SOCIAL SCIENCES

Quantifying the negative impact of brain drain on the integration of European science

Omar A. Doria Arrieta,[1,2]* Fabio Pammolli,[1,3] Alexander M. Petersen[4]†*

The 2004/2007 European Union (EU) enlargement by 12 member states offers a unique opportunity to quantify the impact of EU efforts to expand and integrate the scientific competitiveness of the European Research Area (ERA). We apply two causal estimation schemes to cross-border collaboration data extracted from millions of academic publications from 1996 to 2012, which are disaggregated across 14 subject areas and 32 European countries. Our results illustrate the unintended consequences following the 2004/2007 enlargement, namely, its negative impact on cross-border collaboration in science. First, we use the synthetic control method to show that levels of European cross-border collaboration would have been higher without EU enlargement, despite the 2004/2007 EU entrants gaining access to EU resources incentivizing cross-border integration. Second, we implement a difference-in-difference panel regression, incorporating official intra-European high-skilled mobility statistics, to identify migration imbalance—principally from entrant to incumbent EU member states—as a major factor underlying the divergence in cross-border integration between Western and Eastern Europe. These results challenge central tenets underlying ERA integration policies that unifying labor markets will increase the international competitiveness of the ERA, thereby calling attention to the need for effective home-return incentives and policies.

INTRODUCTION

Despite positive trends in the globalization of research and development (R&D) and collaboration in science (*1*–*5*), recent studies of international collaboration show that national borders are still a formidable hindrance to cross-border activity. As for scientific publication, this comes as somewhat a surprise in the case of Europe, especially considering its long history of policies aimed at reducing national barriers to promote the development of the European Research Area (ERA) (*6*–*9*). In reality, the globalization of science via international collaboration has not evolved uniformly across all countries and regions. As a relevant example, compare the decade before and after 2004, when Western Europe and North America experienced a 36 to 42% increase in the rate of cross-border collaboration (per publication), whereas Eastern Europe and Asia experienced a much slower 9% growth (see Fig. 1). These diverging trends point to the importance of historical, sociotechnological, and geographical factors affecting the globalization of science (*10*–*13*). In addition, they also point to important issues concerning the future of Europe.

So, why have Western and Eastern Europe followed different paths of cross-border integration in science? To provide insight into this divergence phenomena, we constructed a longitudinal data set for 32 European countries over the 17-year period 1996–2012 by aggregating data from four different sources: (i) publication count, citation count, and international collaboration rate data from SCImago Journal and Country Rank (disaggregated across 14 research subject areas indexed here by s, for example, $s = 1300$: "biochemistry, genetics, and molecular biology"); (ii) government investment in R&D data from the World Bank; (iii) official country-country pairwise counts of incoming/outgoing European Union (EU) high-skilled labor mobility from the EU Single Market Regulated Professions Database (*14*); and (iv) global migration data from Abel and Sander (*15*). See Materials and Methods for further description of these data sets.

We then used these data to compare the levels of cross-border collaboration before and after the 2004 EU enlargement, thereby illuminating the complex relations between the integration of European labor markets, "brain drain" (*16*–*22*), and the arrested development of the ERA (*8*, *9*). A hypothetical mechanism connecting these three elements is rather intuitive, following from the dynamic interpersonal nature of scientific collaboration (*23*): We hypothesize that Europe experienced a significant loss of cross-border integration because as mobile academics pursued international career paths, likely by following their previous collaboration channels, the cross-border links that they previously mediated between their home country and their destination country were subsequently eliminated. Thus, in addition to demonstrating how policy shifts can inadvertently spur high-skilled migration (*24*), we also demonstrate additional negative externalities on subsequent cross-border activity here. Because high-impact research is more likely to occur in multicountry collaborations (*5*), our findings show how migration imbalance can negatively affect the convergence of scientific competitiveness across Europe.

RESULTS

To measure the impact of the EU enlargement on the rate of international collaboration in Europe, we implemented two causal inference methods (*25*)—the synthetic control method (SCM) (*26*) and a difference-in-difference (DiD) panel regression model. In each method, we use the EU enlargement—10 entrants in 2004 (CY, CZ, EE, HU, LT, LV, MT, PL, SK, and SI) and 2 entrants in 2007 (BG and RO)—as a multicountry two-stage policy intervention corresponding to the ("treatment") years $t^* = 2004$ and 2007, respectively (for a list of the expanded forms of the abbreviated European country names used in this study, see Countries analyzed in Materials and Methods). Hence, we separated the European countries into two groups, the first comprising the incumbent 2004 EU members and the second comprising the 12 entrant countries.

The dependent variable in our analysis is a country's level of cross-border activity, operationalized using SCImago's cross-border counting

[1] Laboratory for the Analysis of Complex Economic Systems, Institutions Markets Technologies (IMT) Lucca School for Advanced Studies, Lucca 55100, Italy. [2] The Cambridge Management Consulting Labs S.p.A., Milan 20121, Italy. [3] Department of Management, Economics, and Industrial Engineering, Politecnico di Milano, Milan 20156, Italy. [4] Ernest and Julio Gallo Management Program, School of Engineering, University of California, Merced, Merced, CA 95343, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: apetersen3@ucmerced.edu

scheme, which is derived from the affiliation information listed in each publication's author byline. Specifically, if a publication includes author affiliations from more than one country, the publication is counted as "cross-border" for all countries involved in the publication. Thus, for each country i and year t, SCImago reports the fraction $f_{i,t}$ of the total publications ($D_{i,t}$) involving cross-border collaboration. Then, we calculated the total number of publications $\chi_{i,t} = f_{i,t}D_{i,t}$. Although there may be other cross-border counting schemes, which account differently for the total number of authors, countries, and even the number of affiliations per author, we lack comprehensive publication-level information from Scopus required to implement a sensitivity analysis along these lines; see Publication data in Materials and Methods for a discussion of this and alternative counting schemes. Note that, among alternative counting schemes, the one implemented by SCImago is particularly amenable to statistical analysis because it reduces the interdependency between the observations at the country level. That is, a country receives the same cross-border credit for an international publication, independent of the details of author affiliations and the other countries involved.

SCM estimates

We start by framing our SCM analysis in potential outcomes notation (*27*). Let the outcome variable $Y_{i,t}(1) = f_{i,t}$(or $\chi_{i,t}$) correspond to the cross-border activity of a country directly affected by the EU enlargement and $Y_{i,t}(0) = f_{i,t}$(or $\chi_{i,t}$) correspond to the cross-border activity of a non-European country. Thus, we consider the treatment (or intervention) as the enlargement of the EU, with EU membership status providing unique access to the EU's large funding programs and the "freedom of movement" for persons and workers, which are fundamental tenets of EU policy. The set of units in our SCM analysis is divided into three subsets: (i) a balanced panel of 26 non-European countries (those with nonzero publication counts in each s for all t), (ii) the 10 countries entering in the 2004 enlargement, and (iii) 15 incumbent EU countries along with 4 close EU affiliates (CH, IS, LI, and NO), which have special trade and mobility agreements with the EU. To simplify the demonstration of the SCM, we collected and averaged the countries in the second and third groups. The result is an averaged representative unit for each group, which from here on we refer to as the "EU entrants" and "EU incumbents," respectively.

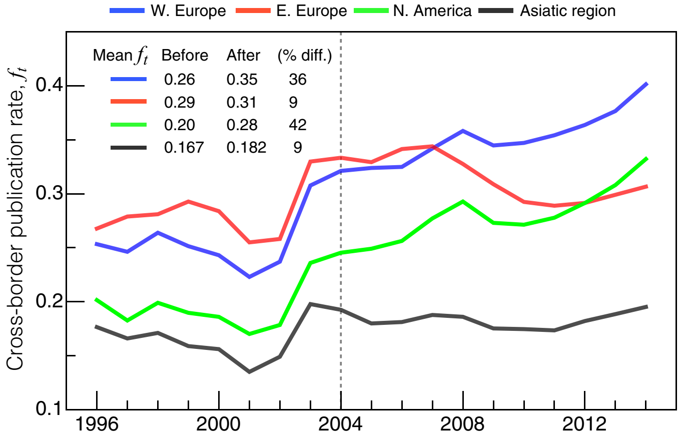

Fig. 1. Eastern-Western European divergence. Global trends in cross-border collaboration by international region: 1996–2014. Source: SCImago Journal and Country Rank based on Scopus (*40*). Notably, the curves for Western (W.) Europe and Eastern (E.) Europe are, before 2004, characterized by a roughly constant offset, thereby satisfying the prior equal slope condition of the DiD framework.

The period before $t^* = 2004$ corresponds to the pretreatment period, for which we assume there to be no difference between $Y_{i,t<t^*}(0)$ and $Y_{i,t<t^*}(1)$; that is, anticipation of the enlargement had no impact on characteristic levels of cross-border activity before 2004. We then use the 2004 enlargement to estimate the counterfactual difference δ between $Y_{i,t\geq t^*}(1)$ and $Y_{i,t\geq t^*}(0)$, which is the unit-level causal effect of EU enlargement on $Y_{i,t}$. Because the entrant and incumbent data are averages, we are actually estimating the mean causal effect for each collection of units, which has been established as an unbiased estimator of the average unit-level causal effect (*27*).

Of course, in reality, we only observe the potential outcome $Y_{i,t\geq t^*}(1)$ for the EU entrants and incumbents. However, the power of the SCM is to estimate the alternative potential outcome $Y_{i,t\geq t^*}(0)$—that is, the counterfactual scenario corresponding to no EU enlargement—by extrapolating synthetic $\hat{f}_{i,t}$ and $\hat{\chi}_{i,t}$ for $t \geq 2004$ using a basis set of countries for which we assume that $Y_{i,t\geq t^*}(1) = Y_{i,t\geq t^*}(0)$. More specifically, our SCM approach estimates $\hat{f}_{i,t}$ and $\hat{\chi}_{i,t}$ using a panel data set composed of four covariates (which are used as matching variables) as input: (i) the total number of publications ($\log_{10}D^s_{i,t}$), (ii) the normalized citations ($R^s_{i,t}$), (iii) the per-capita gross domestic product (GDP) ($\log_{10}GDPpc_{i,t}$), and (iv) government expenditure on R&D as a percentage of GDP, $e_{i,t}$. Provided these data, the SCM estimates an optimal set of weights using the data for $t < 2004$, thereby allowing for the extrapolation of $\hat{f}_>(\hat{\chi}_>)$ for the EU entrant countries based on their projection onto the subspace of covariate data for the 26 non-European control countries [see the study of Varian (*25*) and the Supplementary Materials for further SCM details]. Note that the extent to which the individual covariates explain the dependent variable is not the aim of the SCM; instead, we use a panel regression in the following section to infer the quantitative relations between the covariates and the trends in $f_{i,t}$.

Figure 2 (A and B) shows the empirical curves (f_t and χ_t) measuring the cross-border activity for both the EU incumbents and the EU entrants. In terms of R&D investment and scientific output, the representative entrant county is medium sized (that is, between SG and RU), and the representative incumbent country is large (for example, CA). For this reason, we divided the incumbent χ_t curves by a factor of 10 in Fig. 2 (A and B) to facilitate visual comparison. Along with the real data (indicated by solid lines), each panel also shows the SCM estimates $\hat{f}_t$ and $\hat{\chi}_t$ (indicated by dashed lines).

The difference between $\hat{f}_{i,t}$ and $f_{i,t}$ is an estimate of the country-level causal effect, $\delta \equiv Y_{i,t\geq t^*}(0) - Y_{i,t\geq t^*}(1)$. Because the fraction $f_{i,t}$ is an intensive variable, whereas the total publications $\chi_{i,t}$ is an extensive variable, we measure δ slightly differently for each variable. For f_t, we estimate δ using the mean annual difference between $\hat{f}_t$ and f_t for $t \geq 2005$. For χ_t, we instead measure the percent difference between $\hat{\chi}^> = \Sigma_{t\geq 2005}\,\hat{\chi}_t$ and $\chi^> = \Sigma_{t\geq 2005}\,\chi_t$; that is, $\delta(\%) = 100 \times (\hat{\chi}^> - \chi^>)/\chi^>$. Hence, the summary statistics δ and δ(%) represent the total impact of EU enlargement on the scientific integration and the international competitiveness of the EU in terms of scientific output.

For the 2004 entrants, the SCM results indicate a $\delta = 0.062$ decrease in f_t and a 9% decrease in χ_t relative to the counterfactual (alternative potential outcome). In other words, the positive values indicate that cross-border activity would have been higher had they not entered the EU. In the case of f_t, this corresponds to roughly a 14% decrease over average pre-enlargement levels of $f_t \approx 0.45$. The incumbent EU countries also suffered a 15% decrease in χ_t; however, we also calculated a marginally negative value ($\delta = -0.013$) for the per-publication rate f_t.

This discrepancy calls for a closer look at the statistical significance of the SCM results.

To estimate the likelihood of obtaining these results due to random statistical fluctuations, we also applied the SCM to each of the control countries individually. Figure 2 (C and D) summarizes the results of this "permutation test" (*25*) by showing the differences between the real and synthetic outcome curves after applying the SCM to each country in our analysis—entrant, incumbent, and the controls. For example, the red (blue) dash-dotted curve in Fig. 2C represents the difference $\delta_t = \hat{f}_t - f_t$ between the corresponding red (blue) curves in Fig. 1A. Similarly, the red (blue) dash-dotted curve in Fig. 2D represents the percent difference $\delta_t(\%) = 100(\hat{\chi}_t - \chi_t)/\chi_t$ between the red (blue) curves in Fig. 2B. The other curves represent the individual countries from the control set (see the legend in fig. S1 to identify individual countries by their color).

By comparing the magnitude of posttreatment difference among all permutations of control units, we are able to estimate the statistical significance of the observed SCM result for the entrants and incumbents. To operationalize this comparison, we define the posttreatment difference as the net difference, $\delta_{i,>} = \sum_{t \geq 2005} \delta_{i,t}$ or $\delta_{i,>}(\%) = \sum_{t \geq 2005} \delta_{i,t}(\%)$, over the 8-year period 2005–2012. Hence, under the null hypothesis that there should be no difference in $\delta_{i,>}$ or $\delta_{i,>}(\%)$ among the units, one can estimate the likelihood that the ordered configuration of SCM outcomes could be driven by chance alone. For the case of f_t, the largest $\delta_{i,>} = 0.54$ belongs to the EU entrants, which exceeds the value of the next two largest $\delta_{i,>}$values (for Colombia and China) by 16%. Thus, we can assign a false-positive rate of 1/24 = 0.042 to the observed configuration, meaning that it is rather unlikely, just by chance alone, that the SCM produced such a large effect for the entrants. For the case of χ_t, neither the entrants nor the incumbents have the largest net difference, as several peripheral European countries (Ukraine, Belarus, Russia, and Turkey) show an even larger difference between the real and counterfactual outcome, suggesting that the EU enlargement may have even affected external countries as well. As an additional robustness check, we again applied the permutation test, but we used a "placebo" enlargement year $t^* = 2002$ instead. In this case, fig. S1 (C and D) shows that the entrant and incumbent curves are not distinguished in any way among the control countries. Thus, we can conclude that 2004 is a significant intervention year and that we do not find evidence of preexisting factors that might have driven our result.

To further investigate variation in the SCM results due to disciplinary factors, we repeated our procedure for two disaggregated highly collaborative disciplines, biology and physics, with the latter being the most collaborative of all the subject areas analyzed. High collaboration rates in these two domains can be attributed to the increasing prevalence of large collaborations (*1*) representing multinational consortiums organized around grand scientific challenges, international facilities, and even transnational clinical trials (*3*). Figure S1 (A to D) shows the SCM estimates for these two subject areas, which can be compared directly to the results shown in Fig. 2 (A and B) calculated by pooling all the subject areas together. For the case of f_t, the results are consistent for each subject area, both in sign and in magnitude. For the case of χ_t, the signs of δ(%) are consistent; however, the magnitude for the biology entrants indicates a negligible difference, δ(%) = 1.2%. We leave the

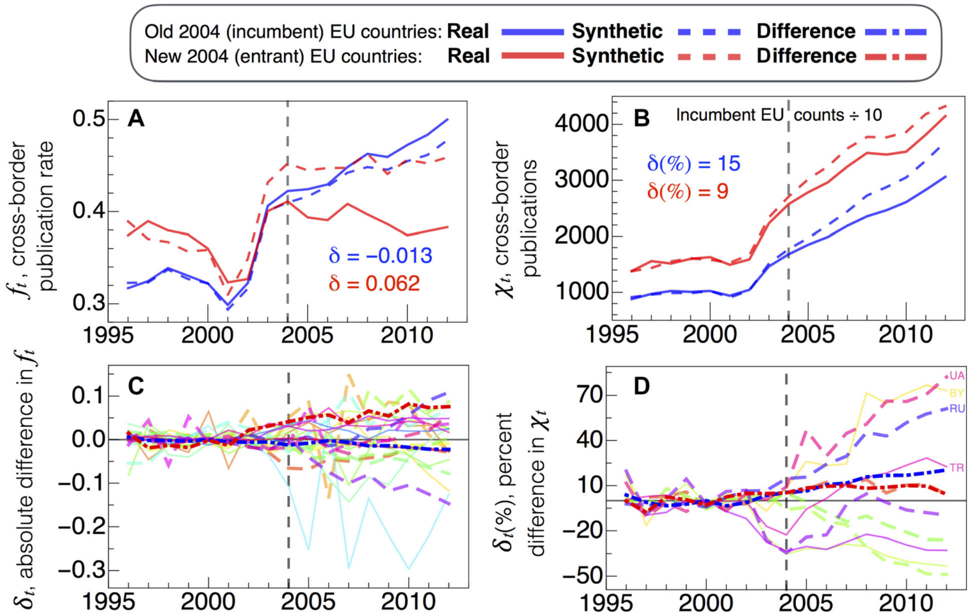

Fig. 2. Comparing synthetic (counterfactual) and real cross-border collaboration after the 2004 EU enlargement. (**A**) SCM results for the fraction f_t of cross-border publications and (**B**) the total number χ_t of cross-border publications. The solid curves represent the real data, whereas the dashed curves represent the estimates, $\hat{\chi}_t$ and $\hat{f}_t$, measuring the counterfactual scenario of no 2004 EU enlargement. Estimates are made using the SCM (*26*), implemented using a control group of 26 non-EU countries to best fit χ_t (f_t) for $t <$ 2004 and then to extrapolate $\hat{\chi}_t(\hat{f}_t)$ for $t \geq$ 2004 (see the Supplementary Materials). Note that the χ_t that represent the incumbent pre-2004 EU countries are divided by 10 to facilitate visualizing all the curves on the same scale. δ and δ(%) represent the difference between the real and synthetic curves after 2004, providing estimates of the "2004 EU entry" effect on cross-border European integration. (**C** and **D**) Estimation of the significance level of the SCM results using the permutation test (*25*). (C) For the intensive variable f_t, each curve represents the absolute difference $\hat{f}_t - f_t$; dash-dotted red and blue curves correspond to the entrant and incumbent EU curves in (A), respectively. Of the 24 curves, the (red) EU entrant curve has the largest positive net difference after 2004. (D) For the extensive variable χ_t, each curve represents the percent difference $100(\hat{\chi}_t - \chi_t)/\chi_t$. The four countries that exceed the average EU entrant curve are peripheral countries bordering the EU. (C and D) The additional colored curves correspond to the SCM difference calculated for each of the non-European control countries. Only the control country curves that passed SCM goodness-of-fit criteria for $t < t^*$ based on the mean squared error between the synthetic and real curve are shown (that is, to eliminate control countries with synthetic estimates that are either unreasonably noisy or not estimable).

exploration of discipline-dependent variations, as well as their potential causes, for future research.

To demonstrate the robustness of our results at the country level, we show in figs. S2 (for $f_{i,t}$) and S3 (for $\chi_{i,t}$) the analogous SCM time series calculated for the 10 individual 2004 entrant countries along with the two 2007 entrants. The country-level SCM results support the overall results observed at the aggregate level while further identifying some caveats. For example, CY is distinguished as the only country that benefited from the enlargement with both $\delta < 0$ and $\delta(\%) < 0$, most likely because of its strategic tax laws, which we will discuss later in terms of high-skilled mobility.

To summarize, we used the SCM to estimate the levels of European cross-border activity in the hypothetical counterfactual scenario corresponding to no 2004 EU enlargement. Before we move to a panel regression framework, which facilitates identifying the role and significance of individual covariates in explaining trends in $f_{i,t}$, we first introduce high-skilled mobility data, which are central to our main result connecting cross-border activity and mobility. Because the mobility data are only available for European countries, we were not able to incorporate it into the SCM analysis.

High-skilled mobility networks

We collected and analyzed publicly available official EU records to estimate the intra-European flow of high-skilled labor over the period 1997–2012. These data are derived from the formal request procedure of a certified professional in a given "host country" seeking cross-border validation of their degree certificate in a particular "destination country" and are collected and reported by each European country in the EU Single Market Regulated Professions Database (*14*). Thus, by aggregating these data across all countries, we constructed a mobility matrix, $M_{ij,t}$, capturing the total high-skilled mobility (head counts) from country i to country j in a given year t. The total across any given row i (or column j) of $M_{ij,t}$ gives the total outgoing $O^{+}_{i,t}$ mobility (or incoming $I^{+}_{j,t}$) from (or to) country i (j). Similarly, we define the net mobility matrix as $\Delta_{ij,t} \equiv M_{ij,t} - M_{ji,t}$, when there is positive net mobility from i to j ($M_{ij,t} > M_{ji,t}$) and $\Delta_{ij,t} \equiv 0$ otherwise. See the Supplementary Materials and figs. S4 to S11 for additional in-depth analysis of the aggregate (European), country, and dyadic (country-country) patterns of high-skilled mobility—before versus after the 2004 enlargement.

Figure 3 shows the mobility matrices before ($M_{ij,<}$) and after ($M_{ij,>}$) the 2004 enlargement. The patchy asymmetry of both matrices demonstrates the uneven geographic distribution of high-skilled mobility across Europe. Moreover, comparison of the mobility matrices and the corresponding $O^{+}_{i,t}$ and $I^{+}_{i,t}$ by country (see fig. S4) together demonstrate the drastic sevenfold increase in intra-European high-skilled mobility after the 2004 enlargement.

To account for the variation in the countries that constitute our mobility network, we also calculated country-level and dyadic (country-country) measures. For example, we calculated the Gini index $G^{in}_{j,t}$ ($G^{out}_{i,t}$) applied to the distribution of incoming (outgoing) mobility to (from) each country in each year to measure the dispersion of incoming (outgoing) mobility. We also measured the net mobility at the country level in two ways, first as the "absolute" net mobility $\Delta_{i,t} = O^{+}_{i,t} - I^{+}_{i,t}$ and second as the mobility polarization or "relative" net mobility, $B_{i,t} = \Delta_{i,t}/(O^{+}_{i,t} + I^{+}_{i,t})$. To illustrate the differences, we show in Fig. 3 the range of Δ_i and B_i calculated for each country, before and after 2004. Note that all of the mobility matrices in our analysis are shown with the countries ordered according to decreasing B_i (calculated over the entire period 1997–2012). For this reason, most of the entrant countries, with the exception of CZ and CY, appear in the upper left quadrant.

To illustrate additional structural information contained in the mobility matrices, we also produced several network visualizations derived from $M_{ij,t}$ and $\Delta_{ij,t}$. Figure 3 shows the minimum spanning tree (MST) representation, indicating that UK and DE were each at the core of the MST network before the enlargement, whereas the UK became the principal root vertex afterward. Figure S5 shows a circular network visualization of $M_{ij,t}$, before and after 2004, combining the directionality and magnitude of the mobility flow. Figure S6 shows that the community structure of the mobility matrix M_{ij} and that of the net mobility matrix Δ_{ij} are nearly identical. Together, we identify UK and DE as the two countries that have gained the most from high-skilled labor mobility: the former being the main "brain gain" hub for the Western European countries and the latter playing the same role for the Eastern European countries.

Figure 4 facilitates the direct comparison of $\Delta_{ij,<}$ and $\Delta_{ij,>}$, illustrating the drastic evolution of intra-European mobility imbalance (see also fig. S9). Of principal importance is the marked shift in the east-west mobility following the EU enlargement. Over the 8-year period 2005–2012, we observe 29% of the mobility to be from Eastern Europe (E; defined here as the 2004/2007 EU entrants) to Western Europe (W; defined here as the incumbent EU plus the four non-EU countries CH, IS, LI, and NO, which have notable trade, free movement, and other political agreements with the EU). By comparison, this percentage represents a significant increase, more than the 5% value observed from east to west ($E \rightarrow W$) over the 8-year period 1997–2004. Nevertheless, despite the drastic increase in the $E \rightarrow W$ mobility after 2004, the increasing weight of both the $E \rightarrow W$ and $W \rightarrow E$ mobility channels, relative to the intraregion mobility $E \rightarrow E$ and $W \rightarrow W$, represents progress toward brain circulation within Europe, which is fundamental for the competitiveness of its knowledge-based economies (*28*).

We use $B_{i,t}$ as a central explanatory variable in a panel regression model in the following section. Similar to f, mobility polarization is also an intensive variable, thereby facilitating the comparison of countries that range considerably in size. It is also a symmetric variable, centered around the value 0 corresponding to equal incoming and outgoing mobility, meaning that the sign of the corresponding coefficient in our regression model has a clear interpretation. By way of example, consider the mean $B_{i,t}$ values after 2004 for the incumbent and entrant countries, $\langle B^{\text{incumbent}}_{>} \rangle = 0.06$ and $\langle B^{\text{entrant}}_{>} \rangle = 0.53$, respectively. Comparatively, the countries with significant net immigration ($B_{i,>} \leq -0.5$) were CY, LU, and UK, whereas PT, GR, MT, HU, SK, BG, PL, RO, EE, and LT were the countries with the largest relative levels of emigration ($B_{i,>} \geq 0.5$); for further comparison of $B_{i,<}$ and $B_{i,>}$, see fig. S10A. Thus, combined with additional extensive (for example, $O^{+}_{i,t}$ and $I^{+}_{j,t}$) and intensive (for example, $G^{in}_{j,t}$ and $G^{out}_{i,t}$) mobility covariates, these variables provide an additional level of variation among the entrants and incumbent EU members.

Panel data regression model

The dependent variable in our model is $f^{s}_{i,t}$—the fraction of publications from country i in subject area s in year t that involve cross-border collaboration. We modeled this variable using a panel regression including country fixed effects ($\beta_{i,0}$) to control for time-invariant, country-level characteristics (for example, national language and geography). We also included a year variable to control for the overall increasing trend in cross-border collaboration. For example, the sharp increase in $f_{i,t}$ around 2002, within the EU and abroad, may stem from the 6th EU Framework Programme, which was the first to broadly include specific international collaboration criteria in its funding schemes.

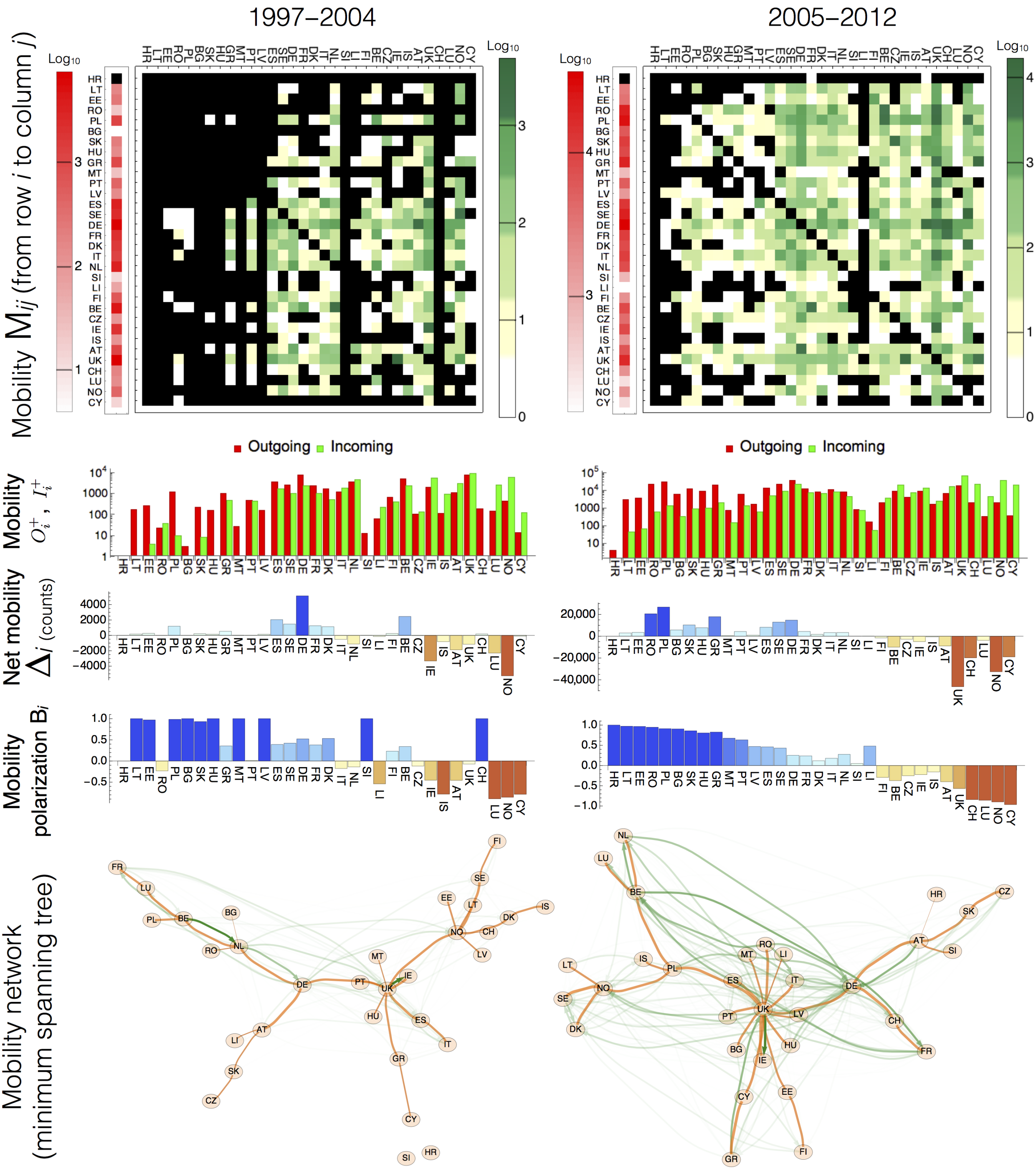

Fig. 3. High-skilled mobility before and after the 2004 enlargement. (**Top**) Mobility matrices (M_{ij}) showing the total mobility (head counts) from country i to j, with black cells indicating 0 observations. The red color scale to the left of each M_{ij} represents $\log_{10}O_i^+$, with black cells indicating $\Delta_i < 4$ for 1997–2004 and $\Delta_i < 155$ for 2005–2012; the green color scale indicates $\log_{10}M_{ij}$ and is split into six equally spaced regimes in logarithmic scale. (**Middle**) Aggregate mobility by country: total outgoing O_i^+, incoming I_i^+, net mobility $\Delta_i = O_i^+ - I_i^+$, and mobility polarization $B_i = (O_i^+ - I_i^+)/(O_i^+ + I_i^+)$. (**Bottom**) MST representation of the mobility networks indicated by the orange links, with green links providing an overlay of the non-MST links. The thickness and opacity of links are nonlinearly related to $\log_{10}M_{ij}$ so that only the most prominent links are visible; color values are not comparable between the two time periods.

In addition to the variables included in the SCM analysis (scientific productivity and impact, R&D investment, and GDPpc), we also include controls for publication subject area and additional covariates that control for cross-border activity, namely, mobility. In compact form, highlighting the two most important explanatory variables, our linear panel model is given by

$$f^{s}_{i,t} = \beta_T T_{EU,i,t} + \beta_B B_{i,t} + \vec{\beta} \cdot \vec{x}_{s,i,t} + \beta_{i,0} + \varepsilon_{i,t} \qquad (1)$$

The binary variable $T_{EU,i,t}$ represents a country's EU membership status (equal to 1 if i is an EU member in year t and 0 otherwise), so that a 0 → 1 shift captures the "EU entry effect" quantified by β_T. The coefficient β_B captures the linear response of $f^{s}_{i,t}$ depending on the inward or outward polarization of mobility. Finally, the inner product $\vec{\beta} \cdot \vec{x}_{s,i,t}$ represents the rest of the model control variables, and $\varepsilon_{i,t}$ is the white noise capturing country-level shocks. Further details on the high-skilled mobility data and the total migration (high skilled + low skilled) data can be found in the Supplementary Materials; for the full specifications of the panel model, see eq. S4.

Within this DiD framework, which comprises the relative comparison of a control and treatment group before and after the EU enlargement, we ran two models to test the significance of β_T and β_B, the results of which are summarized in Fig. 5. In both models, we leverage the fact that the enlargement occurred in two stages: 10 countries entered in 2004, and another 2 countries entered in 2007. In the first model, we use the incumbent EU countries as the baseline for comparison, and thus, β_T represents the change in $f^{s}_{i,t}$ relative to countries already within the EU. In the second model, we use the delayed 2007 entry of BG and RO to provide a second estimate of the EU entry effect, whereby the baseline for comparison in this case are these two delayed entrants. Thus, in the second model, β_T represents the change in $f^{s}_{i,t}$ relative to other non-EU countries whose membership was under consideration.

Starting with the first model, where we compare the entrants and incumbent groups, we find that the entrant countries suffered a decrease in $f_{i,t}$ upon entry into the EU ($\beta_T = -0.058 \pm 0.019$, $P = 0.004$). Moreover, the divergence in $f_{i,t}$ between the entrant and incumbent EU countries was further exacerbated by mobility polarization ($\beta_B = -0.043 \pm 0.013$, $P = 0.002$). In relative terms, comparing the corresponding standardized β coefficients $\hat{\beta}_T$ and $\hat{\beta}_B$, we conclude that the EU entry effect is roughly twice as large as the brain drain effect (see table S1, column 2). This β_T value is similar in magnitude to the counterfactual difference $\delta = 0.062$ produced by the SCM, pointing to the consistency of these two methods.

In all, the net difference in $f_{i,t}$ explained by these two effects is $-0.058 + (-0.043) \times ((\langle B^{\text{new EU}}_{>}\rangle - \langle B^{\text{old EU}}_{>}\rangle)) = -0.078$. The actual DiD in the mean collaboration rates before and after is $(\langle f^{\text{new EU}}_{>}\rangle - \langle f^{\text{old EU}}_{>}\rangle) - (\langle f^{\text{new EU}}_{<}\rangle - \langle f^{\text{old EU}}_{<}\rangle) = -0.085$ [calculated from the mean f_t curves in fig. S4 (A and B) for the 14 s]. Thus, we estimate that T_{EU} and $B_{i,t}$ explain

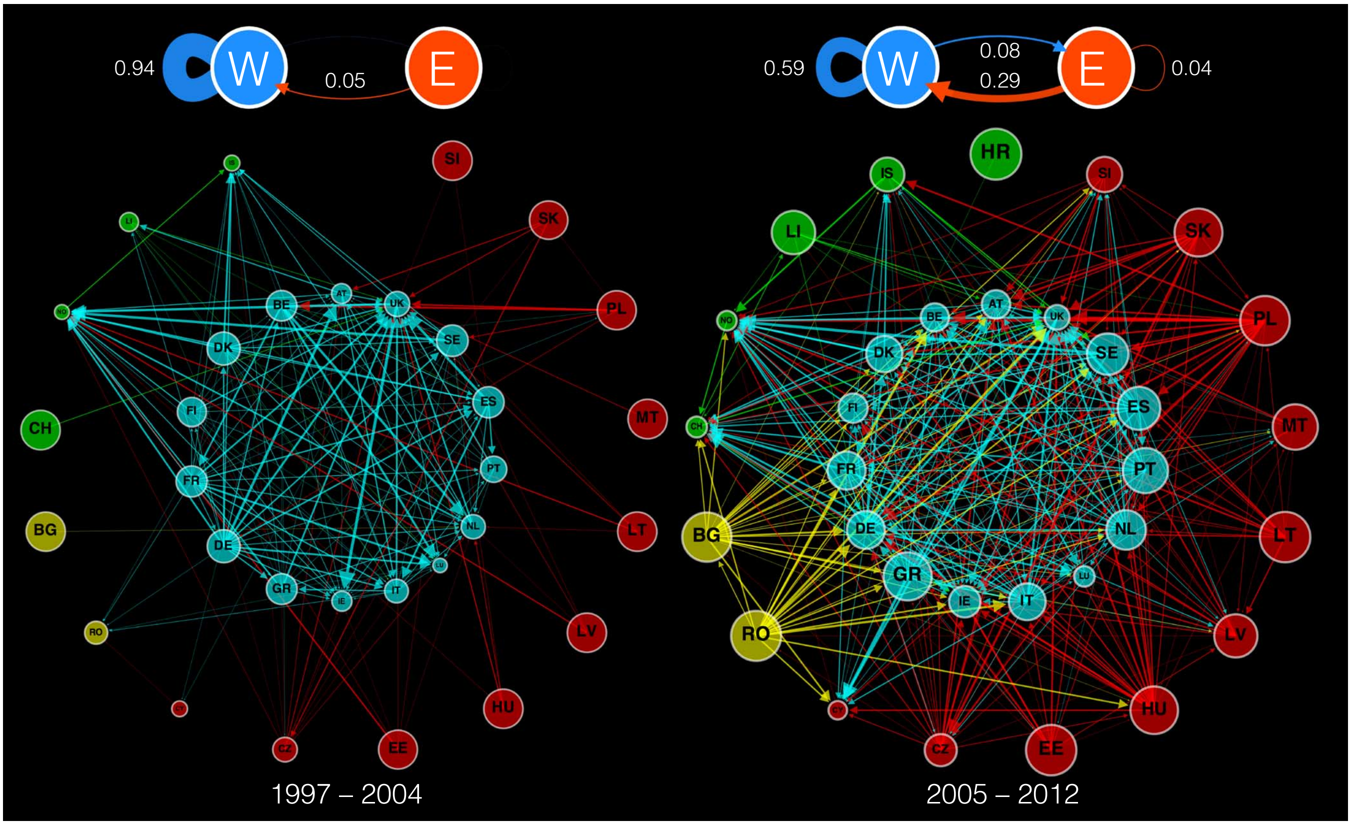

Fig. 4. High-skilled net mobility networks (Δ_{ij}), before and after the 2004 EU enlargement. (**Top**) Mobility between the 2004/2007 entrant countries ("E") and the rest of the incumbent European countries ("W"). The networks in each period are calculated from a total of 43,075 head counts (1997–2004) and 272,813 head counts (2005–2012), respectively. Link thickness represents the fraction of the total mobility. (**Bottom**) Node color represents EU entry year group ($g_{EU,i}$); node size is proportional to the mobility polarization, $1 + B_i$ (with larger values indicating larger mobility out of country i); link thickness is proportional to $\log(|\Delta_{ij}|)^2$ between countries i and j, with the arrow pointing in the direction of the net flow and link color corresponding to the source node. The size/thickness scales used for both networks are the same, facilitating direct comparison.

roughly 92% of the European east-west divergence in f_t over the period of analysis.

We report the full list of parameter estimates in the first three columns of table S1, which also contains partial model results in the additional columns, together demonstrating the robustness of our main results. Also, fig. S7 shows the results of our panel model including interaction terms between year and country group, demonstrating that our results cannot be explained by preexisting confounding factors (that is, before 2004) among the new EU entrant countries. This result is further supported by the null results of our placebo test for the SCM model shown in fig. S1 (C and D), which we implemented using the premature enlargement year $t^* = 2002$.

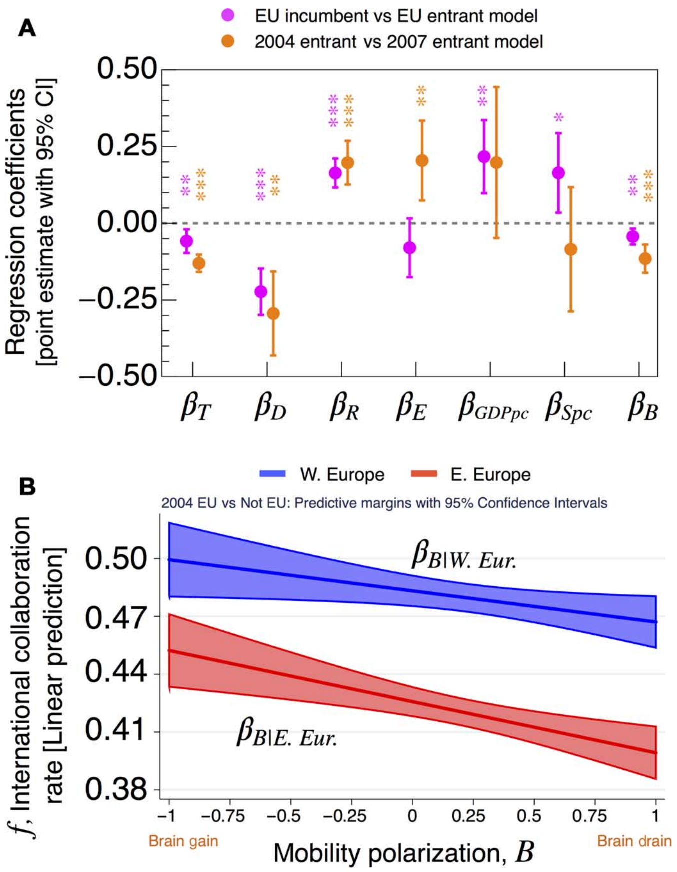

Fig. 5. The impact of EU enlargement and mobility on cross-border collaboration. (**A**) Point estimates with 95% confidence intervals. Two variables are particularly important to our analysis: (i) β_T captures the interaction between dummy values for before/after 2004 and a country's EU membership status—that is, the impact of EU entry; (ii) β_B captures the variation due to mobility polarization, $B_{i,t}$. We ran two regression models, each with a different baseline set of countries to demonstrate the robustness of our results. In the first model (magenta data, "full model"), the incumbent EU members serve as the baseline because their EU membership status does not change over the period of the analysis ($N_{obs.} = 4494$, adjusted $R^2 = 0.66$, and $N_c = 31$ countries). In the second model (orange data, three-period model), we used the 2007 entrants (BG and RO) as the baseline comparison for the 2004 entrants over three periods from 2001 to 2006 ($N_{obs.} = 504$, adjusted $R^2 = 0.60$, and $N_c = 12$ countries). See eq. S4 for the model specification and table S1 for the full set of controls, as well as other partial models demonstrating robustness of our results. Parameters are estimated using country fixed effects and robust SEs implemented by the Huber/White/sandwich estimator, which accounts for cross-sectional heteroscedasticity and within-panel (serial) correlation. As a visual aid, asterisks indicate the level of significance for each coefficient estimate: $^*P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$. (**B**) The marginal effect of $B_{i,t}$ on $f_{i,t}$, calculated using an interaction term between EU membership status and $B_{i,t}$. The main results of this model are two partial coefficients: $\beta_{B|E.Eur.}$ for the entrants and $\beta_{B|W.Eur.}$ for the other Western countries. Holding all other covariates at their mean value, comparison of the marginal linear predictions indicates that a country in the Western European group with $B = 1$ still has a higher expected level of international collaboration than a country from the Eastern European group with $B = -1$. Shaded interval indicates the 95% confidence interval calculated using the δ method.

In the second model, we restricted the number of years included in our analysis to the 6-year period 2001–2006 and consider only the 12 2004 and 2007 entrant countries. Over this subperiod, we define the treatment group as the 2004 entrants and the control group as the 2007 entrants, because the countries in the latter group had not entered the EU by the end of 2006. Hence, this model represents a more traditional treatment framework because all countries are initially not EU members, whereas for the final period (2005–2006), RO and BG were still waiting for their membership approval. As in the first model, we find that the effect of EU entry is negative ($\beta_T = -0.13$; $P < 0.001$), indicating that the 10 2004 entrants suffered a significant decrease in $f^s_{i,t}$ relative to the future entrants (BG and RO) that remained outside the EU in 2005–2006. To put this in perspective, because the average f value during 2002–2004 for the EU entrants is roughly 0.45 (see fig. S4B), then the reduction due to β_T represents roughly a $(-0.13/0.45) \times 100 = -29$ % overall effect. Moreover, the impact of emigration is also negative ($\beta_B = -0.115$; $P < 0.001$), which is consistent in sign with the estimate from the first model. The full list of parameter estimates is shown in the final two columns of table S1 denoted as the "three-period model (G)."

The most important parameter estimates are summarized in Fig. 5A for both models. The positive relation between cross-border activity and a country's scientific reputation ($\beta_R > 0$) was significant in both models. Also, the country-level economic covariates—government expenditure in R&D (β_E), per-capita GDP (β_{GDPpc}), and per-capita researchers (β_{Spc})—also played significant roles in explaining the variation in f. For example, in the second model (which only considers EU entrants), increased levels of government R&D expenditure were positively related to levels of cross-border collaboration. This relation was not significant in the first model, likely due to saturation in the impact of R&D expenditure on cross-border embeddedness. Additional covariates that we were unable to include in our model that also may contribute to the divergence between Eastern and Western Europe cross-border collaboration rates are inequality in R&D funding within the EU Framework Programmes, institutions, and the location of central scientific facilities (*29*, *30*).

To estimate the partial effect of mobility on the Eastern versus the Western European countries, we also ran the first model including an interaction term between $B_{i,t}$ and a binary dummy variable equal to 1 if a country was a member of the EU in 2003 (Western) and 0 otherwise (Eastern). Hence, this model produces a mobility parameter estimate for each region: $\beta_{B|W.Eur.} = -0.0161 \pm 0.0073$ ($P = 0.029$) and $\beta_{B|E.Eur.} = -0.0265 \pm 0.074$ ($P < 0.001$). Figure 5B shows the marginal effects of mobility on $f_{i,t}$ by region, with all other covariates evaluated at their mean values. Because B_i shifts from $B_i < 0$ (net immigration) to $B_i > 0$ (net emigration), there is a significant decline in the international collaboration rate for each country group, with Eastern European countries showing lower levels of $f_{i,t}$ and a slightly more negative marginal effect (the difference between the coefficients $\beta_{B|E.Eur.} - \beta_{B|W.Eur.} = -0.0104 \pm 0.0061$; $P = 0.088$).

We conclude by considering potential limitations and sources of error in our analysis. First, we are limited to using SCImago's definition of cross-border collaboration rate because we lack the publication-level microdata that are necessary to perform a sensitivity analysis of different affiliation counting schemes. Along these lines, we are also unable to account for variations across authors and across time in the number and type of affiliations per author, which would be difficult even with perfectly annotated author byline metadata. In all, because SCImago data are based on the full Scopus publication corpus, comprising millions of publications from numerous subject areas, we assume that there are no substantial country-level biases induced by their cross-border counting scheme and that year-level statistical errors affect all countries equally.

Second, the high-skilled mobility data are limited to intra-European flow; thus, we were unable to incorporate it into the SCM model analysis, which uses a global set of countries as the control set. Instead, our panel regression approach uses the incumbent EU members as the control set in the first panel model, unlike other studies, which use the rest of the world as the control (*8*, *9*).

Third, the EU mobility data are constructed from official records tracking certificate-based professions, such as law, health, business, and education. As a result, these data do not incorporate the mobility of publishing academics (Academia has its own longstanding system of certifying Ph.D. training and credentials based on the evaluation of publication output and institutional reputation; thus, unfortunately, there was no need for the EU to track this profession). To our knowledge, there is no comprehensive publication database that would provide accurate coverage of academic mobility across discipline and time, namely, due to the author name disambiguation problem and the difficulty in directly linking author names with author affiliations in author byline metadata. Thus, we must assume that the mobility patterns of academics are correlated, and thus substitutable, by high-skilled labor mobility patterns. Because mobility is likely related to underlying patterns of collaboration, this assumption may not apply to all disciplines to the same degree, because collaboration itself can vary widely by discipline. For this reason, we included subject area dummies in our panel regression models.

Fourth, we encountered difficulty in using the SCM to estimate the extensive variable $\hat{\chi}_t$ because of the skew in the size distribution of scientific output at the country level. As a result, the SCM was unable to obtain a suitable set of matches for the larger countries in our permutation test across the control countries, and thus, the SCM failed to produce $\hat{\chi}_t$ for the largest countries. Thus, the largest countries within the control set (for example, United States and Japan) contributed differently for each of the SCM estimates. Nevertheless, because neither the entrant nor the incumbent unit was within this size regime, the results for $\hat{f}_t$ and $\hat{\chi}_t$ are largely consistent and are not driven by overfitting due to the largest countries. In all, these considerations point to several open avenues for future research to identify the microlevel mechanisms linking mobility and cross-border collaboration.

DISCUSSION

In summary, our analysis reveals the counterintuitive decrease in cross-border activity by the new member states following their entry into the EU. That is, despite gaining access to EU resources incentivizing cross-border integration, we find that both the number of cross-border publications and rate of cross-border collaboration would have increased for the Western entrant countries had they not joined the EU. Our results explain the divergence in the cross-border collaboration rates for Eastern and Western Europe according to two complementary factors: (i) the regional difference in the impact of emigration ($\beta_{B|E.Eur.} < \beta_{B|W.Eur.}$) combined with (ii) higher levels of emigration among Eastern European countries, principally in the direction of east to west, which markedly increased following the 2004 enlargement.

It is important for EU policy-makers to consider the possible unintended consequences of EU labor market integration, especially considering the EU goals for a unified industrial and academic R&D innovation system (*6*, *31*, *32*). When a researcher not only moves abroad but also brings his or her international links along, this represents a loss of social capital—in addition to human capital and tacit knowledge (*33*, *34*)—that may further reduce the potential for knowledge spillovers across countries. Hence, this net flow of high-skilled labor to the large GDPpc countries (UK, CH, and NO) may negatively affect the convergence of human and technological capital within Europe (*21*), especially when considering the long-term impacts of losing elite scientists (*20*).

However, the EU should be commended for implementing "twinning" and "teaming" policies within the Horizon 2020 framework to counter the divergence in scientific competitiveness and specialization between regions (*35*). It is important to highlight some of the positive effects associated with brain drain, such as increased educational incentives within the source country and positive network externalities on trade and technological adoption (*16*, *19*, *36*).

We identified a link between mobility and cross-border collaboration in science that is rather general and not necessarily specific to Europe. Despite the negative externalities we observed, we emphasize that the opportunities for talented researchers to study abroad are extremely important and are a key component of a globally "open" and "competitive" science system. However, concerning the development of a competitive and sustainable ERA, there should be a concerted effort to address long-term migration from $E \rightarrow W$, so that a "brain recovery" follows organically from "brain circulation" (*17*, *22*, *28*, *37*, *38*), thereby fostering the right conditions for home-return knowledge transfer (*39*). The starting point is within the existing framework of mobility fellowships (for example, Marie Curie and other cross-border fellowships), possibly via implementing stronger incentives and criteria for home-country return and encouraging countries to implement profession-specific strategies for maintaining ties with their high-skilled expatriates (*13*).

MATERIALS AND METHODS

Countries analyzed

We analyzed 32 European countries over the 17-year period 1996–2012: Austria (AT), Belgium (BE), Bulgaria (BG), Croatia (HR), Cyprus (CY), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (GR), Hungary (HU), Iceland (IS), Ireland (IE), Italy (IT), Latvia (LV), Liechtenstein (LI), Lithuania (LT), Luxembourg (LU), Malta (MT), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Slovakia (SK), Slovenia (SI), Spain (ES), Sweden (SE), Switzerland (CH), and United Kingdom (UK). These countries can be grouped according to EU entry year: $g_{EU,i} = 1$ if existing EU member in 2004, $g_{EU,i} = 2$ if part of the 2004 EU enlargement (CY, CZ, EE, HU, LT, LV, MT, PL, SK, and SI), $g_{EU,i} = 3$ if part of the 2007 EU enlargement (BG and RO), and $g_{EU,i} = 4$ (CH, HR, IS, LI, and NO) if not part of the EU as of the end of 2012, corresponding to the final year of our analysis.

Brief description of the four data resources (see the Supplementary Materials for more details)

Publication data.

We downloaded comprehensive publication data from SCImago Journal and Country rank (*40*), which is calculated using comprehensive

Scopus data. From this data repository, we gathered four time series for each country i and each subject area s: (i) the total number of publications, $D^s_{i,t}$; (ii) the total number of citations received in year t, $C^s_{i,t}$; (iii) the fraction of publications involving international collaboration, $f^s_{i,t}$; and (iv) the total number of publications involving international collaboration $\chi^s_{i,t} = f^s_{i,t} D^s_{i,t}$; fig. S4 (A and B) shows the rate of cross-border publication $f^{All}_{i,t}$, combined across all subject areas ("All"), for the 32 EU countries over the period 1996–2012.

We analyzed 14 subject areas (indexed by s): "agricultural and biological sciences" (1100); "biochemistry, genetics, and molecular biology" (1300); "business management and accounting" (1400); "chemical engineering" (1500); "chemistry" (1600); "computer science" (1700); "decision sciences" (1800); "energy" (2100); "engineering" (2200); "environmental science" (2300); "materials science" (2500); "medicine" (2700); "pharmacology, toxicology, and pharmaceutics" (3000); and "physics and astronomy" (3100).

To account for the censoring bias associated with the measurement of citations (that is, publications from recent years have had less time to accrue citations than older publications), we normalized $C^s_{i,t}$ within s and t according to the logarithmic transform, $R^s_{i,t} \equiv (\ln C^s_{i,t} - \langle \ln C^s_{i,t} \rangle)/\sigma[\ln C^s_{i,t}]$, where $\langle ... \rangle$ and $\sigma[...]$ are the mean and SD calculated within each s and t group, respectively. Thus, $R^s_{i,t}$ measures the scientific reputation of country i in subject area s in year t. Moreover, $R^s_{i,t}$ is a time-independent and discipline-independent citation measure that has been shown to be closely distributed according to the Normal(0, 1) baseline distribution (*23*). Hence, $R^s_{i,t}$ is comparable across both s and t, being independent of the disciplinary and censoring bias that are problematic in the comparison of raw citation counts.

Different cross-border counting schemes could be used to define $f^s_{i,t}$, varying in how the share of credit for each cross-border publication is distributed to the affiliated countries. Although we were able to find three consistent verbal descriptions of the counting scheme on the SCImago website describing how $f^s_{i,t}$ is calculated, for example, "the ratio of a journal's documents signed by researchers from more than one country, that is, including more than one country address," we failed to find a precise mathematical description.

To clarify the counting scheme underlying the SCImago data for $f^s_{i,t}$ used in our study, which could conceivably affect the interpretation of our results, we performed a large-scale analysis of article-level data using data from the American Physical Society (APS). This data set is openly available in well-documented XML data files, which record article-level author byline data for more than a hundred years of articles from the *Physical Review* journal family (*Physical Review A*, *Physical Review B*, *Physical Review C*, *Physical Review D*, *Physical Review E*, *Physical Review Letters*, and *Reviews of Modern Physics*). This is the only large, open database we are aware of with comprehensive author affiliation information, which we used to geolocate the individual articles at the country level by using string matches for country names, ISO2 and ISO3 country codes; we were able to assign countries to more than 99.5% of all articles analyzed over the period 1997–2009.

Thus, we used this publication-level APS data to see what kind of counting scheme best reproduces the SCImago data for the "physics and astronomy (s = 3100)" category, that is, $f^{s=3100}_{i,t}$. First, we counted the total number of articles, $A^{APS}_{i,t}$, for each country i and year t. For each article p in the APS data, we also counted the total number of countries, n_p, among the affiliations. We then used two counting schemes to assign a cross-border weight $w_{i,p}$ to each p, calculating each county's total weight by summing across all p for a given t, given by $K^{APS}_{i,t} = \sum_p w_{i,p}$. Independent of the counting scheme, we assigned articles with $n_p = 1$ the weight $w_{i,p} = 0$. However, for articles with $n_p > 1$, we attributed a nonzero weight $w_{i,p}$ depending on the scheme: In method (i), we assigned the n_p-independent weight $w_{i,p} = 1$; in method (ii), we assigned the equipartition weight $w_{i,p} = 1/n_p$, thereby discounting the weight from (i) by the total number of countries involved, which better accounts for the possibility that n_p is growing over time. Then, for each counting scheme, country, and year, we calculated $f^{Est.}_{i,t} = K^{APS}_{i,t}/A^{APS}_{i,t}$.

Figure S12 shows the results of our counting scheme test, indicating that the n_p-independent weighting scheme is the one used by SCImago. Namely, the mean $f^{Est.}_{i,t}$ value (0.70) for the homogenous counting scheme was much closer to the mean $f^{s=3100}_{i,t}$value (0.58), whereas the mean value for the equal-share counting scheme (0.28) was significantly smaller, by nearly a factor of 2. However, despite the first method performing better, the agreement between $f^{Est.}_{i,t}$ and $f^{s=3100}_{i,t}$ suffered from a systematic deviation arising from two main sources of error. First, there was a consistent deviation, $f^{s=3100}_{i,t} < f^{Est.}_{i,t}$, likely arising from the international prestige of the PR journals, which thus attracted publications that are more international on average. We tested the difference between the ranked data using the Kendall τ test, which rejected the null hypothesis that the data are independent ($P < 10^{-21}$); the Spearman rank test and the Kolmogorov-Smirnov test (using mean-centered distributions) also confirmed the statistical relation between the data at the same level of significance. Second, because the noise levels in the calculation of $f^{Est.}_{i,t}$ increased for countries with lower publication rates, we attributed a second source of error to sample size fluctuations. In other words, the PR journals analyzed here represent just seven of the numerous journals categorized within SCImago's physics and astronomy classification.

R&D investment data.

As controls for country investment in R&D, which are, for example, related to the level of internationalization of higher educational institutions (*13*), we used researcher population, government spending, and GDP data from the World Bank data repository (*41*). In particular, for each country, we used government expenditure on R&D data ($E_{i,t}$), per-capita GDP data ($GDPpc_{i,t}$), and per-capita researcher data ($Spc_{i,t}$). See the Supplementary Materials for further details.

Mobility data (EU high-skilled).

As a proxy for trends in intra-European researcher mobility, we used official EU Commission "professionals moving abroad (establishment)" data from Single Market Regulated Professions Database (*14*). This database tracks the number of (high-skilled) professionals who obtained official certification in a given country of qualification (source country) and then applied for official recognition of their professional certification in a particular host country (destination country).

The data are grouped into 13 periods indexed here by t = 1...13 corresponding to 1997/1998, 1999/2000, 2001/2002, 2003/2004, 2005/2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, and 2014. We did not include the final 2 years of data in our analysis because the mobility data were either incomplete or still being updated and because the World Bank R&D data are incomplete for many countries after 2012. It is also worth explicitly stating that we divided the mobility head count variables for periods in $t \leq 2006$ by a factor of 2 so that these count values refer to mean annual rates. Hence, to combine observations across these three data sets, we also aggregated the count data for publications and country-level economic indicators across the specified 2-year periods and then divided the counts by a factor of 2, resulting in 2-year annual averages.

Thus, for each year period t, we recorded $M_{ij,t}$, the total high-skilled mobility ("total positive decisions") from country i ("country of qualification") to country j ("host country"). In all, the total mobility (head counts) for a given time period x, $M_x = \sum_{ij} M_{ij,x}$, is 315,888 (1997–2012), 43,075 (1997–2004), and 272,813 (2005–2012). We also recorded the number of "total negative decisions," $N_{ij,t}$, corresponding to those applications that were denied (for a variety of reasons). The total number of negative decisions by period is 24,046 (1997–2012), 4734 (1997–2004), and 19,312 (2005–2012), representing roughly 7% of the total (positive and negative) decisions made. See the Supplementary Materials for further details and specification of the mobility measures that we derived from these longitudinal country-level data.

Total international migration data.

We used data from Abel and Sander (*15*) to capture the net patterns of migration from country to country (that is, including high- and low-skilled labor) over three 5-year periods, $\tau = 1$ (1995–2000), $\tau = 2$ (2000–2005), and $\tau = 3$ (2005–2010). These data were included in our regression models so that significant coefficients related to high-skilled mobility are in excess of gross migration patterns.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/4/e1602232/DC1
Supplementary Materials and Methods
fig. S1. Supplementary SCM results.
fig. S2. SCM: cross-border publication rate *f*.
fig. S3. SCM: total cross-border publications χ.
fig. S4. International collaboration rates and high-skilled labor mobility.
fig. S5. Country-country mobility networks before and after the 2004 enlargement.
fig. S6. Community structure of the high-skilled mobility networks.
fig. S7. Consistency check for the significance of the EU entry effect.
fig. S8. Comparison of high-skilled to total migration by country-country pair.
fig. S9. Net flow of high-skilled labor: before and after the 2004 enlargement.
fig. S10. Mobility success rates by country: before and after the 2004 enlargement.
fig. S11. Mobility success rates by country-country pair: before and after the 2004 enlargement.
fig. S12. Validation of the SCImago cross-border counting scheme.
table S1. Full panel regression model results.
References (*43*–*52*)

REFERENCES AND NOTES

1. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
2. B. F. Jones, S. Wuchty, B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008).
3. A. M. Petersen, I. Pavlidis, I. Semendeferi, A quantitative perspective on ethics in large team science. *Sci. Eng. Ethics* **20**, 923–945 (2014).
4. S. Milojevic, Principles of scientific research team formation and evolution. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3984–3989 (2014).
5. D. Hsiehchen, M. Espinoza, A. Hsieh, Multinational teams and diseconomies of scale in collaborative research. *Sci. Adv.* **1**, e1500211 (2015).
6. P. Boyle, Policy: A single market for European research. *Nature* **501**, 157–158 (2013).
7. J. Hoekman, K. Frenken, F. van Oort, The geography of collaborative knowledge production in Europe. *Ann. Reg. Sci.* **43**, 721–738 (2009).
8. A. Chessa, A. Morescalchi, F. Pammolli, O. Penner, A. M. Petersen, M. Riccaboni, Is Europe evolving toward an integrated research area? *Science* **339**, 650–651 (2013).
9. A. Morescalchi, F. Pammolli, O. Penner, A. M. Petersen, M. Riccaboni, The evolution of networks of innovators within and across borders: Evidence from patent data. *Res. Policy* **44**, 651–668 (2015).
10. H. Delanghe, U. Muldur, L. Soete, *European Science and Technology Policy: Towards Integration or Fragmentation?* (Edward Elgar, 2009).
11. T. E. Scherngell, *The Geography of Networks and R&D Collaborations* (Springer International Publishing, 2013).
12. A. E. Geuna, *Global Mobility of Research Scientists* (Academic Press, 2015).
13. B. Lepori, M. Seeber, A. Bonaccorsi, Competition for talent. Country and organizational-level effects in the internationalization of European higher education institutions. *Res. Policy* **44**, 789–802 (2015).
14. European Commission: The EU single market regulated professions database (professionals moving abroad); http://ec.europa.eu/growth/tools-databases/regprof/ [accessed August 2015].
15. G. J. Abel, N. Sander, Quantifying global international migration flows. *Science* **343**, 1520–1522 (2014).
16. M. Beine, F. Docquier, H. Rapoport, Brain drain and economic growth: Theory and evidence. *J. Dev. Econ.* **64**, 275–289 (2001).
17. L. Ackers, Moving people and knowledge: Scientific mobility in the European Union. International migration. *Int. Migr.* **43**, 99–131 (2005).
18. L. Ackers, B. Gill, *Moving People and Knowledge Scientific Mobility in an Enlarging European Union* (Edward Elgar, 2008).
19. J. Gibson, D. McKenzie, Eight questions about brain drain. *J. Econ. Perspect.* **25**, 107–128 (2011).
20. B. A. Weinberg, Developing science: Scientific performance and brain drains in the developing world. *J. Dev. Econ.* **95**, 95–104 (2011).
21. V. Grossmann, D. Stadelmann, Does international mobility of high-skilled workers aggravate between-country inequality? *J. Dev. Econ.* **95**, 88–94 (2011).
22. S. P. Kerr, W. Kerr, C. Özden, C. Parsons, Global talent flows. *J. Econ. Perspect.* **30**, 83–106 (2016).
23. A. M. Petersen, Quantifying the impact of weak, strong, and super ties in scientific careers. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4671–E4680 (2015).
24. M. Marx, D. Strumsky, L. Fleming, Mobility, skills, and the Michigan non-compete experiment. *Manage. Sci.* **55**, 875–889 (2009).
25. H. R. Varian, Causal inference in economics and marketing. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7310–7315 (2016).
26. A. Abadie, A. Diamond, J. Hainmueller, Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Am. Stat. Assoc.* **105**, 493–505 (2010).
27. D. B. Rubin, Causal inference using potential outcomes. *J. Am. Stat. Assoc.* **100**, 322–331 (2005).
28. A. M. Petersen, M. Puliga, High-skilled labour mobility in Europe before and after the 2004 enlargement. *J. R. Soc. Interface* **14**, 20170030 (2017).
29. J. Hoekman, T. Scherngell, K. Frenken, R. Tijssen, Acquisition of European research funds and its effect on international scientific collaboration. *J. Econ. Geogr.* **13**, 23–52 (2013).
30. A. Abbott, Q. Schiermeier, After the Berlin Wall: Central Europe up close. *Nature* **515**, 22–25 (2014).
31. M. Nedeva, M. Stampfer, From "Science in Europe" to "European Science". *Science* **336**, 982–983 (2012).
32. European Commission, Directorate-General for Research and Innovation, *European Research Area Progress Report 2013; European Research Area Facts and Figures 2013.* (Publications Office, 2013).
33. A. Agrawal, I. Cockburn, J. McHale, Gone but not forgotten: Labor flows, knowledge spillovers, and enduring social capital. *J. Econ. Geogr.* **6**, 571–591 (2006).
34. A. Agrawal, D. Kapur, J. McHale, A. Oettl, Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *J. Urban Econ.* **69**, 43–55 (2011).
35. European Commission Horizon 2020: Spreading excellence and widening participation; http://ec.europa.eu/programmes/horizon2020/en/h2020-section/spreading-excellence-and-widening-participation.
36. F. Docquier, H. Rapoport, Globalization, brain drain, and development. *J. Econ. Lit.* **50**, 681–730 (2012).
37. C. Dustmann, I. Fadlon, Y. Weiss, Return migration, human capital accumulation and the brain drain. *J. Dev. Econ.* **95**, 58–67 (2011).
38. T. Wiesel, Fellowships: Turning brain drain into brain circulation. *Nature* **510**, 213–214 (2014).
39. D. Wang, Activating crossborder brokerage: Interorganizational knowledge transfer through skilled return migration. *Admin. Sci. Quart.* **60**, 133–176 (2015).
40. SCImago: SJR SCImago Journal and Country Rank; www.scimagojr.com [accessed September 2015].
41. World Bank data sources; http://data.worldbank.org/indicator. [accessed August 2015].
42. A. Petersen, European high-skilled mobility data and Scientific publication & collaboration data (UC Merced, 2017); https://dx.doi.org/10.6071/M3RP49.
43. P. Brown, A. Green, H. Lauder, *High Skills: Globalization, Competitiveness, and Skill Formation* (Oxford Univ. Press, 2001).
44. H. F. Moed, M. Aisati, A. Plume, Studying scientific migration in Scopus. *Scientometrics* **94**, 929–942 (2013).
45. R. Van Noorden, Global mobility: Science on the move. *Nature* **490**, 326–329 (2012).
46. P. Deville, D. Wang, R. Sinatra, C. Song, V. D. Blondel, A.-L. Barabási, Career on the move: Geography, stratification, and scientific impact. *Sci. Rep.* **4**, 4770 (2014).
47. A. Abadie, J. Gardeazabal, The economic costs of conflict: A case study of the Basque country. *Am. Econ. Rev.* **93**, 113–132 (2003).
48. M. Kahanec, *International Handbook on the Economics of Migration*, A. F. Constant, K. F. Zimmermann, Eds. (Edward Elgar, 2013), chap. 7, pp. 137–152.
49. M. E. J. Newman, Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577–8582 (2006).

50. F. Giavazzi, G. Tabellini, Economic and political liberalizations. *J. Monet. Econ.* **52**, 1297–1330 (2005).

51. T. Persson, G. Tabellini, Democracy and development: The devil in the detail. *Am. Econ. Rev.* **96**, 319–324 (2006).

52. M. Krzywinski, J. Schein, İ. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, M. A. Marra, Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

Acknowledgments: We thank C. A. Mora, M. Onorato, P. Zacchia, M. Puliga, the Faculty of Economics at the Pontificia Universidad Javeriana, and two anonymous referees for helpful comments. **Funding:** We acknowledge funding from the National Research Program of Italy (PNR) project "CRISIS Lab." **Author contributions:** All authors conceived the research. O.A.D.A. and A.M.P. designed the analyses and conducted the analyses. A.M.P. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. All data used in this manuscript are openly available through references (*14*, *15*, *40*, *41*); please see reference (*42*) for curated and documented high-skilled mobility and cross-border collaboration data.

Submitted 13 September 2016
Accepted 9 February 2017
Published 12 April 2017
10.1126/sciadv.1602232

Citation: O. A. Doria Arrieta, F. Pammolli, A. M. Petersen, Quantifying the negative impact of brain drain on the integration of European science. *Sci. Adv.* **3**, e1602232 (2017).