details on the calculation of  $t_{i,T}^*$ . In particular, our framework relies on the following subscript notation: we split into three time periods (indexed by  $T$ ), drawing on numerous researcher profiles (indexed by  $i$ ), which are then split into three mobility groups (indexed by  $G$ ). Moreover, in order to assess individual researcher profile attributes, we also analysed the collaborators, keywords, and countries associated with each researcher profile, hereafter generically indexed by  $j$ .

For two main reasons, we split the analysis into three mobility observation periods denoted by  $T$ , each defined by non-overlapping lower and upper year limits,  $t_T^-$  and  $t_T^+$ :  $T_1 \equiv [1990–1997]$ ,  $T_2 \equiv [1998–2003]$  and  $T_3 \equiv [2004–2007]$ . First, this separation facilitates drawing period-specific

conclusions; considered together, they facilitate identifying trends in the mobility effect over time. Second, this choice provides a compromise on the issue of how to treat multiple mobility events for a given researcher. In principle, a researcher could migrate multiple times over a short period (e.g. characteristic of a short postdoctoral or sabbatical period). However, we are primarily interested in the shifts in research activity associated with the first mobility event within a reasonable time frame. Thus, for a given analysis period  $T$ , we use just the first  $t_{i,T}^*$  associated with a sequence of mobility events. By partitioning our analysis into multiple  $T$ , we then allow for subsequent mobility events by the same  $i$  to also contribute to our analysis. Electronic supplementary material, figure S1(D) provides a schematic of the separation of the analysis into three observation periods:  $R$  is the number of researcher profiles analysed per  $T$ :  $R_1 = 4124$ ,  $R_2 = 9362$  and  $R_3 = 13457$ . The variable window size accounts for the growth of the profession, and the latter period is chosen to occur after 2004 so to account for the increased international mobility following the EU enlargement [4,16].

Electronic supplementary material, figure S1(E) provides a schematic description of our procedure for classifying each researcher  $i$  that was active in a given period  $T$  according to three groups: (i)  $G_{i,T} = 1$  identifies researchers who did not migrate before or during period  $T$ ;  $G_{i,T} = 2$  identifies a ‘placebo’ mobility group comprised of researchers who were mobile prior to  $T$  but not during  $T$ ; and (iii)  $G_{i,T} = 3$  identifies researchers who were mobile during  $T$ , with specific mobility year  $t_{i,T}^*$ . Researchers in a given  $T$  belonging to either group  $G_{i,T} = 1$  or 2 are prescribed  $t_{i,T}^* \equiv \text{Median}[t_T^-, t_T^+]$ , the midpoint of  $T$  (e.g.  $t_{i,1}^* \equiv 1994$ ). We then aggregated each researcher’s publications in the  $\Delta t \equiv 5$ -year window before  $t_{i,T}^*$ , i.e. over the interval  $[t_{i,T}^* - \Delta t, t_{i,T}^* - 1]$ ; likewise, we aggregated the publications after  $t_{i,T}^*$  over the interval  $[t_{i,T}^*, t_{i,T}^* + \Delta t - 1]$ . This framework facilitates measuring research patterns over a balanced observation period for each  $i$ , thereby leveraging the longitudinal dimension of the researcher profiles. As a final dataset refinement, we excluded researcher profiles with fewer than three publications in either the period before or after  $t_{i,T}^*$  and fewer than four distinct years of publication activity in total.

Electronic supplementary material, figure S1(F) shows the annual distribution of the total 31 075 mobility events, including in this count the multiple movements by a single researcher within a single  $T$ . Refining to just the first mobility event per researcher per period, we observe 6498 profiles belonging to the  $G_{i,T} = 3$  group (i.e. roughly 21% of the total mobility events). These select researcher profiles are rich in data and sufficient in number to implement a matching method approach to estimate the impact of researcher mobility on research outcomes by comparing groups 1 and 3; moreover, we perform a robustness check by comparing researchers in groups 1 and 2 and researchers in groups 2 and 3. Intuitively, we expect that the mobility effect for the groups 1 and 3 comparison should be larger than the other group comparison estimates if our specification is to be considered consistent.

### 2.3. Research activity measures

We provide four complementary perspectives on (i) scientific impact, (ii) collaboration, (iii) research topics and (iv) geographical coordination, defined as follows:

#### 2.3.1. Citation impact

We normalized the standard integer citation count  $n_{i,p,t}$  for each publication  $p$  published in year  $t$  to account for temporal bias. The result is a normalized z-score  $z_{i,p}$ , which is normally distributed for all  $t$ . We then calculate the average citation impact value  $Z_i^{+-} \equiv \langle z_{p,i}^{+-} \rangle$  across the  $N_i^{+-}$  publications in each  $\Delta t \equiv 5$ -year interval, i.e. before ( $-$ ) or after ( $+$ )  $t_{i,T}^*$ . We also use a measure of total citation impact,  $\Sigma_i^{+-}$ , useful for assessing the magnitude of the mobility effect in real terms by applying a ‘citation deflator’ that accounts for ‘citation inflation’ [50].

#### 2.3.2. Co-author diversity

For each  $\Delta t \equiv 5$ -year interval we count the number of articles published by  $i$  with co-author  $j$ , given by  $k_{ij}$ . We then calculate the Shannon entropy for the distribution of  $k_{ij}$  across the set of  $K_i$  co-authors, defined as  $E_{K_i}^{+-}$ . Higher levels of variation correspond to larger entropy values, with  $E \geq 0$ ; the limiting case of no variation,  $k_{ij} = \text{const.}$  for all  $j$ , corresponds to  $E = 0$ .

#### 2.3.3. Research topic diversity

Similar to (§2.3.2), we aggregate the instances of *Physics and Astronomy Classification Scheme* (PACS) codes associated with each publication to calculate a Shannon entropy measuring the variation in research topics,  $E_{Q_i}^{+-}$ .

#### 2.3.4. Geographic reach diversity

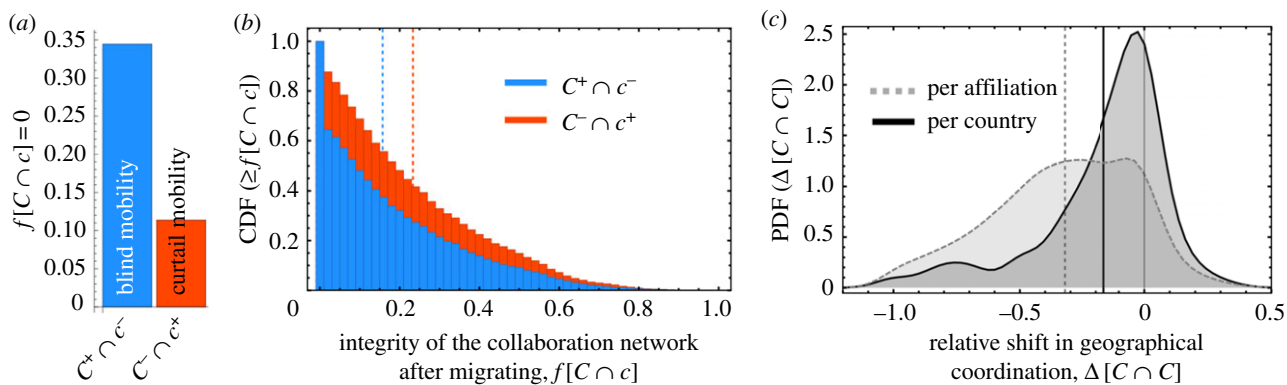
We aggregate the countries indicated in the affiliation bylines and again calculate a Shannon entropy measuring the variation in geographical coordination,  $E_{C_i}^{+-}$ .

See the Material and methods for further details on each measure. In what follows, we use the quantities  $Z_i^+$ ,  $\Sigma_i^+$ ,  $E_{K,i}^+$ ,  $E_{Q,i}^+$  and  $E_{C,i}^+$  as dependent variables in five separate models, and the corresponding  $Z_i^-$ ,  $\Sigma_i^-$ ,  $E_{K,i}^-$ ,  $E_{Q,i}^-$  and  $E_{C,i}^-$  along with five other covariate data measured before  $t_{i,T}^*$  in order to match  $i$  between  $G_{i,T}$  groups. Electronic supplementary material, figures S3–S4 show the distribution of the quantities  $Z_i^{+-}$ ,  $E_{K,i}^{+-}$ ,  $E_{Q,i}^{+-}$  and  $E_{C,i}^{+-}$ ; see electronic supplementary material, A.2 and figure S5 for the analysis of the difference in these variables around  $t_{i,T}^*$ .

### 2.4. Reorganization of social and geographical links

When a researcher relocates, there is an immediate impact on his/her proximity to former collaborators. This reorganization of collaboration networks at the individual level can have far-reaching implications at the national and global scale. A case example is how East–West migration imbalance within Europe following the 2004 enlargement of the European Union (EU) [16] negatively impacted rates of international collaboration; it is not unlikely that high-skilled migration in response to Brexit will have a similar effect on international collaboration patterns in Europe [4].

Against this background, in this section we investigate the extent to which mobility mediates collaboration-based shifts in social and geographical integration, providing insights into the formation and disintegration of social capital. As above,  $C^-$  ( $C^+$ ) denotes the list of countries extracted from the publication affiliation byline (dropping the subscript  $i$  for brevity), with list length denoted by  $|C^-|$  ( $|C^+|$ ). We also define the set of distinct countries as  $\hat{C}^-$  ( $\hat{C}^+$ ), with lengths  $|\hat{C}^-|$  and  $|\hat{C}^+|$ , thereby disregarding the multiplicity



**Figure 2.** The impact of mobility on geographical collaboration networks.  $C^+ \cap c_i^-$  ( $C^- \cap c_i^+$ ) represents the geographical overlap between a researcher's location before (after) migrating, denoted by  $c_i^-$  ( $c_i^+$ ), and the set of collaborator locations after (before) migrating, denoted by  $C^+$  ( $C^-$ ). (a) Shown is the fraction of mobility events corresponding to: (i) 'blind mobility' (blue): in which there is no overlap between a researcher's prior location and the locations of future collaborators ( $f[C^+ \cap c_i^- = 0] = 0.34$ ); and (ii) 'curtail mobility' (red): in which there is no overlap between the locations of a researcher's prior collaborators and his/her destination country ( $f[C^- \cap c_i^+ = 0] = 0.11$ ). (b) The distributions of collaboration network integrity, measured by  $f[C^+ \cap c_i^-]$  and  $f[C^- \cap c_i^+]$ , are right-skewed: on average there is only a 16% overlap between  $C^+$  and  $c_i^-$  and a 23% overlap between  $C^-$  and  $c_i^+$ . (c) The difference  $\Delta[C \cap C]$  measures the change in the amount of geographical overlap (see equation (2.2)), measured in two ways: per affiliation and per country. Negative values indicate more overlap before as compared to after. Both methods indicate relatively high levels of collaboration network disintegration following a mobility event: 89% (70%) of the values are negative when measuring per affiliation (per country). Shown are calculations on data aggregated across all three periods,  $T_{123}$ ; for analogous plots specific to a given period  $T$  see electronic supplementary material, figure S6. Vertical lines indicate distribution mean values. (Online version in colour.)

of the countries in the original lists. By way of example, consider the hypothetical list of countries associated with the affiliations derived from an arbitrary set of publications,  $C^- = \{CA, CA, FR, FR, AU, AU\}$ ; then the corresponding unweighted country list is  $\hat{C}^- = \{CA, FR, AU\}$ . These two definitions provide complementary perspectives on the degree to which geographical collaborations reorganize: the first measure ( $C^{+, -}$ ) is weighted proportional to the number of publications (i.e. per author affiliation), and the second is more aggregate (i.e. per country).

As such, we seek to quantify the degree to which the source country  $c_i^-$  and destination country  $c_i^+$  of a mobile researcher are related to  $C^{+, -}$ . Thus, we define the 'per-affiliation' geographical association of the source (destination) country  $c_i^-$  ( $c_i^+$ ) with  $C^-$  (or  $C^+$ ) as

$$\text{and } \left. \begin{aligned} f[C^+ \cap c_i^-] &= \frac{|C^+ \cap c_i^-|}{|C^+|} \\ f[C^- \cap c_i^+] &= \frac{|C^- \cap c_i^+|}{|C^-|} \end{aligned} \right\} \quad (2.1)$$

where  $|X \cap Y|$  denotes the number of elements in the intersection of the two sets  $X, Y$ . Similarly, the 'per-country' measure of geographical association is defined using  $\hat{C}^{+, -}$  in equation (2.1) instead.

Figure 2a shows that roughly 34% of the mobility events were 'blind', corresponding to the value  $f[C^+ \cap c_i^-] = 0$ , i.e. the destination country was not in the sphere of prior collaborations. Similarly, roughly 11% of the migrations corresponded to the scenario in which the sphere of destination collaborations did not intersect with the source country ( $f[C^- \cap c_i^+] = 0$ ), corresponding to the maximal curtailing of prior collaborations.

In addition to these extreme cases, we find that most other mobility events are followed by significant collaboration reorganization. Figure 2b shows the cumulative distribution function (CDF)  $\text{CDF}(\geq f[C^- \cap c_i^+])$  and  $\text{CDF}(\geq f[C^+ \cap c_i^-])$ , with both distributions more concentrated around small values. In the case of  $f[C^- \cap c_i^+]$ , this indicates a relatively

small likelihood that researchers maintain prior collaborations when they move.

We also analysed the dynamics of geographical coordination, conditional on the source country, defined as the difference

$$\Delta[C \cap C] = f[C^+ \cap c_i^-] - f[C^- \cap c_i^+], \quad (2.2)$$

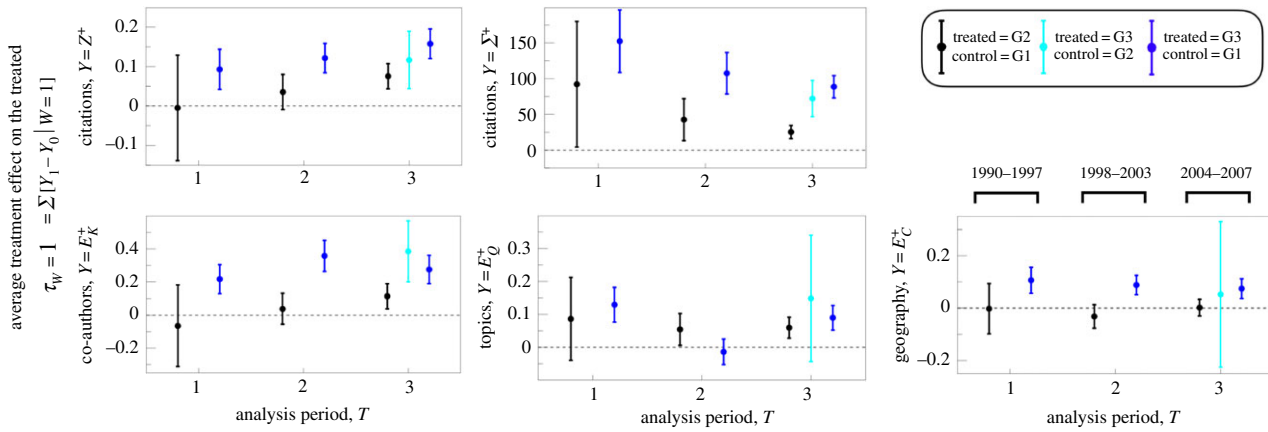
which is negative if there is more overlap between  $c_i^-$  and  $C^-$  than between  $c_i^-$  and  $C^+$ . Figure 2c shows that most  $\Delta[C \cap C]$  are negative—i.e. migration is associated with significant churning within collaboration networks. Electronic supplementary material, figure S6 shows the same distributions for data disaggregated by  $T$ , demonstrating the stability of these observations over time.

## 2.5. Estimating the mobility effect using propensity score matching

To what degree does researcher mobility affect scientific impact, topical direction and collaboration? In this section, we describe the Rubin causal inference framework [51] as it applies to estimating the impact of cross-border mobility on various quantitative researcher career metrics. Using potential outcome notation, we estimate the *average treatment effect on the treated* (ATET),

$$\begin{aligned} \tau_{W=1}[Y] &\equiv E[Y_1 - Y_0 | W = 1] \\ &\approx N_{W=1}^{-1} \sum_{i: W_i=1} (Y_i - \hat{Y}_i(0)), \end{aligned} \quad (2.3)$$

where  $Y_i$  indicates the outcome variable of interest, defined as one of the variables  $Z_i^+$ ,  $\Sigma_i^+$ ,  $E_{K,ir}^+$ ,  $E_{Q,ir}^+$  or  $E_{C,i}^+$  and  $E[Y_1 - Y_0]$  denotes the expected difference between the two counterfactual outcome measures. The indicator  $W_i = 1$  denotes 'treatment', i.e. mobility by individual  $i$  in period  $T$ ; conversely,  $W_i = 0$  corresponds to no mobility during  $T$ . Of course, a researcher is observed with either  $W_i = 0$  or  $W_i = 1$  in a given  $T$ , but not both; thus, the challenge is to estimate



**Figure 3.** Estimation of the effect of mobility on various career measures. For a given variable  $Y$  and period of analysis  $T$ , we applied the PSM method to compare a ‘treated’ group (either  $G_i = 2$  or  $3$ ) to a ‘control’ group (either  $G_i = 1$  or  $2$ ), resulting in three comparison: G1–G2, G2–G3 and G1–G3. Each panel reports the results for a different dependent variable  $Y$ ; shown are point estimates with error bar indicating 95% CI. (Online version in colour.)

the counterfactual outcome, i.e. what would have happened had the researcher  $i$  not migrated. Hence, we use the propensity score matching (PSM) method [52] to identify researcher pairs  $(i, i')$ , where  $i'$  is as similar as possible to  $i$  in terms of likelihood to belong to the treated group based on measured researcher characteristics (denoted by  $X_i^-$ ) prior to  $t_{i,T}^*$ . Equation (2.3) defines the mobility-effect estimate  $\tau_{W=1}[Y]$  according to the PSM method, which approximates the counterfactual outcome  $\hat{Y}_i(0)$  for researcher  $i$  with  $Y_i(0)$  by identifying the closest match  $i'$  from the pool of  $N_{W=0}$  researchers with  $W=0$ . In this way,  $\tau_{W=1}$  is approximated as the average researcher-level effect across the sample of  $N_{W=1}$  researchers with  $W=1$ . For a practical review of the matching estimators and their implementation see [53].

We estimated the treatment effect associated with each dependent variable  $Y_i^+$  using a vector  $X_i^-$  comprised of six covariates. We specified the logit model to calculate the likelihood of treatment for each  $i$ , which is used by PSM to match  $i$  with the closest  $i'$  (results are reported for single matches, which we found to be consistent with nearest-neighbour matching up to the five closest matches). To be specific, for a given dependent variable  $Y_i^+$ , the six covariates we used to match are: (i) the outcome variable calculated for the period before the mobility,  $Y_i^-$ ; (ii) the number of distinct co-authors,  $|k_{ij}^-|$ ; (iii) the number of publications,  $N_i^-$ ; (iv) the average citation impact,  $Z_i^-$ ; (v) the researcher age  $s_i^* = t_{i,T}^* - y_i^0 + 1$  in the year of the mobility event; and (vi) a factor variable  $F_i^-$  which maps the country of residence  $c_i^-$  to one of five geographical regions (N. America, S. & C. America, Europe, Asia & Australasia, Africa; see electronic supplementary material, A.3). Electronic supplementary material, figure S7 shows the distribution of each model variable and the corresponding correlation matrix.

Electronic supplementary material, table S1 shows the estimates of a logit model for the dependent variable  $1_{G_i=3}$ , with value 1 if a researcher  $i$  belongs to the mobile group  $G_i=3$  and 0 otherwise, thereby estimating the mobility likelihood depending on a given researcher’s attributes. See electronic supplementary material, appendix A.4 for analysis of the factors that correlate with cross-border mobility, along with refs. [1,3] which describe additional employment and other factors not feasible to include within our data-driven approach.

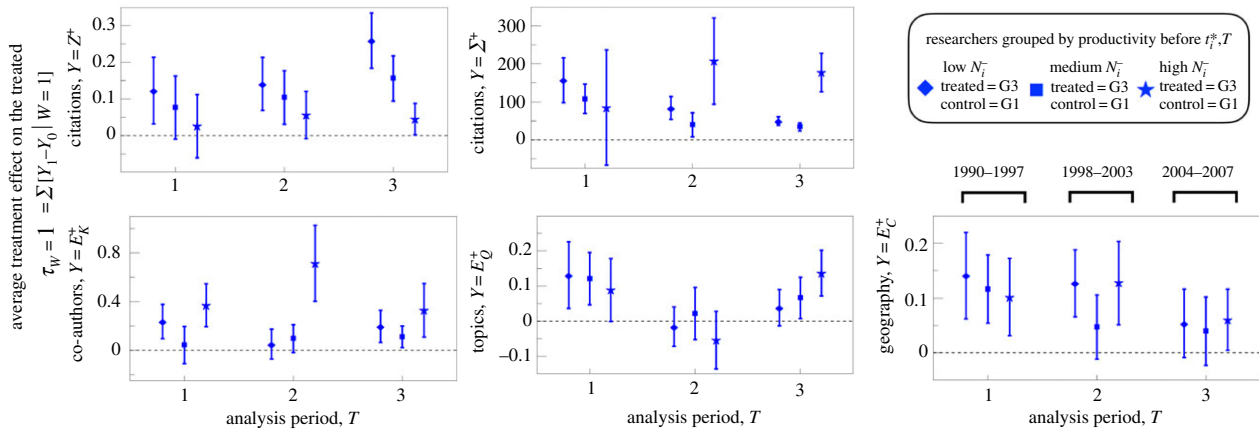
## 2.6. Impact of mobility on research outcomes

We applied the PSM method to five different dependent variables ( $Z^+$ ,  $\Sigma^+$ ,  $E_K^+$ ,  $E_Q^+$ ,  $E_C^+$ ) calculated for three non-overlapping analysis time periods and for three different control-treated subgroup comparisons. More specifically, the matching between the groups G3 (treated, comprised of researchers with  $G_{i,T}=3$ ) and G1 (control, comprised of researchers with  $G_{i,T}=1$ ) provides the core estimate of the mobility effect; the matching between G2 (placebo-treated) and G1 (placebo control) provides a robustness check, as we do not expect there to be significant differences between the researchers in these two groups; and the matching between G3 (treated) and G2 (control), serves as an additional robustness check, since we hypothesize the mobility effect to be present, but to a lesser degree than the G3–G1 estimation.

Figure 3 reports the resulting 35  $\tau_{W=1}[Y]$  mobility-effect estimations; insufficient samples sizes for the G2 and G3 groups limited the PSM matching performance for the periods  $T_1$  and  $T_2$  (see electronic supplementary material, figure S2(D) for the size of G1, G2, G3 for each  $T$ ). To demonstrate model robustness, electronic supplementary material, figure S8 shows the estimates using the nearest-neighbour method (*nnmatch* in STATA13) instead of the PSM method (*psmatch* in STATA13); the important difference is that with the *nnmatch* method one can force an exact match on the source region  $F_i^-$ . The results using *nnmatch* are consistent with the results of PSM, with the exception of the ATET for  $E_K^+$ , which are significantly smaller in significance and magnitude. The Material and methods section summarizes additional robustness checks to further assess the statistical significance of our PSM estimates.

The PSM results indicate that researcher mobility has a significant positive effect on citation impact and increases diversity in research topics and collaboration at the individual and geographical level. Notably, the effect size for mean citation impact,  $\tau_{W=1}[Z^+]$ , increased over each period  $T$ . Because  $z_p \equiv (\ln[n_p + 1] - \mu_i) / \sigma_i$  is a logarithmic transform of the citation count variable  $n_p$  (see Material and methods), the mobility effect for the average publication, in terms of the per cent increase in  $n_p$ , is approximately  $100 \times (\sigma) \times \tau_{W=1}[Z^+] = 9\%$  to  $17\%$  depending on  $T$ .

We also observe a decreasing trend for  $\tau_{W=1}[\Sigma^+]$ , which at first appears inconsistent with the trend for  $\tau_{W=1}[Z^+]$ . However, the  $\Sigma^+$  measure suffers from right-censoring bias



**Figure 4.** Estimation of the mobility effect by productivity subgroup. Researchers were separated into three groups according to the number of publications in the pre-treatment period,  $N_i^-$ : low (bottom tercile, triangles), medium (middle tercile, squares) and high (upper tercile, stars). For a given variable and period of analysis  $T$ , we applied the PSM method to compare the mobile ‘treated’ group ( $G_i = 3$ ) to the not-mobile ‘control’ group ( $G_i = 1$ ) corresponding to the group comparison G1–G3 in figure 3. Each error bar is a point estimate with 95% confidence interval. (Online version in colour.)

(publications analysed for later  $T$  have not the same amount of time to accrue citations as for publications from previous  $T$ ) and so the negative trend is confounded by this bias; the value of the citation premium is nevertheless statistically significant and relevant in magnitude:  $\tau_{W=1}[\Sigma^+] \approx 90$ –150 citations depending on  $T$ , which provides an estimate for the aggregate mobility effect tallying over all research produced in the 5-year period after mobility. This is a lower-bound estimate because the APS dataset does not include citations from articles published in other journals, nor does  $\Sigma^+$  account for the citation tally from researchers’ publications in non-APS journals. Thus, it is likely that  $\tau_{W=1}[\Sigma^+]$  would be significantly larger if non-APS publications and citations were included.

Among the entropy variables, the mobility effect is largest for  $\tau_{W=1}[E_K^+]$  (diversity of co-authors) and  $\tau_{W=1}[E_C^+]$  (diversity of geographical coordination). The effect of mobility on  $E_Q^+$  (diversity of research topics) is statistically significant for periods  $T_1$  and  $T_3$ , but not for  $T_2$ . These results quantify the extent to which mobile researchers facilitate a valuable interface for intellectual exchange. A prime example was the cross-disciplinary interface formed by computing and biology researchers in the genomics revolution, in which this constructive configuration facilitated the cross-pollination of methods and institutional know-how that were crucial for the success of the Human Genome Project [13].

We further estimated the mobility effect using OLS regression on matched researcher pairs ( $i, i'$ ), which facilitates incorporating covariates  $X^-$  to explain the outcome variable  $Y^+$  associated with each  $\tau_{W=1}[Y]$ . See electronic supplementary material, A.5 and tables S2–S6 for the description and results of this additional analysis. For example, we find that increasing researcher age correlates with lower research topic diversity,  $E_Q^+$ , suggesting that the ability to diversify into new research topics decreases with age ( $\beta_{s^*} < 0$ ; electronic supplementary material, table S5). We also observe a negative relation ( $\beta_{s^*} < 0$ ) between researcher age and citation impact variables, e.g.  $Z^+$  and  $\Sigma^+$ , consistent with findings from previous studies on researcher careers [11,13].

Are the observed mobility effects merely a feature of elite scientists in our dataset? To address this question, we separated the researchers into three terciles according to  $N_i^-$  in order to further analyse the degree to which variation in  $\tau_{W=1}[Y]$  is mediated by researcher productivity, a variable that is highly

correlated with a researcher’s characteristic team size, among other factors. Figure 4 reports the PSM estimates after matching researchers only within each subgroup, showing that the mobility effect on mean citation impact is largest for researchers belonging to the low- and medium- $N_i^-$  tercile groups; the mobility effect on co-author diversity is dominated by the high  $N_i^-$  tercile group; and no significant variation is observed for topic diversity or geographical diversity.

### 3. Discussion

The globalization of science and decreasing costs of migration have democratized the opportunities for international mobility [2], thereby becoming an increasingly relevant topic for national innovation system policy [1], the economics of science [5], and the multiscale modelling of the scientific system [6]. Against this backdrop, we assembled and analysed a relatively comprehensive longitudinal dataset of researchers in physics in order to accurately quantify how researcher mobility affects career growth and intellectual exploration at the micro-level, and the churning of the collaboration network facilitating international integration at the macro-level. We used statistical matching techniques to measure the differences between mobile researchers and similar non-mobile researchers, thereby addressing the reverse causality and sample selection biases that hinder estimating causal effects. To this end, we developed a methodical framework for comparing the shifts in a researcher’s publication profile in the 10-year period centred around each mobility event (see electronic supplementary material, figure S2 for a schematic of this framework).

We proceeded by separating our analysis into three time periods ( $T$ ) in order to be able to identify significant temporal trends. Qualitatively, our quantitative results are not sensitive to  $T$ , indicating that the mobility premium identified for each variable ( $Z^+, \Sigma^+, E_K^+, E_Q^+, E_C^+$ ) has not diminished as virtual avenues for exploring new collaborations proliferate. The results that are relevant to individual researchers are as follows. We measure a 9–17% increase in the number of citations received by articles published by mobile researchers, relative to the matched non-mobile control set. Aggregating this differential across all publications in the 5 years after migrating, on average this citation premium tallies up to

the order of 100 citations (figure 3). A hypothetical mechanism that emerges directly from our analysis of the research diversity variables ( $E_k^\pm$ ,  $E_Q^\pm$ ,  $E_C^\pm$ ) is that mobility increases the likelihood of drawing together different intellectual capacities and professional experience distributed across researchers, which when combined, provide a valuable perspective on how to best advance research efforts [17]. As such, this recombinant collaboration emerges as an important factor promoting recombinant innovation [54].

We complement this analysis based upon a broad cross-section of physicists with an alternative perspective based upon the extreme tail of scientific success—the 884 recipients of the Nobel Prize from 1901 to 2016. Analysis of their international trajectories reveals that 23% of these Nobelists were foreign-born, having received the award for work done in a country different than their birth country (figure 1). As such, these success stories provide ample evidence that just a single immigrant can have a monumental positive impact on a destination country, explaining to some degree the substantial international competition for elite scientists [2,3,5,21].

Motivated by recent work connecting migration imbalance (i.e. East–West ‘brain drain’) following the 2004 European Union enlargement to the stalled integration of national innovation systems across Europe [16,55], we also analysed the coevolution of mobility and collaboration from a systems perspective. In this regard, we provide additional micro-level evidence for such a mobility-mediated disintegration mechanism by showing that researchers completely curtail all former international collaborations roughly 11% of the time they migrate (figure 2). Conversely, we also found that 34% of the mobile researchers in our analysis moved to a location that was not in their sphere of prior collaborations. In addition to these exceptional cases, we observed remarkably high levels of churning following most mobility events. In this way, migration affords a prime opportunity to phase out old (possibly stagnant) collaborations and research topics in order to make way for exploring new (potentially lucrative) avenues. Other scholars have also noted that having migrated once likely increases the likelihood to migrate again, not merely due to additional experience and openness in regards to future opportunities, but also because recently shuffled collaboration networks are likely to be less restricting [31].

In conclusion, we analysed the dynamic interpersonal and international interface mediated by mobile researchers, one that facilitates the exchange of knowledge, as well as other valuable professional and cultural perspectives. While our focus was on mobile researchers in physics, our results are generalizable to other scientific and intellectual domains where cross-border experience confers a considerable comparative advantage by virtue of increased exposure to group diversity [56] and opportunities to broker knowledge [30], thereby promoting career development along an otherwise uncertain career path [24,29,42–44].

Against this background, recent scholarly efforts have provided considerable new insights into the professional, social and familial factors influencing scientists’ decisions to stay or go [3,31,49]. For example, survey responses from more than 15 000 respondents indicated that ‘opportunities to improve future career prospects’ and ‘prestige/excellence of the institution’ were the two most important factors underlying their decision to move abroad [31]; however, a separate large study found that researchers were significantly less likely to relocate if they recently obtained competitive funding or if

their children were in high school [3]. Related empirical work has also sought to determine whether other types of mobility—i.e. institutional, social, or inter-sectoral—produce a measurable effect on subsequent productivity and citation impact [28,57,58]; altogether these studies consistently report positive shifts in productivity, but a little if not insignificant effect on citation impact, when moving to a better institution.

International recruitment and home-return policy are typically geared around elite scientists [2,21,32], yet our results suggest that the career benefits of mobility are common to all ranks. Another important consideration is immigration policy, which can have a significant impact on the attraction and retention of talented researchers, thereby affecting the development of national innovation systems. An important example is the 2000 change in USA H-1B visa policy, which nullified the cap on available visas for non-profit sponsors, thereby eliminating the competition between industry and academia over this critical type of visa that facilitates high-skilled immigration [5]. By analysing the annual changes in the number of H-1B visa recipients in industrial science and engineering, scholars found that higher rates of Chinese and Indian H-1B recipients correlated with higher levels of employment and patenting in innovation hubs that depend on immigrant high-skilled labour; moreover, the scholars report limited effects of the increase in H-1B immigrant population on native inventors’ employment and patenting productivity [59]. These are important results showing that immigration of high-skilled labour does not appear to crowd out native employment opportunities and innovation capacity. And finally, countries with travel visa restrictions are negatively associated with international researcher mobility, in particular because they significantly increase the cost of travel, thereby inhibiting short-term collaboration visits that may precede the opportunity and ultimate decision to migrate long-term [49].

Considered in this way, our results contribute to these discussions by suggesting that national innovation systems and high-skilled immigration policy should not be so exclusively focused on the attraction and retention of elite scientists, but rather, should extend strategies to develop competitiveness by fostering international community and cultural diversity that serve all ranks of scientists. Moving forward on these issues, especially in consideration of the inherent difficulties—both ethical and statistical—in testing and empirically measuring the impacts of immigration policy [60], there is a need to develop better data-driven analytical and systems modelling methods to inform science policy on these issues. Indeed, anticipating the multiscale impacts of science policy is a formidable challenge calling for continued trans-disciplinary efforts to better understand the scientific enterprise [5,6,13,61].

## 4. Material and methods

### 4.1. Network-based author disambiguation method

We use the rich publication metadata in the APS dataset as input for a network-based author disambiguation method [37] that groups publications into researcher profiles. The Helbing algorithm leverages three key pieces of information available for each publication: (i) the co-author names, (ii) the publications listed in the reference list, i.e. outgoing citations, and (iii) the list of incoming citations from other publications (note that incoming and outgoing citations are restricted to publications within the APS dataset). Importantly, we did not provide any



geographical affiliation information to the disambiguation algorithm, and so the resolved researcher profiles are free of geographical biases, and thus, well-suited for the study of cross-border mobility.

In an effort to assess statistically significant changes in publication profile attributes, and to reduce the frequency of spurious fluctuations arising from small sample size, we only analysed profiles with  $N_i \geq 10$  publications in the APS dataset and initial publication year  $y_i^0 \geq 1985$ , resulting in 26 170 researcher profiles; electronic supplementary material, figure S1(C) shows the distribution  $y_i^0$  for our final dataset, with 1995 as the median starting-year value.

## 4.2. Specification of research activity measures

### 4.2.1. Citation impact

In order to compare the citation impact of publications (indexed by  $p$ ) from different years ( $t$ ), we apply a normalization method that maps the integer citation count  $n_{i,p,t}$  to a z-score using a logarithmic transform. To be specific,  $z_{i,p} \equiv (\ln(n_{i,p,t} + 1) - \mu_t) / \sigma_t$ , where  $\mu_t = \langle \ln(n_t + 1) \rangle$  and  $\sigma_t = \sigma[\ln(n_t + 1)]$  denote the mean and standard deviation calculated over the set of publications from a specific year cohort ( $t$ ). By mapping citations to their log value (adding +1 to avoid the divergence associated with  $\ln 0$ ), and accounting for the age-cohort-specific mean and standard deviation,  $z_{i,p}$  follows a normal distribution  $N(0, 1)$  that is stable across  $t$  [11,13]. Publications with  $z_{i,p} > 0$  are thus above the average log citation impact,  $\mu_t$ , and since they are measured in units of  $\sigma$ , standard intuition and statistics of z-scores apply. As such,  $z_{i,p}$  is well-suited for cross-temporal analysis, e.g. OLS regression, as well as averaging and summing within profiles. For this reason we calculate the average citation impact value  $Z_i^{+,-} \equiv \langle z_{p,i}^{+,-} \rangle = (1/N_i^{+,-}) \sum_{p \in N} z_{p,i}^{+,-}$  across the  $N_i^{+,-}$  publications in each  $\Delta t \equiv 5$ -year interval, before ( $-$ ) or after ( $+$ )  $t_{i,T}^*$ . Moreover, due to the properties of logs, the per cent difference in  $n_t$  associated with the mobility effect is given by  $100\Delta n_p/n_p = 100 \times \sigma_t \times (\partial z/\partial W) \approx 100 \times \langle \sigma \rangle \times \tau_{W=1}[Z^+]$ , which follows because  $\sigma_t$  is approximately constant over time, and so we approximate  $\sigma_t$  with the average value  $\langle \sigma \rangle = 1.05$ .

For comparison, we also considered an ‘extensive’ citation impact measure, as opposed to  $z_{p,i}$ , which is an ‘intensive’ impact measure. The deflated citation count  $\hat{n}_{i,p,t}$  accounts for the fact that the total number of references produced by science is steadily growing with time (electronic supplementary material, figure S1(A)), the result of which is a ‘citation inflation’ measurement bias associated with the nominal citation count  $n_{i,p,t}$  [50]. However, the ‘deflated’ variable  $\hat{n}_{i,p,t}$  is well-suited for comparisons of citation counts for articles published in different years, and so we instead tally  $\hat{n}_{i,p,t}$  for each interval, defining the total as  $\Sigma_i^{+,-} \equiv \sum_{p \in +,-} \hat{n}_{i,p,t}^{+,-}$ . As such,  $\Sigma_i^{+,-}$  is amenable to modelling differences in total citation impact before and after  $t_{i,T}^*$ .

### 4.2.2. Co-authors

Within each  $\Delta t \equiv 5$ -year interval,  $K_i$  denotes the number of distinct co-authors we count for each  $i$ . Similarly, the number of publications including central researcher  $i$  and co-author  $j$  is  $k_{ij}$ ; the total number of co-author instances across all  $N_i$  publications is  $A_i = \sum_j k_{ij}$ . We then use the Shannon entropy, a diversity index measuring the variety across the  $K_i$  co-authors, defined

as  $E_{K,i} = -\sum_{j=1}^{K_i} (k_{ij}/A_i) \ln(k_{ij}/A_i)$ . More specifically, we calculate the entropy  $E_{K,i}^-$  ( $E_{K,i}^+$ ) using data in the  $\Delta t$ -year interval before (after)  $t_{i,T}^*$ . Higher levels of variation correspond to larger entropy values  $E \geq 0$ ; the limiting case of no variation,  $k_{ij} = \text{const.}$  for all  $j$ , corresponds to  $E = 0$ .

### 4.2.3. Research topics

Similar to (§4.2.2), we aggregate the PACS codes associated with each publication, a system used by the American Institute of Physics and implemented broadly in physics journals since 1975; see <https://publishing.aip.org/publishing/pacs/pacs-2010-regular-edition>. This five-level classification is comprised of more than 5000 individual PACS codes, which authors self-assign to their publications (e.g. ‘89.75.-k’ corresponds to ‘Complex systems’); we observe on average 2.5 unique PACS per publication with only 1% of publications having five or more PACS. We aggregated the PACS codes from all publications in each observation period into the two lists denoted by  $q_j^{+,-}$ . We then define the variation  $E_{Q,i}^{+,-}$  in each  $q_j^{+,-}$  list of PACS codes using the same Shannon entropy measure in (ii) above.

### 4.2.4. Geographical reach

As above, we aggregate the country codes associated with each publication affiliation into two lists,  $C_i^{+,-}$ . We define the variation  $E_{C,i}^{+,-}$  in the categorical country code lists using the Shannon entropy (e.g. the list  $C = \{\text{NL}, \text{NL}, \text{NL}, \text{UK}, \text{IT}, \text{JP}, \text{AU}\}$  has entropy  $E_C = 1.475$ ).

## 4.3. Robustness check of propensity score matching method

We randomized the treatment/control group classification variable  $1_{G_i} = 3$  to test the likelihood that we could obtain an effect size as large as the observed  $\tau_{W=1}[Y]$  by chance. To be specific, we shuffled the treatment indicator  $1_{G_i} = 3$  within each G3–G1 estimation, without replacement, thereby conserving the total number of observations (researchers) classified as being mobile during a specific period (i.e. belonging to group  $G_i = 3$ ). We applied this shuffling procedure 10 000 different times, each time recording the value of the ‘placebo’ estimate for  $\tau_{W=1}[Y]$ . Figure S9 shows the distribution of placebo estimates,  $P(\tau_{W=1}[Y])$ , for each  $Y$ ; in estimates, with one exception (panel D for  $\tau_{W=1}[E_Q^+]$ ), we find the observed treatment effect to be significantly larger and outside the 99% CI bounds of the placebo estimate, thereby demonstrating the robustness of our PSM specification to spurious correlations.

**Data accessibility.** We used two principal open-source datasets provided by the American Physical Society [36] and WolframAlpha [35]; raw mobility data will be made available upon request.

**Competing interests.** We declare we have no competing interests.

**Funding.** We received no funding for this study.

**Acknowledgements.** We thank the A.P.S. for making their data available. We are also grateful to two anonymous referees and F. Arroyave Bermudez for helpful comments, and acknowledge the developers of the open-source *Circos* circular network layout Table Viewer software [62] used to produce the Nobelist mobility network.

## References

- Geuna A (ed.). 2015 *Global mobility of research scientists*. Oxford, UK: Academic Press.
- Hunter RS, Oswald AJ, Charlton BG. 2009 The elite brain drain. *Econ. J.* **119**, F231–F251. (doi:10.1111/j.1468-0297.2009.02274.x)
- Azoulay P, Ganguli I, Zivin JG. 2017 The mobility of elite life scientists: professional and personal determinants. *Res. Policy* **46**, 573–590. (doi:10.1016/j.respol.2017.01.002)
- Petersen AM, Puliga M. 2016 High-skilled labour mobility in Europe before and after the 2004 enlargement. *J. R. Soc. Interface* **14**, 20170030. (doi:10.1098/rsif.2017.0030)

5. Stephan P. 2012 *How economics shapes science*. Cambridge, MA: Harvard University Press.
6. Fortunato S *et al.* 2018 Science of science. *Science* **359**, eaao0185. (doi:10.1126/science.aao0185)
7. Agrawal A, Cockburn I, McHale J. 2006 Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *J. Econ. Geogr.* **6**, 571–591. (doi:10.1093/jeg/lbl016)
8. Breschi S, Lissoni F. 2009 Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *J. Econ. Geogr.* **9**, 439–468. (doi:10.1093/jeg/lbp008)
9. Petersen AM, Fortunato S, Pan RK, Kaski K, Penner O, Rungi A, Riccaboni M, Stanley HE, Pammolli F. 2014 Reputation and impact in academic careers. *Proc. Natl Acad. Sci. USA* **111**, 15 316–15 321. (doi:10.1073/pnas.1323111111)
10. Mazloumian A, Eom YH, Helbing D, Lozano S, Fortunato S. 2011 How citation boosts promote scientific paradigm shifts and Nobel prizes. *PLoS ONE* **6**, e18975. (doi:10.1371/journal.pone.0018975)
11. Petersen AM. 2015 Quantifying the impact of weak, strong, and super ties in scientific careers. *Proc. Natl. Acad. Sci. USA* **112**, E4671–E4680. (doi:10.1073/pnas.1501444112)
12. Franzoni C, Rossi-Lamastra C. 2017 Academic tenure, risk-taking and the diversification of scientific research. *Ind. Innovation* **24**, 691–712. (doi:10.1080/13662716.2016.1264067)
13. Petersen AM, Majeti D, Kwon K, Ahmed ME, Pavlidis I. 2018 Cross-disciplinary evolution of the genomics revolution. *Sci. Adv.* **4**, eaat421. (doi:10.1126/sciadv.aat4211)
14. Jonkers K, Cruz-Castro L. 2013 Research upon return: the effect of international mobility on scientific ties, production and impact. *Res. Policy* **42**, 1366–1377. (doi:10.1016/j.respol.2013.05.005)
15. Scellato G, Franzoni C, Stephan P. 2015 Migrant scientists and international networks. *Res. Policy* **44**, 108–120. (doi:10.1016/j.respol.2014.07.014)
16. Doria Arrieta OA, Pammolli F, Petersen AM. 2017 Quantifying the negative impact of brain drain on the integration of European science. *Sci. Adv.* **3**, e1602232. (doi:10.1126/sciadv.1602232)
17. Fleming L, Sorenson O. 2004 Science as a map in technological search. *Strategic Manage. J.* **25**, 909–928. (doi:10.1002/smj.384)
18. Fleming L, Mingo S, Chen D. 2007 Collaborative brokerage, generative creativity, and creative success. *Adm. Sci. Q.* **52**, 443–475. (doi:10.2189/asqu.52.3.443)
19. Foster JG, Rzhetsky A, Evans JA. 2015 Tradition and innovation in scientists' research strategies. *Am. Sociol. Rev.* **80**, 875–908. (doi:10.1177/0003122415601618)
20. Jia T, Wang D, Szymanski B. 2017 Quantifying patterns of research-interest evolution. *Nat. Hum. Behav.* **1**, 0078. (doi:10.1038/s41562-017-0078)
21. Laudel G. 2005 Migration currents among the scientific elite. *Minerva* **43**, 377–395. (doi:10.1007/s11024-005-2474-7)
22. Hoisl K. 2007 Tracing mobile inventors—the causality between inventor mobility and inventor productivity. *Res. Policy* **36**, 619–636. (doi:10.1016/j.respol.2007.01.009)
23. Marx M, Strumsky D, Fleming L. 2009 Mobility, skills, and the Michigan non-compete experiment. *Manage. Sci.* **55**, 875–889. (doi:10.1287/mnsc.1080.0985)
24. Cruz-Castro L, Sanz-Menendez L. 2010 Mobility versus job stability: assessing tenure and productivity outcomes. *Res. Policy* **39**, 27–38. (doi:10.1016/j.respol.2009.11.008)
25. Edler J, Fier H, Grimpe C. 2011 International scientist mobility and the locus of knowledge and technology transfer. *Res. Policy* **40**, 791–805. (doi:10.1016/j.respol.2011.03.003)
26. Franzoni C, Scellato G, Stephan P. 2014 The mover's advantage: the superior performance of migrant scientists. *Econ. Lett.* **122**, 89–93. (doi:10.1016/j.econlet.2013.10.040)
27. Baruffaldi SH, Landoni P. 2012 Return mobility and scientific productivity of researchers working abroad: the role of home country linkages. *Res. Policy* **41**, 1655. (doi:10.1016/j.respol.2012.04.005)
28. Deville P, Wang D, Sinatra R, Song C, Blondel VD, Barabasi AL. 2014 Career on the move: geography, stratification, and scientific impact. *Sci. Rep.* **4**, 4770. (doi:10.1038/srep04770)
29. Baker A. 2015 Non-tenured post-doctoral researchers' job mobility and research output: an analysis of the role of research discipline, department size, and coauthors. *Res. Policy* **44**, 634–650. (doi:10.1016/j.respol.2014.12.012)
30. Wang D. 2015 Activating cross-border brokerage: interorganizational knowledge transfer through skilled return migration. *Adm. Sci. Q.* **60**, 133–176. (doi:10.1177/0001839214551943)
31. Franzoni C, Scellato G, Stephan P. 2015 International mobility of research scientists: lessons from globsci. In *Global mobility of research scientists*, pp. 35–65. Amsterdam, the Netherlands: Elsevier.
32. Kahn S, MacGarvie M. 2016 Do return requirements increase international knowledge diffusion? Evidence from the Fulbright program. *Res. Policy* **45**, 1304–1322. (doi:10.1016/j.respol.2016.02.002)
33. Sugimoto CR, Robinson-García N, Murray DS, Yegros-Yegros A, Costas R, Larivière V. 2017 Scientists have most impact when they're free to move. *Nature* **550**, 29–31. (doi:10.1038/550029a)
34. Fernandez-Zubieta A, Geuna A, Lawson C. 2015 What do we know of the mobility of research scientists and impact on scientific production? In *Global mobility of research scientists*, pp. 1–33. Amsterdam, the Netherlands: Elsevier.
35. Wolfram Alpha LLC. 2016 Nobel Prize in physics. See <https://www.wolframalpha.com/input/?i=Nobel+Prize+in+Physics>.
36. American Physical Society. 2009 APS data sets for research. See <http://journals.aps.org/datasets>.
37. Schulz C, Mazloumian A, Petersen AM, Penner O, Helbing D. 2014 Exploiting citation networks for large-scale author name disambiguation. *EPJ Data Sci.* **3**, 11. (doi:10.1140/epjds/s13688-014-0011-3)
38. Chait R. 2009 *The questions of tenure*. Cambridge, MA: Harvard University Press.
39. Kaplan K. 2010 Academia: the changing face of tenure. *Nature* **468**, 123–125. (doi:10.1038/nj7320-123a)
40. Cyranoski D, Gilbert N, Ledford H, Nayar A, Yahia M. 2011 Education: the PhD factory. *Nature* **472**, 276–279. (doi:10.1038/472276a)
41. Powell K. 2015 The future of the postdoc. *Nature* **520**, 144–147. (doi:10.1038/520144a)
42. Kaminski D, Geisler C. 2012 Survival analysis of faculty retention in science and engineering by gender. *Science* **335**, 864–866. (doi:10.1126/science.1214844)
43. Petersen AM, Riccaboni M, Stanley HE, Pammolli F. 2012 Persistence and uncertainty in the academic career. *Proc. Natl Acad. Sci. USA* **109**, 5213–5218. (doi:10.1073/pnas.1121429109)
44. Way SF, Morgan AC, Clauset A, Larremore DB. 2017 The misleading narrative of the canonical faculty productivity trajectory. *Proc. Natl Acad. Sci. USA* **114**, E9216–E9223. (doi:10.1073/pnas.1702121114)
45. Duch J, Zeng XHT, Sales-Pardo M, Radicchi F, Otis S, Woodruff TK, Amaral LAN. 2012 The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS ONE* **7**, e51332. (doi:10.1371/journal.pone.0051332)
46. Jadidi M, Karimi F, Lietz H, Wagner C. 2017 Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *Adv. Complex Syst.* **21**, 1750011. (doi:10.1142/s0219525917500114)
47. Way SF, Larremore DB, Clauset A. 2016 Gender, productivity, and prestige in computer science faculty hiring networks. In *Proc. of the 25th Int. Conf. on World Wide Web*, pp. 1169–1179.
48. Lerchenmueller MJ, Sorenson O. 2018 The gender gap in early career transitions in the life sciences. *Res. Policy* **47**, 1007–1017. (doi:10.1016/j.respol.2018.02.009)
49. Appelt S, van Bezuekom B, Galindo-Rueda F, de Pinho R. 2015 Which factors influence the international mobility of research scientists? In *Global mobility of research scientists*, pp. 177–213. Amsterdam, the Netherlands: Elsevier.
50. Petersen AM, Pan RK, Pammolli F, Fortunato S. 2018 Methods to account for citation inflation in research evaluation. SSRN e-print: 3193712.
51. Rubin DB. 2005 Causal inference using potential outcomes. *J. Am. Stat. Assoc.* **100**, 322–331. (doi:10.1198/016214504000001880)
52. Rosenbaum PR, Rubin DB. 1983 The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55. (doi:10.1093/biomet/70.1.41)
53. Abadie A, Drukker D, Herr JL, Imbens GW. 2004 Implementing matching estimators for average treatment effects in STATA. *STATA J.* **4**, 290–311.

54. Fleming L. 2001 Recombinant uncertainty in technological search. *Manage. Sci.* **47**, 117–132. (doi:10.1287/mnsc.47.1.117.10671)
55. Chessa A, Morescalchi A, Pammolli F, Penner O, Petersen AM, Riccaboni M. 2013 Is Europe evolving toward an integrated research area? *Science* **339**, 650–651. (doi:10.1126/science.1227970)
56. Page SE. 2008 *The difference: how the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.
57. Allison PD, Long JS. 1990 Departmental effects on scientific productivity. *Am. Sociol. Rev.* **55**, 469–478. (doi:10.2307/2095801)
58. Fernández-Zubieta A, Geuna A, Lawson C. 2015 Mobility and productivity of research scientists. In *Global mobility of research scientists*, pp. 105–131. Amsterdam, the Netherlands: Elsevier.
59. Kerr WR, Lincoln WF. 2010 The supply side of innovation: H-1B visa reforms and US ethnic invention. *J. Labor. Econ.* **28**, 473–508. (doi:10.1086/651934)
60. Simon M, Schwartz C, Hudson D, Johnson SD. 2018 A data-driven computational model on the effects of immigration policies. *Proc. Natl Acad. Sci. USA* **115**, E7914–E7923. (doi:10.1073/pnas.1800373115)
61. Fealing KH (eds) 2011 *The science of science policy: a handbook*. Stanford, CA: Stanford Business Books.
62. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009 Circos: an information aesthetic for comparative genomics. *Genome. Res.* **19**, 1639–1645. (doi:10.1101/gr.092759.109)