INFORMATION ASYMMETRY IN THE SCIENTIFIC LABOR MARKET: THE ROLE OF RELATIONAL LINKS

Documents de travail GREDEG
GREDEG Working Papers Series

Pierre Boutros
Michele Pezzoni
Lionel Nesta
Sonia Paty

GREDEG WP No. 2026-14
https://ideas.repec.org/s/gre/wpaper.html

Information Asymmetry in the Scientific Labor Market: The Role of Relational Links

Pierre Boutros[*], Michele Pezzoni[†], Lionel Nesta[‡], Sonia Paty[§]

Abstract

This paper investigates the factors predicting the destination choice of mobile researchers. To do so, we use a unique dataset on researchers' mobility between labs within the largest French public research organization from 2012 to 2022. We find that relational links, namely citation and co-authorship links between mobile researchers and destination lab members, are among the strongest predictors of researchers' destination choices. Specifically, a citation link prior to mobility between a researcher and a lab is associated with a 3.7 percentage-point higher probability that the researcher chooses that lab as a destination, while a co-authorship link is associated with a 9.8 percentage-point higher probability. We argue that citation and co-authorship links are highly relevant because they serve as information channels to help address substantial information asymmetry between researchers and potential destination labs before mobility. We further find that citation links are better predictors of destination choice when the cognitive distance between the researcher and the lab is high, whereas co-authorship plays a stronger role when the cognitive distance is low. Finally, we find that other lab characteristics, such as the size, productivity, and funding availability, are less relevant to the destination choice.

JEL Codes: D83, J61, J24

Keywords: Mobile researchers; Internal labor markets; Information asymmetry; Destination choice

[*]Corresponding author. Université Côte d'Azur, CNRS, GREDEG (UMR 7321), France. Email: pierre.boutros@etu.unice.fr.

[†]Université Côte d'Azur, CNRS, GREDEG (UMR 7321), France and Observatoire des Sciences et Techniques, HCERES, France. Email: michele.pezzoni@univ-cotedazur.fr.

[‡]Université Côte d'Azur, CNRS, GREDEG (UMR 7321), France, OFCE, Sciences Po, France, and SKEMA Business School, France. Email: lionel.nesta@univ-cotedazur.fr.

[§]Université Lumière Lyon 2, CNRS, Université Jean Monnet Saint-Etienne, emlyon business school, GATE (UMR 5824), Lyon, France. Email: sonia.paty@cnrs.fr.

1 Introduction

In modern science, the discovery process increasingly relies on teamwork to acquire the knowledge needed to study complex scientific problems (Wuchty et al. 2007, Jones 2009). As a result, researchers and labs exchange knowledge primarily through the mobility of researchers who bring their accumulated knowledge stocks with them when they relocate (Zucker et al. 1998, Kaiser et al. 2015). Knowledge stocks are not limited to codified outcomes, such as scientific publications, but also include tacit components, such as non-codifiable scientific competencies, intellectual ability, commitment to work, teamwork skills, and problem-solving capacities (Nelson & Winter 1985). Researchers' mobility generates a scientific labor market in which researchers' and labs' knowledge stocks constitute their most valuable assets (Stephan 2012).

In this scientific labor market, the optimal researcher-lab match is the one that yields the highest innovative outcomes for both parties (Jovanovic 1979, Nooteboom et al. 2007). However, optimal matches do not always occur, and researchers may relocate to labs where their knowledge does not integrate well with the lab's knowledge stock, resulting in limited innovative outcomes. Despite the importance of studying the researchers' destination choices, previous literature has focused extensively on the outcomes of researchers' mobility after it occurs. From these studies, we know that mobility enhances researchers' productivity, broadens research networks, fosters knowledge exchange, and provides better access to resources (Levin & Stephan 1999, Ackers 2005, Edler et al. 2011, Franzoni et al. 2012), showing that an effective researcher-lab match is crucial for the scientific outcomes of both parties. However, we have little knowledge about how the destination choice occurs.

To fill this gap, we study researchers' destination choices, arguing that these choices are largely driven by information asymmetry stemming from researchers' and labs' imperfect information on each other's knowledge stocks. We also argue that researchers and labs attempt to overcome this information asymmetry problem by relying on relational links formed through scientific publications, as indicated by citation links, and temporary professional collaborations, as indicated by co-authorship links, which convey information about codified and tacit aspects of each other's knowledge stocks before mobility. Given the central role of knowledge stocks and the related information asymmetry problem, references to scientific publications (i.e., citation links) and temporary professional collaborations (i.e., co-authorship links) are expected to be among the strongest predictors of destination choice. In this paper, we aim to empirically test the latter statement by addressing the following question: *Are citation and co-authorship links prior to mobility strong predictors of researchers' destination choices?*

To answer this question, we use a unique dataset covering the mobility and destination choices of 2,033 researchers across labs within the *National Center for Scientific Research* (CNRS), the largest public research organization in France, from 2012 to 2022. Our empirical results are consistent with our theoretically grounded hypothesis that researchers and labs rely on citations to scientific works and co-authorships to address information asymmetry, as these are among the strongest predictors of destination choice. Specifically, we find that having citation links to a potential destination lab is associated with a 3.7 percentage-point higher probability of choosing that lab, and having a co-authorship link is associated with a 9.8 percentage-point higher probability of choosing that lab. Moreover, we show that the effectiveness of these information channels is largely moderated by the cognitive distance between the researcher and the potential destination lab. Consistent with our theory, we find that other lab characteristics, such as size, productivity, and funding availability, are less relevant. Finally, we find discipline heterogeneity. The effect of co-authorship links is mainly observed in STEM disciplines, whereas it does not appear to predict destination choices in humanities and social sciences.

The remainder of the paper is structured as follows. Section 2 outlines a review of the literature. Section 3 describes the data used in the analysis and illustrates the empirical strategy. Section 4 reports the results, and Section 5 concludes by summarizing the key findings.

2 Previous literature

Previous literature has studied researchers' mobility without focusing on the destination choice. Specifically, a first body of literature has focused on the determinants of mobility decisions, such as career incentives, institutional characteristics, or access to research resources (Azoulay et al. 2017, Verginer & Riccaboni 2021). A second body has examined mobility outcomes, showing that mobility can enhance researchers' productivity, broaden professional networks, foster knowledge exchange, and improve access to scientific resources (Levin & Stephan 1999, Ackers 2005, Edler et al. 2011, Franzoni et al. 2012).

Both these literature bodies paid little to no attention to how researchers choose their destination organization. The limited attention paid to the destination choice may be due to data constraints. Indeed, studying the destination choice requires detailed information not only on the researchers' origin and destination labs, but also on the full set of potential labs to which they could have moved. This information is typically unobserved because it is difficult to define a meaningful set of potential institutions to which a researcher can relocate (Azoulay et al. 2017). In this paper, we address this limitation by exploiting a unique empirical setting that allows us to construct, for each mobile researcher, a complete set of potential destination labs. A notable exception is De la Croix et al. (2024), who study the location decisions of scholars in the European academic market over the period 1000–1800 using a database of thousands of scholars. Their analysis shows that scholars tended to disproportionately relocate to the most prestigious universities, that those with greater human capital were particularly drawn to high-quality institutions, and that they tended to migrate over greater distances. Different from De la Croix et al. (2024), we focus on the micro-level determinants of destination choice within a modern public research organization.

The match between a researcher and a lab can be viewed as a process of knowledge combination between two parties (Kogut & Zander 1992, Weitzman 1998, Fleming 2001), where knowledge stocks are the most distinctive traits of individuals, i.e., researchers, and organizations, i.e., labs. In this context, researchers accumulate knowledge through education and work experience, while labs accumulate knowledge by recruiting new members (Nelson & Winter 1985, Jones 2009). Knowledge recombination is complete when researchers move to labs and recombine their knowledge stocks with those of lab members (Zucker et al. 1998, Stephan 2012). Therefore, it is crucial to understand the process that leads to the formation of researcher-lab matches and the subsequent recombination of their knowledge stocks (Stephan 2012, Baruffaldi & Landoni 2016).

One difficulty in identifying the most innovative researcher-lab matches is that knowledge stocks cannot be fully observed by researchers and labs before mobility, mainly because of the inherent challenges of observing the tacit aspects of each other's knowledge stocks (Cohen et al. 1990). Indeed, the tacit aspects of the knowledge stocks, such as non-codifiable scientific competencies, intellectual ability, commitment to work, access to resources, teamwork skills, and problem-solving capacities, become observable only through repeated interactions and collaboration (Nelson & Winter 1985, Zucker et al. 1998). On the contrary, information about codified aspects of knowledge stocks can, in principle, be collected from written documents, such as scientific articles, without direct contact or physical mobility (Jaffe et al. 1993). However, to interpret the knowledge embedded in codified documents effectively, researchers must have a basic understanding of the topic (Cowan et al. 2000). Indeed, when the cognitive distance between the reader and the codified document is large, the capacity to interpret the information included in the codified documents tends to be low (Nooteboom et al. 2007). As a result, documents allow researchers and labs to collect information about the codified aspects of each other's knowledge stocks only when their cognitive distance is low.

Ideally, researchers and labs should assess their respective knowledge stocks by exploring multiple potential researcher-lab matches and selecting the one that is the best fit between their knowledge stocks. However, this solution is not feasible because exploring each potential researcher-lab match through mobility is costly. Researchers face both private and professional relocation costs associated with mobility (Greenwood 1975), while labs bear the risks of long-term employment commitments (Lazear & Gibbs 2014). This situation is commonly referred to in the labor literature as employee-employer matching under conditions of imperfect information (Burdett & Coles 1999). In this literature, the quality of an employment relationship is typically not fully observable at the outset. Instead, employees and employers learn over time about the quality of the match (Jovanovic 1979, Mortensen & Pissarides 1994, Sengul 2017). As a result, information asymmetries are usually reduced through observable characteristics, such as contracts that describe the job content and the conditions offered by the organization, and interviews that assess the candidate's skills and competencies, and the candidate's resume or referrals (Spence 1978, Gürtzgen & Pohlan 2024).

Unlike most labor markets, where many factors are at play, the scientific labor market is characterized by the crucial role of individuals' and organizations' knowledge stocks as their most valuable asset (Stephan 2012, Baruffaldi & Landoni 2016). This is especially true in our empirical context, which examines researchers' destination choices following mobility between the labs of France's largest public research organization, i.e., the CNRS. In this internal labor market, wages, contracts, working conditions, and organizations' rules are largely fixed across all CNRS labs, making researchers' and labs' knowledge stocks important determinants of destination choices. Consequently, information asymmetries regarding the researcher-lab match mainly concern the content of their knowledge stocks.

In this paper, we argue that researchers and labs seek to reduce information asymmetries before mobility using two main channels. First, they collect information about each other's knowledge stocks from written documents, i.e., scientific publications (Jaffe et al. 1993, Narin et al. 1997). Indeed, scientific publications provide a paper trail of researchers' and labs' research activities, describing the evolution of their knowledge that can be codified. A second source of information comes from temporary professional collaborations, i.e., co-authorships. By working together on specific research projects, researchers observe each other's working practices, research abilities, their capacity to coordinate scientific activities, and other tacit aspects of their knowledge stocks (Katz & Martin 1997, Nakajima et al. 2010).

Given the importance of knowledge stocks in destination choices and the use of citations to scientific works and co-authorship as channels to reduce information asymmetries, we expect both citation and co-authorship links to be strong predictors of destination choice. However, we also expect that the effectiveness of these information channels depends on researchers' and labs' ability to interpret the information they convey. Indeed, differences in the cognitive distance between researchers and potential destination labs can influence how useful publications and prior collaborations are for assessing the potential researcher-lab match.

3 Data, model, and variables

3.1 Data

In this paper, we focus on CNRS, the principal public research organization in France and one of the world's leading research institutions.[1] It ranks among the top 15 research institutions worldwide.[2] CNRS offers tenured research positions with no compulsory teaching duties. It is organized into several thematic institutes, each dedicated to broad scientific domains (e.g., life sciences, physics, chemistry, environmental sciences, mathematics, humanities and social sciences, etc.). In fact, it is the only French organization active in all scientific disciplines. Established in 1939, CNRS is France's largest public research organization, employing over 33,000 personnel, including more than 11,000 permanent researchers distributed across more than 1,000 labs and research units co-operated in partnership with French universities, engineering schools, and other public research institutes such as Inserm (the National Institute of Health and Medical Research), CEA (the Atomic Energy Commission), or INRAE (the National Research Institute for Agriculture, Food and Environment).[3] Therefore, lab members typically consist of a mix of permanent

[1] More information about CNRS's structure can be found in Appendix D.

[2] https://www.nature.com/nature-index/research-leaders/2025/institution/all/all/global.

[3] https://www.cnrs.fr/en/le-cnrs/organisation.

CNRS researchers, university faculty members, postdoctoral fellows, PhD students, engineers, and technical staff. Internal mobility at CNRS is possible at any time and must be initiated by the researcher. The procedure requires the researcher to submit a scientific project, which is assessed by the directors of both the origin and destination labs.[4] Moreover, working conditions across CNRS labs are largely fixed. CNRS researchers hold the same type of contracts, receive a nationally pre-determined salary, and face similar research obligations regardless of their lab affiliation.

Our empirical analysis relies on three data sources. First, we use administrative records from CNRS[5] covering researchers between 2012 and 2022, which provide information on their recruitment dates, current affiliation lab, rank, gender, and scientific discipline. Second, we use the mobility records, which document researchers' mobility events, including the type (internal move, external departures, and retirement) and date of each move, as well as the destination lab if an internal move occurs. Finally, we collect bibliometric data by matching[6] researchers' names to author records in *OpenAlex* and retrieve information such as the publication year, co-author list, number of publications, and citations (Priem et al. 2022).

We start by filtering the administrative records to include only researchers recruited after 2012, ensuring that we have the complete employment history of the researchers at CNRS. Next, we filter the mobility records to only keep internal moves, that is, transfers from one CNRS lab to another in France. The sample includes 2,033 CNRS researchers, of whom 207 (10.2%) experienced at least one internal mobility between CNRS labs during our study period, i.e., 2012–2022. These correspond to 197 unique movers, with 10 researchers moving more than once. We focus on these 197 researchers, and for each mobile researcher, we list all CNRS labs in the same discipline as potential destinations,[7] dropping the labs that were inactive at the time of the move (i.e., labs with zero researchers). In other words, our dataset consists of all possible researcher-lab pairs for each mobility event, and among these pairs, only one represents each researcher's actual destination choice. As a result, we obtain 6,875 possible researcher-lab pairs. On average, each researcher chooses among 33 potential destination labs, with a minimum of 4 and a maximum of 90.

[4] https://www.cnrs.fr/sites/default/files/pdf/DRH2020-11_LDG-mobilite.pdf

[5] Researchers' administrative and mobility records were provided by DAPP-CNRS (Direction d'appui aux partenariats publics).

[6] In some cases, OpenAlex assigns multiple author IDs to the same researcher. We manually filter those to ensure that we are attributing the correct publications to every researcher in our dataset.

[7] We checked for interdisciplinary mobility and confirmed that 83% of all moves occurred within the same discipline. More details can be found in Appendix B.

3.2 Model and variables

We model the probability that researcher i chooses destination lab l among all potential labs as in Equation 1.

$$\Pr(\text{Choice}_{i,l} = 1) = \Lambda\Big(\beta_0 + \mathbf{B_\Gamma \Gamma_{i,l}} + \mathbf{B_L L_l} + \mathbf{B_R R_i} + \beta_\lambda \lambda_i\Big), \tag{1}$$

where Λ is the logistic link function. We define the destination choice $Choice_{i,l}$ as a dummy variable that equals one if lab l is the actual destination, and zero otherwise. Destination choice is conditioned upon three main vectors of explanatory variables: a vector $\mathbf{\Gamma}_{i,l}$ of variables characterizing the relational links between a mobile researcher i and a potential destination lab l; a vector $\mathbf{L}_l$ of characteristics describing the potential destination lab; and a vector $\mathbf{R}_i$ of characteristics describing the mobile researcher.

Vector $\mathbf{\Gamma}$ includes three variables capturing different forms of links between mobile researcher i and potential destination lab l. The first two variables measure direct relational links. Variable *Citation link*$_{i,l}$ is a dummy variable set to unity if at least one citation link exists between researcher i and any researcher affiliated with lab l during the five years preceding mobility, and zero otherwise. Variable *Co-authorship link*$_{i,l}$ is similarly defined as a dummy variable equal to one if researcher i has co-authored at least one paper with any researcher affiliated with lab l over the same five-year period. The third variable captures cognitive rather than relational links. We define *Low cognitive distance*$_{i,l}$ as a dummy variable equal to one when the cognitive distance between researcher i and lab l lies below the sample mean, and zero otherwise. This needs a prior measure of cognitive distance based on the distance between the journals in which mobile researchers published and the journals in which researchers affiliated with the potential destination lab published during the five years prior to mobility. The journal distance is computed using a co-citation matrix that captures how often pairs of journals are cited together within the same article.[8] In this matrix, journals that are more frequently co-cited are considered cognitively closer, and vice versa.

Before detailing the components of vector $\mathbf{\Gamma}$, we clarify the role of cognitive distance in the specification. Citation and co-authorship links capture direct relational ties between mobile researcher i and potential destination lab l. Cognitive proximity, by contrast, is a more latent form of connection, since it reflects similarity in knowledge space rather than an observed interpersonal relation. Although we

[8]Appendix C provides additional details on the construction of this measure.

allow *Low cognitive distance*$_{i,l}$ to affect destination choice directly, its main role in the specification is to moderate the effect of direct relational ties. Whether low cognitive distance strengthens or weakens the effect of citation and co-authorship links remains an open empirical question. On the one hand, proximity in knowledge space may make interpersonal ties more informative, more relevant, and easier to mobilize. On the other hand, it may render such ties partly redundant, since cognitively close labs may already be identifiable as suitable destinations even in the absence of direct relational links.

Given the above, we specify the contribution of $\mathbf{\Gamma}_{i,l}$ to destination choice as follows:

$$\begin{aligned}\mathbf{B_\Gamma \Gamma_{i,l}} &= \beta_{\Gamma,1}\,\text{Citation link}_{i,l} \\ &+ \beta_{\Gamma,2}\,\text{Co-authorship link}_{i,l} \\ &+ \beta_{\Gamma,3}\,\text{Low cognitive distance}_{i,l} \\ &+ \beta_{\Gamma,4}\,\text{Citation link}_{i,l} \times \text{Low cognitive distance}_{i,l} \\ &+ \beta_{\Gamma,5}\,\text{Co-authorship link}_{i,l} \times \text{Low cognitive distance}_{i,l}\end{aligned} \tag{2}$$

The terms $\mathbf{L}_l$ and $\mathbf{R}_i$ in Equation 1 are vectors of variables controlling for the lab and researcher characteristics, respectively. Vector $\mathbf{L}_l$ includes the *Geographic distance* that represents the distance between the origin and the potential destination lab l in kilometers.[9] It also includes the *Difference in lab size*, the *Difference in the number of journal publications*, the *Difference in administrative support staff*, and the *Difference in scientific support staff* between the potential destination lab l and the origin lab, measured in a time window of five years prior to the mobility. We also control for location with a *Paris* dummy that equals one if the potential destination lab l is located in Paris, and zero otherwise. Lastly, we proxy for the lab's ability to secure resources with a dummy variable *Funding capacity* that equals one if potential lab l is affiliated with a university[10] that held an excellence grant[11] in the year preceding mobility, and zero otherwise.

As for the vector of researcher characteristics $\mathbf{R}_i$, it includes the career rank

[9]Coordinates are obtained from OpenStreetMap, and distances are straight-line kilometers between the origin and potential destination city centers. Pairs in the same municipality are set to 0 km (Cambon et al. 2021).

[10]In France, a single university hosts many research labs across disciplines. Affiliation here refers to the lab's formal institutional link to the university.

[11]IdEx and I-SITE are competitive excellence grants providing substantial, multi-year institutional funding for selected universities that meet certain criteria. Those awards can be renewed or discontinued based on performance metrics.

using CNRS research fellow categories, *CR (Chargé.e de recherche)* and *DR (Directeur.trice de recherche)*,[12] with *CR* being the reference group. The variable *Female researcher* is a dummy variable that equals one if the researcher i is female, and zero otherwise. We also control for the *Age* of the researcher at the time of the move and her *Relative productivity*, measured as the researcher's total number of journal publications in the five years before mobility divided by the average number of journal publications per researcher in the potential destination lab. Finally, we control unobserved heterogeneity across disciplines and years by including *Discipline* fixed effects corresponding to CNRS institutes[13] and *Year* fixed effects capturing the time of the move.

Studying destination choices might raise a sample selection problem (Dahl 2002, Verginer & Riccaboni 2021). Indeed, we observe the choice of a destination lab only for researchers who decide to move. Hence, the sample of mobile researchers is not random, and limiting the study sample to mobile researchers may lead to biased and inconsistent estimates. To address this issue, we implement a two-step Heckman correction (Heckman 1979). In the first stage, reported in Appendix F, we estimate a selection equation considering the whole population of 2,033 CNRS researchers, including both mobile and non-mobile researchers. Specifically, we model the probability of researchers' mobility as a function of the researcher's and origin lab's characteristics, as reported in Equation F1. After estimating Equation F1, we compute the correction term λ_i included in Equation 1 to correct for potential selection bias that might arise from the non-random decision to move.

3.3 Descriptive statistics

Table 1 shows the descriptive statistics separately for non-chosen *(Choice = 0)* and chosen *(Choice = 1)* labs. The chosen labs show higher levels of *Citation link* activity between the researcher and the chosen lab before mobility. Specifically, on average, researchers have citation links with chosen labs in 37.1% of the cases, compared to only 13.5% of the cases for non-chosen labs. A very similar pattern is observed for co-authorship links. Researchers show a *Co-authorship link* with chosen labs in 25.8% of the cases, compared to only 2.9% for non-chosen labs. *Low cognitive*

[12] The CNRS ranks do not follow the typical academic titles. However, a CR usually corresponds in seniority to an assistant/associate professor, while DR usually corresponds to a full professor. Both positions correspond to a civil servant status.

[13] CNRS has ten institutes: chemistry (INC), physics (INP), biological sciences (INSB), information sciences & interactions (INS2I), engineering systems (INSIS), humanities and social sciences (INSHS), ecology & environment (INEE), nuclear & particle physics (IN2P3), earth & universe sciences (INSU), and mathematical sciences (INSMI).

distance and *Geographic distance* also show some differences between the two groups, but to a lesser extent. Having a *Low cognitive distance* between the researcher and the lab is more common among chosen labs than among non-chosen labs (50.6% vs. 40.1%), and researchers also seem to choose geographically closer labs (285 km vs. 381 km). Other lab and researcher characteristics show very similar means across both groups.

Table 1: Descriptive statistics

	$Choice = 0 (N = 6,658)$				$Choice = 1 (N = 217)$			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Relational links								
Citation links	0.135	0.342	0.000	1.000	0.371	0.484	0.000	1.000
Co-authorship links	0.029	0.169	0.000	1.000	0.258	0.439	0.000	1.000
Low cognitive distance	0.401	0.490	0.000	1.000	0.516	0.501	0.000	1.000
Lab characteristics								
Geographic distance	381.847	231.614	0.000	1029.905	285.375	244.295	0.000	958.361
Difference in lab size	-3.459	105.790	-439.000	419.000	-3.127	101.711	-301.000	319.000
Difference in number of publications	-22.740	429.258	-1965.000	1950.000	-38.761	399.727	-1643.000	1570.000
Difference in admin support	-0.236	6.182	-29.000	34.000	-0.028	5.774	-26.000	22.000
Difference in science support	0.378	15.768	-89.000	108.000	0.254	15.700	-79.000	72.000
Paris	0.227	0.419	0.000	1.000	0.235	0.425	0.000	1.000
Funding capacity	0.659	0.474	0.000	1.000	0.615	0.488	0.000	1.000
Researcher characteristics								
Research Director (DR)	0.122	0.327	0.000	1.000	0.127	0.333	0.000	1.000
Female researcher	0.340	0.474	0.000	1.000	0.371	0.484	0.000	1.000
Age at movement	37.691	4.399	31.000	55.000	38.019	4.624	31.000	55.000
Relative productivity	3.804	5.883	0.000	181.138	4.050	8.039	0.000	95.464
Movement year	2019.580	2.049	2014.000	2022.000	2019.610	2.001	2014.000	2022.000
Correction term λ	1.548	0.341	0.768	2.721	1.562	0.344	0.768	2.721

Notes: Descriptive statistics are reported separately for observations where the dependent variable *choice* equals 0 (non-chosen labs) and 1 (chosen labs).

4 Results

4.1 Main results

Table 2 reports the average marginal effects[14] calculated based on the estimates of Equation 1. Columns (1) and (2) include only the main independent variables *Citation link* and *Co-authorship link*, along with *Low cognitive distance*, without including any additional control variables. In column (1), these variables enter the model independently, while column (2) allows for their interaction with Low cognitive distance. Column (3) adds to the specification of column (2) by introducing a set of controls, which include researcher and lab characteristics, as well as discipline and year fixed effects to account for unobserved heterogeneity across disciplines and years. Finally, column (4) changes the definition of citation links by focusing on *researcher-to-lab citations* only instead of the bidirectional citation links. The correction term from the first-stage selection equation is also included in all four models.[15]

Columns (1) and (2) show that both citation and co-authorship links are strong predictors of destination choice. In other words, mobile researchers are more likely to choose labs with which they have pre-existing citation links and co-authorship ties. The coefficients on both variables remain positive and statistically significant even after controlling for researcher and lab characteristics in column (3). In terms of magnitude, the average marginal effects in column (3) show that having at least one citation link with a potential destination lab increases the probability of choosing that lab by 3.7 percentage points, while having prior co-authorship ties increases this probability by 9.8 percentage points. Although these marginal effects seem small in absolute terms, they are in fact large relative to the unconditional probability of choosing a lab, which is about 3%.

We now turn to column (4), where citation links are restricted to researcher-to-lab citations only. This distinction is relevant in the CNRS context because destination choices are driven by the researcher rather than by destination labs. Therefore, focusing on citations made by the researcher captures whether she is aware of the destination lab's scientific output. We find that the coefficient on researcher-to-lab citations remains positive and statistically significant, although its

[14] The logit coefficients are reported in Table E1 in Appendix E.

[15] To show that our results do not depend critically on the Heckman correction, we re-estimate the outcome equation in Appendix A without including the correction term and end up with similar results.

marginal effect is smaller than that of bidirectional citation links in column (3). The effect of prior co-authorship remains positive and statistically significant. Finally, the coefficient on *Low cognitive distance* is positive and statistically significant once we control for researcher and lab characteristics in columns (3) and (4), meaning that researchers are more likely to choose labs that are working on similar or closely related topics to their own.

Regarding the control variables, only three show statistically significant coefficients. *Geographic distance* appears to be negatively correlated with destination choice, indicating that researchers are less likely to favor labs located farther away from their lab of origin. This shows that physical distance remains a barrier to mobility, even in internal labor markets in a single country. Moreover, the *Paris* dummy is negative and statistically significant, reflecting the high living costs and housing constraints which make labs in Paris less attractive to researchers. Finally, *Relative productivity* is negatively correlated with destination choice, suggesting that researchers are less likely to move to labs where they are more productive compared to the lab average. All remaining controls show no statistical significance. Our results support our hypotheses by showing that destination choices are mainly driven by citation and co-authorship links.

Table 2: Average marginal effects of the second-stage outcome equation

	(1)	(2)	(3)	(4)
Relational links				
Citation links	0.013*	0.027**	0.037***	
	(0.007)	(0.012)	(0.014)	
Researcher-to-lab citations				0.024*
				(0.014)
Co-authorship links	0.059***	0.117***	0.098***	0.108***
	(0.007)	(0.027)	(0.024)	(0.027)
Low cognitive distance	0.005	0.004	0.013**	0.014**
	(0.003)	(0.003)	(0.006)	(0.006)
Lab characteristics				
Geographic distance (in natural logs)			-0.006***	-0.006***
			(0.001)	(0.001)
Difference in lab size$^{(a)}$			0.004	0.005
			(0.004)	(0.004)
Difference in number of publications$^{(a)}$			-0.003	-0.006*
			(0.004)	(0.003)
Difference in admin support$^{(a)}$			0.002	0.002
			(0.002)	(0.002)
Difference in science support$^{(a)}$			-0.002	-0.003
			(0.003)	(0.003)
Paris			-0.011*	-0.010*
			(0.006)	(0.005)
Funding capacity			-0.004	-0.004
			(0.005)	(0.005)
Researcher characteristics				
Research Director (DR)			0.003	0.004
			(0.003)	(0.003)
Female researcher			0.001	0.002
			(0.002)	(0.002)
Age at movement			0.000	0.000*
			(0.000)	(0.000)
Relative productivity (in natural logs)			-0.006***	-0.003***
			(0.001)	(0.001)
Correction term λ	0.002	0.002	0.007**	0.005*
	(0.003)	(0.002)	(0.003)	(0.003)
Discipline	No	No	Yes	Yes
Movement year	No	No	Yes	Yes

Notes: $N = 6,875$. The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. Superscript a indicates standardized variables. Column (1) includes the main independent variables without interactions or controls. Column (2) introduces interactions with low cognitive distance. Column (3) adds researcher and lab controls, as well as discipline and year fixed effects. Column (4) uses researcher-to-lab citations instead of the bidirectional citation links. All four models include a correction term from the first-stage selection equation reported in Table F2 in Appendix F. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2 Placebo exercise

To assess whether our main results are driven by mechanical correlations between researcher and lab characteristics rather than by the informational content of citation and co-authorship links, we implement several placebo tests.[16] These tests randomize the correlation structure between destination choices, relational variables, namely *Citation links, Co-authorship links* and *Low cognitive distance*, and the remaining covariates in the model. The objective is to alter any meaningful informational content embedded in these links while preserving key features of the original data structure. If the estimated effects were merely driven by spurious correlations between the variables of interest and destination choice, they should survive a randomization of the data structure. By contrast, if our baseline results capture a genuine systematic alignment between researcher characteristics and destination labs, the estimated relationships should vanish once this alignment is broken.

We perform four placebo exercises. In the first exercise we randomize the destination choice within each researcher-specific choice set, while leaving the matrix of explanatory variables unchanged. This procedure preserves researcher-specific choice sets and researcher-level characteristics, while breaking any systematic relationship between the observed covariates and the realized destination. In two additional exercises, we randomize the relational structure of the data. Exercise 2 permutes *Citation links, Co-authorship links* and *Low cognitive distance* across potential destination labs within each researcher's choice set, while Exercise 3 permutes these variables jointly in order to preserve their correlation structure.

These first three placebo tests are informative about the role of unobserved researcher heterogeneity: since the randomization occurs within each researcher's choice set, it preserves systematic differences across researchers in all three variables describing relational ties. If the baseline results were driven solely by such differences, the estimated relationships should persist under randomization. By contrast, if they reflect the fact that chosen destinations are more strongly connected to researchers than non-chosen alternatives within a researcher's set, the placebo estimates should collapse toward zero. This placebo therefore mimics the logic of a within-researcher comparison, although it is not a substitute for including researcher fixed effects in the baseline specification.

Finally, Exercise 4 implements a final placebo that reverses the assignment logic by randomizing researchers across labs. Specifically, for each lab, we randomly assign

[16] Appendix H provides a detailed discussion of the placebo exercises, including the underlying hypotheses.

researchers by selecting the same number of realized matches as observed in the data. This procedure preserves the distribution of realized matches at the lab level, while breaking the correspondence between labs and the characteristics of the researchers matched to them. Again, if the baseline findings are driven by systematic alignment between researcher characteristics and destination labs, the estimated relationships should disappear under this reassignment procedure.

Table 3 reports the results of the four placebo exercises based on $B = 2,000$ replications. Across all four exercises, the estimated marginal effects of *Citation links, Co-authorship links* and *Low cognitive distance* collapse toward zero and are statistically insignificant. This supports the interpretation that the baseline results capture meaningful relational links rather than spurious correlations.

Table 3: Placebo marginal effects over $B = 2,000$ placebo replications

	(1)	(2)	(3)	(4)
Citation link	0.001	-0.000	-0.001	0.000
	(0.008)	(0.006)	(0.007)	(0.008)
	[-0.016;0.018]	[-0.013;0.012]	[-0.016;0.015]	[-0.016;0.016]
Co-authorship link	0.004	0.002	0.005	0.005
	(0.014)	(0.011)	(0.015)	(0.015)
	[-0.025;0.033]	[-0.021;0.024]	[-0.025;0.035]	[-0.024;0.034]
Low cognitive distance	0.003	0.003	0.002	0.003
	(0.005)	(0.004)	(0.004)	(0.005)
	[-0.006;0.012]	[-0.005;0.011]	[-0.005;0.010]	[-0.007;0.012]

Notes: Table 3 reports results from Equation 1 pertaining to the three variables of interest: citation links, co-authorship links, and low cognitive distance. The reported results are the average marginal effects, the standard errors in parentheses, and the 95% confidence intervals in square brackets. Column (1) randomizes destination choices within researcher-specific choice sets. Column (2) permutes relational links across potential destinations within choice sets. Column (3) jointly permutes relational links to preserve their correlation structure. Column (4) reassigns researchers across labs while preserving the number of realized matches per lab.

4.3 Further analyses

Our baseline results indicate that, while both citations and prior co-authorship links are strong predictors of destination choice, the magnitude of the effect of prior co-authorship is almost three times larger than that of citations. In Section 2, we argued that the role of citations and prior co-authorship is likely moderated by the cognitive distance between the researcher and the potential destination lab. However, the marginal effects reported in Table 2 correspond to the sample average across the

Low cognitive distance variable. In the following analyses, we report the average marginal effects of *Citation* and *Co-authorship links* calculated separately for *Low cognitive distance* equal to 0 and 1. This allows us to assess whether the importance of *Citation* and *Co-authorship links* varies with different levels of cognitive distance.

Table 4 reports the average marginal effects of the variables of interest, with all control variables included. Columns (1) and (2) use the same specification as column (3) of Table 2, and columns (3) and (4) provide a similar exercise using the specification from column (4) of Table 2 but compute the average marginal effects separately for cognitively distant and cognitively proximate pairs.[17] The results show that the two relational links, citations and prior co-authorship, play different roles depending on cognitive distance. For cognitively distant pairs, citation links are the only significant predictor of destination choice, while prior co-authorship is not statistically significant. For cognitively proximate pairs, the opposite pattern emerges: prior co-authorship becomes the only significant predictor, while citation links no longer have a significant effect. This suggests that citations and co-authorship reduce different types of information asymmetry.

When the researcher and the potential destination lab work on cognitively distant topics, the main uncertainty concerns scientific relevance: does the knowledge produced by one party matter for the research activities of the other? Citation links are informative in this context because they reveal actual knowledge flows across distant domains. A citation from the researcher to the potential destination lab indicates that the researcher draws on the lab's scientific output. Conversely, a citation from the potential destination lab to the researcher indicates that the lab relies on the researcher's knowledge stock. In both cases, the citation link provides evidence that the two parties' knowledge stocks may be scientifically complementary despite their cognitive distance. By contrast, prior co-authorship is less informative in cognitively distant contexts. Such collaborations may occur because a specific project requires complementary expertise, but they do not necessarily indicate that the potential destination lab is a natural or likely destination for the researcher. In this case, co-authorship may capture project-specific complementarity rather than a broader match between the researcher and the lab.

The logic changes for cognitively proximate pairs. When the researcher and the potential destination lab work on closely related topics, the scientific relevance of each party's work is easier to assess. The remaining uncertainty is therefore less

[17] Cognitively distant pairs correspond to observations for which *Low cognitive distance* equals 0. Cognitively proximate pairs correspond to observations for which *Low cognitive distance* equals 1.

about whether their knowledge stocks are relevant to one another, and more about the tacit and interpersonal dimensions of collaboration: whether the parties can work together effectively, coordinate their research practices, and sustain a productive working relationship. In this context, prior co-authorship is especially informative because it provides direct evidence of successful collaboration. Citation links, by contrast, add little information once cognitive proximity already establishes the scientific relevance of the match.

Table 4: Average marginal effects, by cognitive distance

	(1) Low cognitive distance = 0	(2) Low cognitive distance = 1	(3) Low cognitive distance = 0	(4) Low cognitive distance = 1
Citation links	0.052***	0.001		
	(0.019)	(0.011)		
Researcher-to-lab citations			0.040*	0.001
			(0.021)	(0.011)
Co-authorship links	0.035	0.192***	0.051	0.193***
	(0.024)	(0.047)	(0.033)	(0.047)

Notes: $N = 6,875$. The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. Columns (1) and (2) report the average marginal effects based on column (3) of Table 2. Columns (3) and (4) report the average marginal effects based on column (4) of Table 2. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Next, we examine whether the role of information channels varies across disciplines. To do so, we distinguish between humanities and social sciences (INSHS) and STEM disciplines[18] and report the average marginal effects separately for cases where *Low cognitive distance* equals 0 and 1. Columns (1) and (2) of Table 5 report the results for INSHS when *Low cognitive distance* equals 0 and 1, respectively. Columns (3) and (4) report the corresponding results for STEM disciplines. All columns are estimated using the same specification as in column (3) of Table 2.

Our results show that the mechanism discussed in Table 4 differs across disciplines. When the cognitive distance between researchers and potential destination labs is high (*Low cognitive distance = 0*), citation links appear to be the main predictor of destination choice across both disciplines, consistent with the results in Table 4. However, when the cognitive distance is low (*Low cognitive distance = 1*), co-authorship links appear to be a predictor of destination choice only in STEM disciplines. These results suggest that the type of information asymmetry researchers

[18]STEM disciplines in our sample correspond to the following CNRS institutes: chemistry (INC), physics (INP), biological sciences (INSB), information sciences & interactions (INS2I), engineering systems (INSIS), ecology & environment (INEE), nuclear & particle physics (IN2P3), earth & universe sciences (INSU), and mathematical sciences (INSMI).

face varies across disciplines. In STEM, knowledge is highly specialized and research is often conducted in large teams and requires access to specific equipment and technical resources (Larivière et al. 2006, 2015). As a result, information related to the knowledge stock, whether codified or tacit, becomes valuable for evaluating potential destination labs. In INSHS, on the other hand, research is more frequently conducted in small teams and relies significantly less on equipment (Larivière et al. 2006, 2015). In this case, observable characteristics of the research environment, such as lab characteristics, provide more relevant information for researchers when evaluating potential destination labs, as shown in Table G1 in Appendix G.

Table 5: Average marginal effects of the second-stage outcome equation, by low cognitive distance and discipline

	INSHS		STEM	
	Low cognitive distance = 0	Low cognitive distance = 1	Low cognitive distance = 0	Low cognitive distance = 1
Citation links	0.081*	0.003	0.050**	0.001
	(0.048)	(0.097)	(0.025)	(0.011)
Co-authorship links	0.063	0.097	0.021	0.202***
	(0.072)	(0.149)	(0.022)	(0.051)
Number of observations	1,197	1,197	5,678	5,678

Notes: The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. INSHS stands for humanities and social sciences. STEM stands for science, technology, engineering, and mathematics. Columns report average marginal effects separately by discipline and by the value of *Low cognitive distance*. All columns are estimated using the same specification as in column (3) of Table 2. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Conclusion

The goal of this paper was to identify the factors predicting destination choices among mobile researchers within an organization. We have used a unique dataset covering all researchers recruited by the French National Center for Scientific Research (CNRS) between 2012 and 2022 and constructed, for each researcher, the complete set of potential destination labs. We then applied a two-stage Heckman selection model to correct for selection bias and account for the fact that destination choices are only observed for researchers who decide to move.

Our econometric analysis shows that researchers rely on scientific publications and prior temporary professional collaborations to address information asymmetries when choosing a destination lab. Specifically, both citation and prior co-authorship links significantly increase the probability that a researcher selects a given destination lab. However, the magnitude of these effects is different, with co-authorship being the dominant driver of destination choice. Having a prior professional collaboration with a potential destination lab increases the probability of choosing that lab by 9.8 percentage points, compared to 3.7 percentage points for citation links. We further show in an additional analysis that the role of information channels is moderated by the cognitive distance between the researcher and the destination lab. When cognitive distance is large, the effect of prior co-authorship disappears, while citation links appear to be sufficient for researchers to make an informed decision. On the other hand, when cognitive distance is low, prior co-authorship appears to provide valuable information for researchers, whereas the effect of citation links disappears. Finally, we find discipline heterogeneity. The effect of co-authorship links is mainly observed in STEM disciplines, whereas it does not appear to predict destination choices in humanities and social sciences.

Taken together, our results indicate that destination choices are shaped primarily by relational links between mobile researchers and potential destination labs. This has a direct policy implication. If the objective is to attract researchers to a specific lab, policies should not focus only on improving the lab's general attractiveness through higher productivity, funding, or scientific support. They should also create opportunities for relational links to emerge between the lab and external researchers. One possible instrument is the funding of short-term visiting periods, which can expose external researchers to the lab, facilitate interaction with local teams, and create the conditions for future collaboration. Such programs may therefore be more effective at influencing destination choices than policies aimed exclusively at increasing the lab's scientific excellence.

A second implication of our results is that relational links should not be treated as interchangeable signals. Citation and co-authorship links convey different types of information and reduce different forms of uncertainty before mobility. Prior co-authorship provides information about tacit and interpersonal compatibility: whether researchers can work together effectively, coordinate their research practices, and trust each other in the production of knowledge. Co-authorship-based moves may therefore correspond to safer matches, with a lower risk of post-mobility integration problems. This safety may, however, come at the cost of novelty, since repeated collaboration can limit exposure to new sources of knowledge (Guimera et al. 2005).

Citation links capture a different type of relation. They reveal scientific relevance between two parties, but do not resolve uncertainty about whether collaboration will work in practice. Citation-based moves may therefore be more exploratory. They may involve a higher risk of unsuccessful recombination, but they may also create more room for serendipity, understood as valuable scientific novelty emerging from unexpected combinations of knowledge (Merton & Barber 2004, Yaqub 2018, Copeland 2019). This interpretation is consistent with evidence that recombining distant knowledge is less successful on average, but can increase the probability of breakthrough outcomes (Fleming 2001). It is also consistent with the finding that high-impact scientific work often combines conventional knowledge with more atypical combinations (Uzzi et al. 2013). Research organizations may therefore face a trade-off between supporting safer mobility based on prior co-authorship and encouraging more exploratory mobility based on citation links. The former may increase the likelihood of successful post-mobility integration; the latter may increase the likelihood of more novel, and potentially serendipitous, scientific outcomes.

Our paper is not without limitations. First, we study destination choices within an internal labor market where the set of potential destination labs is limited and fully observable. This setting allows us to better understand destination choices, but it also implies that our sample is not representative of the behavior of the typical researcher looking to change organizations. In external labor markets, researchers face a much larger choice set, and destination choices may be influenced by labor market frictions that we do not capture here. Second, we introduce citations and prior co-authorship as information channels that reduce information asymmetries. However, these measures capture only formal scientific links, and informal links, such as seminars, are not observed. From an organizational perspective, explaining destination choice is only one part of the story for understanding, and ideally avoiding, mismatches. However, low-quality matches can still occur, and future research could quantify the cost of this mismatch by estimating how much output is lost when researchers do not move to labs that best match their skills.

Acknowledgements

We thank participants of the Workshop on the Organization, Economics, and Policy of Scientific Research (WOEPS), Milan, Italy, in March 2026, for their valuable comments and suggestions. We are especially grateful to our discussant, Martina Lori, for taking the time to carefully read our paper and for providing constructive feedback. We are also grateful to Anne Sigogneau, Laurent Silvera, Alpha Owusu from the "Direction d'appui aux partenariats publics" (DAPP) and Gaëlle Bujan at CNRS for providing and helping us with the dataset. We gratefully acknowledge Cyril Moulin, Etienne Farvaque and all the participants at CNRS-ADSR meetings, Paris, France. We are also grateful to the French government for its financial support of this work through the grant ANR-24-CE26-1154.

Competing interests

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ackers, L. (2005), 'Moving people and knowledge: Scientific mobility in the european union1', *International Migration* **43**(5), 99–131.

Azoulay, P., Ganguli, I. & Zivin, J. G. (2017), 'The mobility of elite life scientists: Professional and personal determinants', *Research Policy* **46**(3), 573–590.

Baruffaldi, S. H. & Landoni, P. (2016), 'Mobility intentions of foreign researchers: The role of non-economic motivations', *Industry and Innovation* **23**(1), 87–111.

Burdett, K. & Coles, M. G. (1999), 'Long-term partnership formation: Marriage and employment', *The Economic Journal* **109**(456), 307–334.

Cambon, J., Hernangómez, D., Belanger, C. & Possenriede, D. (2021), 'tidygeocoder: An r package for geocoding', *Journal of Open Source Software* **6**(65), 3544.

Cohen, W. M., Levinthal, D. A. et al. (1990), 'Absorptive capacity: A new perspective on learning and innovation', *Administrative Science Quarterly* **35**(1), 128–152.

Copeland, S. (2019), 'On serendipity in science: Discovery at the intersection of chance and wisdom', *Synthese* **196**(6), 2385–2406.

Cowan, R., David, P. A. & Foray, D. (2000), 'The explicit economics of knowledge codification and tacitness', *Industrial and Corporate Change* **9**(2), 211–253.

Dahl, G. B. (2002), 'Mobility and the return to education: Testing a roy model with multiple markets', *Econometrica* **70**(6), 2367–2420.

De la Croix, D., Docquier, F., Fabre, A. & Stelter, R. (2024), 'The academic market and the rise of universities in medieval and early modern europe (1000–1800)', *Journal of the European Economic Association* **22**(4), 1541–1589.

Edler, J., Fier, H. & Grimpe, C. (2011), 'International scientist mobility and the locus of knowledge and technology transfer', *Research Policy* **40**(6), 791–805.

Fleming, L. (2001), 'Recombinant uncertainty in technological search', *Management Science* **47**(1), 117–132.

Franzoni, C., Scellato, G. & Stephan, P. (2012), 'Foreign-born scientists: mobility patterns for 16 countries', *Nature Biotechnology* **30**(12), 1250–1253.

Greenwood, M. J. (1975), ‘Research on internal migration in the united states: A survey’, *Journal of Economic Literature* pp. 397–433.

Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. (2005), ‘Team assembly mechanisms determine collaboration network structure and team performance’, *Science* **308**(5722), 697–702.

Gürtzgen, N. & Pohlan, L. (2024), ‘Do employers learn more from referrals than from other recruitment channels?’, *Labour Economics* **89**, 102574.

Heckman, J. J. (1979), ‘Sample selection bias as a specification error’, *Econometrica* pp. 153–161.

Jaffe, A. B., Trajtenberg, M. & Henderson, R. (1993), ‘Geographic localization of knowledge spillovers as evidenced by patent citations’, *The Quarterly Journal of Economics* **108**(3), 577–598.

Jones, B. F. (2009), ‘The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder?’, *The Review of Economic Studies* **76**(1), 283–317.

Jovanovic, B. (1979), ‘Job matching and the theory of turnover’, *Journal of Political Economy* **87**(5, Part 1), 972–990.

Kaiser, U., Kongsted, H. C. & Rønde, T. (2015), ‘Does the mobility of r&d labor increase innovation?’, *Journal of Economic Behavior & Organization* **110**, 91–105.

Katz, J. S. & Martin, B. R. (1997), ‘What is research collaboration?’, *Research Policy* **26**(1), 1–18.

Kogut, B. & Zander, U. (1992), ‘Knowledge of the firm, combinative capabilities, and the replication of technology’, *Organization Science* **3**(3), 383–397.

Larivière, V., Gingras, Y. & Archambault, É. (2006), ‘Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities’, *Scientometrics* **68**(3), 519–533.

Larivière, V., Gingras, Y., Sugimoto, C. R. & Tsou, A. (2015), ‘Team size matters: Collaboration and scientific impact since 1900’, *Journal of the Association for Information Science and Technology* **66**(7), 1323–1332.

Lazear, E. P. & Gibbs, M. (2014), *Personnel economics in practice*, John Wiley & Sons.

Levin, S. G. & Stephan, P. E. (1999), 'Are the foreign born a source of strength for us science?'.

Merton, R. K. & Barber, E. (2004), *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*, Princeton University Press, Princeton, NJ.

Mortensen, D. T. & Pissarides, C. A. (1994), 'Job creation and job destruction in the theory of unemployment', *The review of economic studies* **61**(3), 397–415.

Nakajima, R., Tamura, R. & Hanaki, N. (2010), 'The effect of collaboration network on inventors' job match, productivity and tenure', *Labour Economics* **17**(4), 723–734.

Narin, F., Hamilton, K. S. & Olivastro, D. (1997), 'The increasing linkage between us technology and public science', *Research Policy* **26**(3), 317–330.

Nelson, R. R. & Winter, S. G. (1985), *An evolutionary theory of economic change*, harvard university press.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V. & Van den Oord, A. (2007), 'Optimal cognitive distance and absorptive capacity', *Research Policy* **36**(7), 1016–1034.

Priem, J., Piwowar, H. & Orr, R. (2022), 'Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts', *arXiv preprint arXiv:2205.01833* .

Sengul, G. (2017), 'Learning about match quality: Information flows and labor market outcomes', *Labour Economics* **46**, 118–130.

Spence, M. (1978), Job market signaling, *in* 'Uncertainty in economics', Elsevier, pp. 281–306.

Stephan, P. (2012), *How economics shapes science*, Harvard University Press.

Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. (2013), 'Atypical combinations and scientific impact', *Science* **342**(6157), 468–472.

Verginer, L. & Riccaboni, M. (2021), 'Talent goes to global cities: The world network of scientists' mobility', *Research Policy* **50**(1), 104127.

Weitzman, M. L. (1998), 'Recombinant growth', *The Quarterly Journal of Economics* **113**(2), 331–360.

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.

Wuchty, S., Jones, B. F. & Uzzi, B. (2007), 'The increasing dominance of teams in production of knowledge', *Science* **316**(5827), 1036–1039.

Yaqub, O. (2018), 'Serendipity: Towards a taxonomy and a theory', *Research Policy* **47**(1), 169–179.

Zucker, L. G., Darby, M. R. & Brewer, M. B. (1998), 'Intellectual human capital and the birth of us biotechnology enterprises', *The American Economic Review* **88**(1), 290–306.

Appendix A. Outcome equation estimated without the Heckman correction

In Appendix A, we re-estimate the outcome equation without including the correction term to assess the robustness of our baseline results. All coefficients in Tables A1, A2, and A3 remain stable.

Table A1: Average marginal effects of the second-stage outcome equation, excluding the correction term

	(1)	(2)	(3)	(4)
Relational links				
Citation links	0.013*	0.027**	0.032***	
	(0.007)	(0.012)	(0.012)	
Researcher-to-lab citations				0.025*
				(0.014)
Co-authorship links	0.059***	0.116***	0.099***	0.108***
	(0.007)	(0.027)	(0.024)	(0.027)
Low cognitive distance	0.005	0.004	0.012**	0.013**
	(0.003)	(0.003)	(0.006)	(0.006)
Lab characteristics				
Geographic distance (in natural logs)			-0.006***	-0.006***
			(0.001)	(0.001)
Difference in lab size$^{(a)}$			0.006	0.005
			(0.004)	(0.004)
Difference in number of publications$^{(a)}$			-0.007*	-0.006*
			(0.003)	(0.003)
Difference in admin support$^{(a)}$			0.002	0.002
			(0.002)	(0.002)
Difference in science support$^{(a)}$			-0.003	-0.003
			(0.003)	(0.003)
Paris			-0.010*	-0.010*
			(0.006)	(0.005)
Funding capacity			-0.003	-0.004
			(0.005)	(0.005)
Researcher characteristics				
Research Director (DR)			0.004	0.004
			(0.003)	(0.003)
Female researcher			0.002	0.002
			(0.002)	(0.002)
Age at movement			0.000	0.001*
			(0.000)	(0.000)
Relative productivity (in natural logs)			-0.004**	-0.003**
			(0.001)	(0.001)
Discipline	No	No	Yes	Yes
Movement year	No	No	Yes	Yes

Notes: $N = 6,875$. The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. Superscript a indicates standardized variables. Column (1) includes the main independent variables without interactions or controls. Column (2) introduces interactions with low cognitive distance. Column (3) adds researcher and lab controls, as well as discipline and year fixed effects. Column (4) uses researcher-to-lab citations instead of the bidirectional citation links. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Average marginal effects of the second-stage outcome equation, by low cognitive distance and excluding the correction term

Dependent variable: choice	(1) Low cognitive distance = 0	(2) Low cognitive distance = 1	(3) Low cognitive distance = 0	(4) Low cognitive distance = 1
Citation links	0.053***	0.002		
	(0.019)	(0.011)		
Researcher-to-lab citations			0.040*	0.002
			(0.022)	(0.011)
Co-authorship links	0.035	0.192***	0.052	0.191***
	(0.024)	(0.047)	(0.033)	(0.046)

Notes: $N = 6,875$. The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. Columns (1) and (2) report the average marginal effects based on the same specification as in column (3) of Table 2. Columns (3) and (4) report the average marginal effects based on the same specification as in column (4) of Table 2. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Average marginal effects of the second-stage outcome equation, by low cognitive distance and discipline and excluding the correction term

	(1) INSHS Low cognitive distance = 0	(2) INSHS Low cognitive distance = 1	(3) STEM Low cognitive distance = 0	(4) STEM Low cognitive distance = 1
Citation links	0.081*	0.001	0.051**	0.001
	(0.048)	(0.096)	(0.025)	(0.011)
Co-authorship links	0.060	0.103	0.021	0.202***
	(0.068)	(0.154)	(0.022)	(0.051)
Number of observations	1,197	1,197	5,678	5,678

Notes: The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. INSHS stands for humanities and social sciences. STEM stands for science, technology, engineering, and mathematics. Columns (1) and (2) report the average marginal effects for humanities and social sciences when *Low cognitive distance* equals 0 and 1, respectively. Columns (3) and (4) report the corresponding results for STEM disciplines. All columns are estimated using the same specification as in column (3) of Table 2. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B. Distribution of moves across disciplines

To construct the potential destination choice set of each researcher, we needed to define the set of labs that could be considered as realistic options. Since CNRS is organized into broad scientific disciplines, it is reasonable to assume that researchers move within their own discipline. To validate this assumption, Table B1 shows the distribution of moves between the origin and destination disciplines of all moves. Rows correspond to the origin discipline and columns correspond to the destination discipline. We can notice from the diagonal that the vast majority of researchers relocate to a lab within the same discipline as their own lab of origin. Only about 17% of researchers switch to a different discipline.

Table B1: Distribution of moves by origin and destination disciplines

	INC	INP	INSB	INS2I	INSIS	INSHS	INEE	IN2P3	INSU	INSMI
INC	**18**	1	1	0	0	0	0	0	0	0
INP	2	**12**	1	0	1	0	0	0	1	0
INSB	6	0	**40**	0	0	1	3	0	0	0
INS2I	0	0	0	**21**	2	0	0	0	0	0
INSIS	1	1	0	0	**1**	0	0	0	0	1
INSHS	0	0	2	3	0	**44**	3	0	1	0
INEE	0	0	0	0	0	0	**13**	0	1	0
IN2P3	1	1	0	0	0	0	0	**3**	0	0
INSU	0	1	0	0	0	0	1	0	**10**	0
INSMI	0	0	0	0	0	0	0	0	0	**15**

Notes: CNRS has ten institutes: chemistry (INC), physics (INP), biological sciences (INSB), information sciences & interactions (INS2I), engineering systems (INSIS), humanities and social sciences (INSHS), ecology & environment (INEE), nuclear & particle physics (IN2P3), earth & universe sciences (INSU), and mathematical sciences (INSMI).

Appendix C. Cognitive distance measure

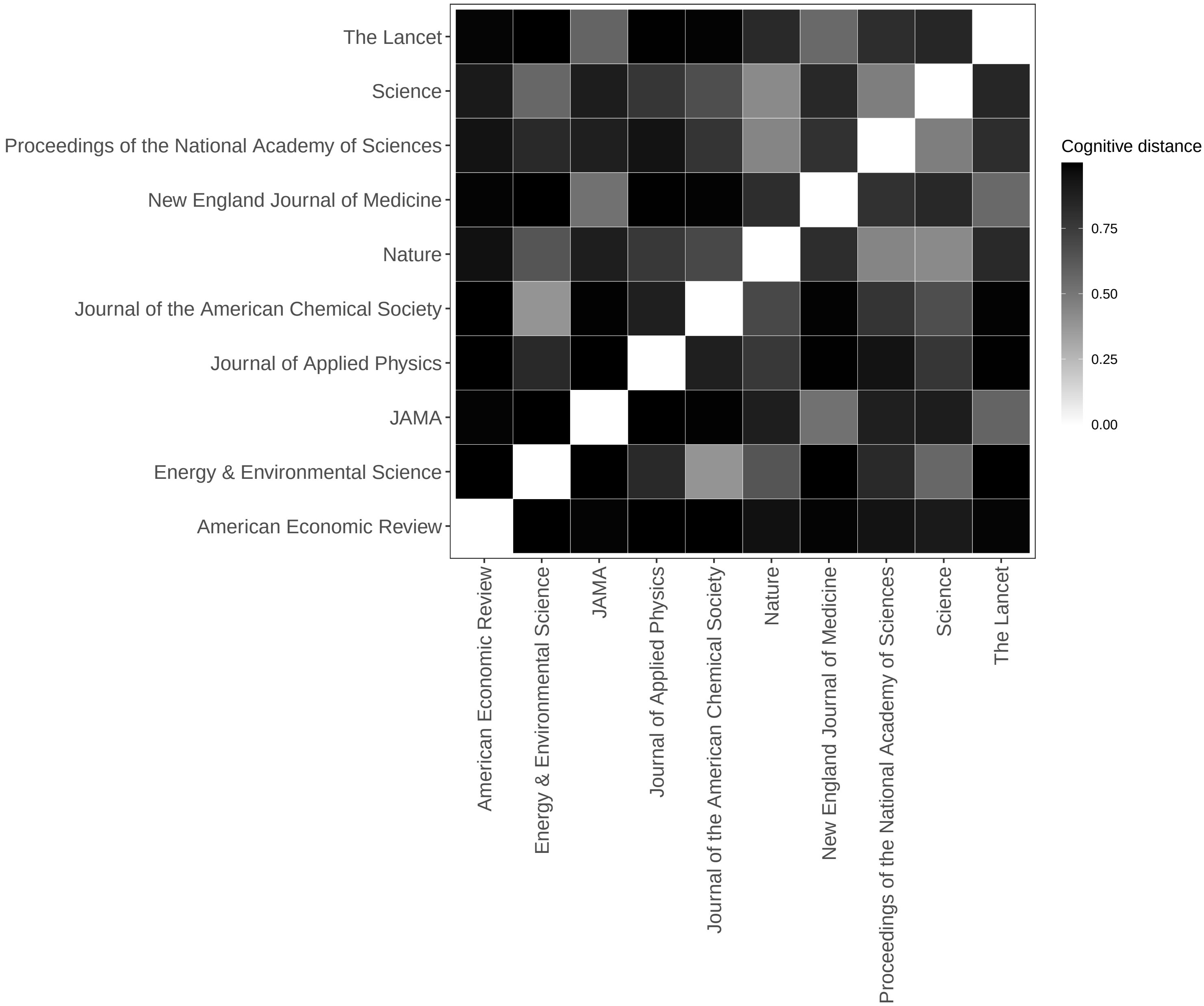
In Appendix C, we describe the construction of the journal distance matrix used to measure the cognitive distance between researchers and labs. The goal is to quantify how similar two journals are in terms of content. Figure C1 represents a heatmap of the distances between 10 journals. Darker colors indicate larger distances, while lighter colors indicate smaller distances. For example, the distance between *The Lancet* and *JAMA*, as well as between *The Lancet* and *The New England Journal of Medicine*, is represented by a relatively light color, indicating that these journals are cognitively close to one another since these journals all belong to the broad medical field. However, the distance between *The Lancet* and the *American Economic Review* is represented by a much darker color, indicating a large cognitive distance and suggesting little to no overlap in their scientific content. The heatmap is extracted from a larger matrix that reports the distances between all possible pairs of journals among the 41,189 journals cited in the reference lists of a random sample of 1,462,877 articles published between 2000 and 2020 with at least one French author. If two journals co-appear frequently in the reference list of the same articles, we consider that they must be close in terms of content. To operationalize this idea, we compute $\text{freq}_{A,B}$, which represents the number of articles in which a pair of journals A and B are cited together. We then define the distance between journals A and B as:

$$\text{Distance}_{A,B} = 1 - \frac{\text{freq}_{A,B}}{\min(\text{freq}_A, \text{freq}_B)}, \tag{C1}$$

where freq_A and freq_B are the number of articles in which journals A and B appear individually. We divide by $\min(\text{freq}_A, \text{freq}_B)$ to account for the maximum number of times in which journals A and B could be cited together. For example, if journal A is cited in 10 articles and journal B in 100, the maximum possible co-appearance is 10. If they co-appear in all 10 of these articles, the cognitive distance then equals 0:

$$\text{Distance}_{A,B} = 1 - \frac{10}{\min(10, 100)} = 0 \tag{C2}$$

Figure C1: Journal distance matrix: heatmap representation for 10 journals



To measure the cognitive distance between a researcher and each potential destination lab, we use the journal distance matrix described above. For each researcher-lab pair, we first identify the set of distinct journals in which the researcher published during the five years prior to mobility. Similarly, we identify the set of distinct journals in which researchers affiliated with the potential destination lab published during the same five-year window. We then use the journal distance matrix to compute the distance for each journal combination. The cognitive distance between researcher i and potential lab l can be defined as:

$$D_{il} = \frac{1}{|J_i^{res}| \times |J_l^{lab}|} \sum_{j \in J_i^{res}} \sum_{k \in J_l^{lab}} d_{jk}, \tag{C3}$$

where D_{il} is a unit-interval measure of distance, 0 indicating full proximity and 1 maximal distance. J_i^{res} is the set of distinct journals in which researcher i published during the five years prior to mobility, J_l^{lab} is the set of journals in which lab l published during the same time window, and d_{jk} is the distance between journals j and k obtained from the journal distance matrix. When no journal-level distance is available for a given journal pair, we assign the maximum distance value of 1. For example, suppose researcher i published in journals A and B, while researchers affiliated with potential lab l published in journals C and D. We retrieve the four distances d_{AC}, d_{AD}, d_{BC}, and d_{BD} from the journal distance matrix. The cognitive distance between researcher i and lab l is then:

$$D_{il} = \frac{d_{AC} + d_{AD} + d_{BC} + d_{BD}}{2 \times 2} \tag{C4}$$

Appendix D. The structure of CNRS

Founded in 1939, the French National Center for Scientific Research (French: *Centre national de la recherche scientifique*, CNRS) is the French state research organization and is one of the largest fundamental science agencies in Europe. CNRS operates based on research units, which are mainly of two kinds: *proper units* (UPRs) are operated solely by CNRS, and *joint research units* (UMRs) are run in association with other institutions, such as universities or other public research agencies (INSERM, INRIA, INRAE, etc.). It is important to mention that members of joint research units may be either CNRS employees, university employees (assistant professor, professor, etc.), or other research agencies members (PhDs, postdocs, etc.). The CNRS also has support units (USR), which may, for instance, supply administrative, computing, library, or engineering services. Each research unit has a unique numeric code. In 2022, CNRS had 927 research units, 210 support units, and 80 international units. It employed around 33,000 staff, including 11,880 tenured researchers, 14,850 engineers and technical staff, and 9,570 contractual workers.[19]

[19] Rapport d'activité CNRS 2022, https://www.cnrs.fr/sites/default/files/page/2023-10/RA_CNRS_2022.pdf

Appendix E. Logit coefficients of the outcome equation

In the paper, we report the average marginal effects of the second-stage outcome equation in Table 2 to interpret the results. In this appendix, we present the corresponding logit coefficients.

Table E1: Logit coefficients of the second-stage outcome equation

	(1)	(2)	(3)	(4)
Relational links				
Citation links	0.446*	1.198***	1.426***	
	(0.243)	(0.319)	(0.331)	
Citation links × low cognitive distance		-1.192***	-1.283***	
		(0.453)	(0.451)	
Researcher-to-lab citations				1.087***
				(0.399)
Researcher-to-lab citations × low cognitive distance				-1.053**
				(0.506)
Co-authorship links	2.052***	1.093**	0.966**	1.245**
	(0.269)	(0.455)	(0.477)	(0.519)
Co-authorship links × low cognitive distance		1.443**	1.435**	1.154*
		(0.579)	(0.591)	(0.618)
Low cognitive distance	0.163	0.252*	0.563***	0.510***
	(0.116)	(0.143)	(0.197)	(0.195)
Lab characteristics				
Geographic distance (in natural logs)			-0.208***	-0.209***
			(0.033)	(0.033)
Difference in lab size$^{(a)}$			0.148	0.196
			(0.143)	(0.130)
Difference in number of publications$^{(a)}$			-0.118	-0.223*
			(0.130)	(0.120)
Difference in admin support$^{(a)}$			0.085	0.082
			(0.088)	(0.087)
Difference in science support$^{(a)}$			-0.064	-0.091
			(0.103)	(0.104)
Paris			-0.381*	-0.350*
			(0.197)	(0.195)
Funding capacity			-0.136	-0.140
			(0.174)	(0.173)
Researcher characteristics				
Research Director (DR)			0.116	0.130
			(0.124)	(0.121)
Female researcher			0.054	0.078
			(0.073)	(0.074)
Age at movement			0.012	0.018*
			(0.009)	(0.010)
Relative productivity (in natural logs)			-0.202***	-0.113***
			(0.047)	(0.050)
Constant	-3.952***	-3.973***	-2.471***	-3.693***
	(0.493)	(0.514)	(0.584)	(0.492)
Correction term λ	0.070	0.062	0.258**	0.188*
	(0.089)	(0.085)	(0.121)	(0.113)
Discipline	No	No	Yes	Yes
Movement year	No	No	Yes	Yes
Pseudo R^2	0.079	0.084	0.118	0.112
Log likelihood	-874.65	-869.66	-837.34	-843.34

Notes: $N = 6,875$. The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. Superscript a indicates standardized variables. Column (1) includes the main independent variables without interactions or controls. Column (2) introduces interactions with low cognitive distance. Column (3) adds researcher and lab controls, as well as discipline and year fixed effects. Column (4) uses researcher-to-lab citations instead of the bidirectional citation links. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix F. Selection equation and sample selection correction

This appendix presents the first-stage selection equation used to correct for potential selection bias that might arise from the fact that destination choices are not random but observed only for researchers who experience a mobility. In the first stage, we estimate a probit of the mobility decision used to calculate the correction term, which will be included as a control term in the second-stage outcome equation.

$$\Pr(\text{Mobility}_i^*) = \Phi\left(\alpha_0 + \mathbf{A}_R R_i + \mathbf{A}_L L_l + \alpha_z Z_i\right) \tag{F1}$$

Let Mobility_i^* denote researcher i's latent propensity to undertake an internal move between CNRS labs in France. $\mathbf{R}_i$ is a vector of variables controlling for researcher characteristics. We include the variables *Female researcher, Age*, and rank *(DR)* of the researcher at the time of recruitment at CNRS. Additionally, we control for the researcher's *Productivity at recruitment* by summing their cumulative number of journal publications before their entry to CNRS, and include the dummy variable *Researcher above average*, which takes the value of one if researcher i's productivity before entry exceeds the average productivity of the other researchers affiliated with the same lab, and zero otherwise. *Cognitive distance* represents the intellectual distance between the researcher's publications and those of the lab of origin. It is computed using publications produced during the period ranging from five years before recruitment to two years after recruitment. Finally, we include *Discipline* and *Year* fixed effects to control for unobserved heterogeneity across disciplines and years.

$\mathbf{L}_i$ is a vector of lab characteristics for the researcher's origin lab. It includes the *Lab size* and the total number of *Administrative* and *Science support* staff observed one year prior to entry. We also proxy for the lab's performance and funding ability with two dummy variables. *Lab above average* takes the value of one if the lab's total number of publications exceeds the average productivity of other labs in the same discipline over the five-year window prior to the entry year, and zero otherwise. *Funding capacity* takes the value of one if the origin lab is affiliated with a university that received an excellence grant in the year preceding entry, and zero otherwise. Lastly, we control for the lab's location with a *Paris* dummy that equals one if the lab of origin is located in Paris, and zero otherwise.

Finally, we include Z_i, an exclusion restriction that affects the selection process (mobility decision) but not the outcome of interest (destination choice). However, finding valid exclusion restrictions is often challenging because Z_i must satisfy both relevance and exclusion (Heckman 1979, Wooldridge 2010). In other words, Z_i must at the same time predict the probability of moving but not the destination choice. We use *Co-authorship origin* as our exclusion restriction. *Co-authorship origin* is a dummy variable that equals 1 if the researcher has co-authored at least one article with a colleague in the lab of origin during the period ranging from five years before recruitment to two years after, and zero otherwise. This variable serves as a proxy for the researcher's attachment to her lab. The rationale behind using it as our exclusion restriction is that strong ties are likely to reduce the probability of moving. However, conditional on moving, co-authorship with colleagues in the origin lab has no direct effect on destination choices. After estimating Equation F1, we compute the correction term, λ_i, and augment the outcome equation (Equation 1) with it to correct for selection bias.

Table F1 presents the summary statistics for the sample used in the first stage. The dataset includes 2,033 CNRS researchers, of whom 10.2% experienced at least one *Mobility* event during the observation period, i.e., 2012–2022. Approximately 34.8% of the sample are *Female researchers*, and the average *Age at recruitment* is 34.2 years. As expected, very few researchers (3.1%) enter CNRS at the rank *DR*, meaning that the majority of our sample joined CNRS early in their research career. Interestingly, the variable *Researcher above average* shows that 85.7% of newly recruited CNRS researchers have a publication record above the average of their peers in the recruitment lab, which also includes university researchers. This suggests that CNRS takes the recruitment process very seriously and values highly productive candidates at the time of recruitment. This is also reflected in the average *Productivity at recruitment*, which is around 20 publications. Finally, 78.8% of the researchers co-authored at least one article with a colleague in the lab of origin, and the average *Cognitive distance* between the researcher's publications and those of the lab of origin is 0.924.

Regarding the lab characteristics, the variable *Lab above average* shows that around 49.2% of CNRS labs have a research productivity level above the disciplinary average. The average *Lab size* is about 117 researchers, ranging from labs with only 2 researchers to very large research units with up to 462 researchers. In terms of resources, labs provide on average 5.1 *Administrative support staff* and around 13.2 *Scientific support staff*, and nearly 70% of the origin labs are affiliated with a university that received an excellence grant. Finally, 20.3% of researchers are recruited into labs located in *Paris*, and the average *Entry year* is 2016.7.

Table F1: Descriptive statistics for the first-stage selection equation

	Mean	S.D.	Min	Max
Mobility	0.102	0.302	0.000	1.000
Female researcher	0.348	0.477	0.000	1.000
Age at recruitment	34.186	4.501	25.000	61.000
Research Director (DR)	0.031	0.175	0.000	1.000
Productivity at recruitment	20.002	28.152	0.000	302.000
Researcher above average	0.857	0.350	0.000	1.000
Co-authorship at origin	0.788	0.409	0.000	1.000
Cognitive distance	0.924	0.059	0.553	1.000
Lab above average	0.492	0.500	0.000	1.000
Lab size	117.090	91.991	2.000	462.000
Admin support	5.113	5.303	0.000	79.000
Science support	13.253	17.775	0.000	234.000
Funding capacity	0.698	0.459	0.000	1.000
Paris	0.203	0.402	0.000	1.000
Entry year	2016.770	2.559	2013.000	2021.000

Notes: $N = 2,033$.

Table F2 reports the marginal effects of the first-stage selection equation (Equation F1), which identifies the determinants of researchers' mobility decisions. Column (1) includes only researchers' demographic characteristics. Column (2) adds researcher characteristics. Column (3) further includes lab characteristics, while column (4) presents the full model by including *discipline* and *year* fixed effects to control for unobserved heterogeneity across disciplines and years. *Entry age* is negative and statistically significant across all columns, suggesting that older researchers are less likely to move than their younger counterparts. Gender on the other hand, is not statistically significant, reflecting the equal mobility opportunities that CNRS offers to all eligible researchers regardless of their gender.[20]

Productivity proxies do not significantly affect mobility decisions in our case, which contradicts the literature showing that more productive researchers tend to be more mobile (Azoulay et al. 2017). However, this is only true in external labor markets where productivity acts as a signal of researchers' quality and facilitates access to new positions. In our case, researchers are simply reallocating within the same organization rather than competing for external positions. In this case, productivity plays a weaker signaling role. The *Co-author origin* variable, which acts as a proxy for the researcher's tie with her current lab, is negatively associated with mobility. In other words, researchers who collaborate with their current peers are less likely to move. Finally, *Cognitive distance* is weakly negatively associated with mobility when lab characteristics are included in column (3), but the effect becomes statistically insignificant in column (4) after controlling for *discipline* and *year* fixed effects. This suggests that knowledge proximity matters mainly across disciplines rather than within the same discipline.

Lab resources do not appear to affect mobility decisions in our case. *Funding capacity, Administrative support*, the *Lab size*, location (captured by the *Paris* variable), and the lab's overall productivity (*Lab above average*) show no significant effects on the probability of moving. The only exception is *Scientific support*, which plays a statistically significant role in mobility decisions. This is consistent with the idea that scientific staff, such as technicians and research assistants, directly contribute to research projects, and a higher level of such support reduces the incentive to move.

[20] https://www.cnrs.fr/en/update/2024-2026-gender-equality-action-plan-aiming-transform-collective-dynamic-tangible-action

Table F2: Average marginal effects of the first-stage selection equation

	(1)	(2)	(3)	(4)
Demographic characteristics				
Female	0.011	0.010	0.008	0.014
	(0.014)	(0.014)	(0.014)	(0.014)
Entry age	-0.004**	-0.004**	-0.005**	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
Researcher characteristics				
DR		-0.001	-0.002	0.000
		(0.049)	(0.049)	(0.049)
Productivity at recruitment (in natural logs)		-0.005	0.003	0.011
		(0.010)	(0.010)	(0.010)
Researcher above average		0.032	0.007	-0.003
		(0.029)	(0.029)	(0.028)
Co-author origin		-0.047***	-0.039**	-0.035**
		(0.018)	(0.018)	(0.018)
Cognitive distance$^{(a)}$		-0.004	-0.013*	-0.005
		(0.007)	(0.007)	(0.008)
Lab characteristics				
Lab above average			0.025	0.004
			(0.017)	(0.018)
Lab size (in natural logs)			-0.017	0.007
			(0.011)	(0.011)
Admin support (in natural logs)			0.000	-0.012
			(0.012)	(0.013)
Science support (in natural logs)			-0.027***	-0.024**
			(0.009)	(0.011)
Funding capacity			-0.011	0.015
			(0.015)	(0.015)
Paris			0.008	0.002
			(0.016)	(0.016)
Discipline	No	No	No	Yes
Entry year	No	No	No	Yes

Notes: $N = 2,033$. The dependent variable is *mobility*, a dummy variable equal to 1 if the researcher undertakes an internal mobility and 0 otherwise. Superscript a indicates standardized variables.Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix G. Additional results, by discipline

In this appendix, we rerun the analysis by discipline without calculating the marginal effects at different values of *Low cognitive distance*. We distinguish between humanities and social sciences (INSHS) and STEM disciplines. Columns (1) and (2) of Table G1 report the average marginal effects based on the same specification as in column (3) of Table 2, estimated separately for INSHS and STEM disciplines, respectively. Columns (3) and (4) report the average marginal effects based on the same specification as in column (4) of Table 2, again for INSHS and STEM disciplines, respectively. The results are largely consistent with those reported in Table 5. The main difference is that citation links in INSHS become insignificant. This result is expected because the reported estimates correspond to the average marginal effects. As shown in Table 5 in the results section, the effect of citation links in INSHS only emerges when the marginal effects are calculated separately at different values of *Low cognitive distance*. The results also confirm that lab characteristics, such as the size and the quality of the lab, appear to give more information for researchers in INSHS, whereas citation and co-authorship links remain the main predictors of destination choices for researchers in STEM disciplines.

Table G1: Average marginal effects of the second-stage outcome equation, by discipline

	(1) INSHS	(2) STEM	(3) INSHS	(4) STEM
Relational links				
Citation links	0.072	0.027*		
	(0.045)	(0.014)		
Researcher-to-lab citations			0.041	0.028*
			(0.046)	(0.016)
Co-authorship links	0.067	0.106***	0.128	0.102***
	(0.066)	(0.026)	(0.097)	(0.026)
Low cognitive distance	0.039*	0.006	0.042*	0.008
	(0.021)	(0.005)	(0.022)	(0.005)
Lab characteristics				
Geographic distance (in natural logs)	-0.008***	-0.006***	-0.007***	-0.006***
	(0.003)	(0.001)	(0.003)	(0.001)
Difference in lab size$^{(a)}$	0.064**	0.003	0.057***	0.002
	(0.023)	(0.004)	(0.021)	(0.004)
Difference in number of publications$^{(a)}$	-0.051**	-0.005	-0.041*	-0.004
	(0.023)	(0.003)	(0.023)	(0.003)
Difference in admin support$^{(a)}$	-0.014	0.004	-0.012	0.004
	(0.012)	(0.003)	(0.012)	(0.003)
Difference in science support$^{(a)}$	0.023	-0.003	0.020	-0.003
	(0.016)	(0.003)	(0.016)	(0.003)
Paris	-0.000	-0.012**	-0.001	-0.012**
	(0.014)	(0.006)	(0.014)	(0.006)
Funding capacity	-0.001	-0.005	-0.004	-0.005
	(0.014)	(0.005)	(0.013)	(0.005)
Researcher characteristics				
Research Director (DR)	0.020	0.006	0.019	0.006
	(0.018)	(0.004)	(0.018)	(0.004)
Female researcher	-0.002	0.004	-0.004	0.004
	(0.009)	(0.003)	(0.009)	(0.003)
Age at movement	0.001	-0.000	0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)
Relative productivity (in natural logs)	-0.009*	-0.004**	-0.008	-0.004**
	(0.005)	(0.002)	(0.005)	(0.001)
Correction term λ	-0.012	0.006	-0.012	0.005
	(0.020)	(0.004)	(0.018)	(0.004)
Discipline	Yes	Yes	Yes	Yes
Movement year	Yes	Yes	Yes	Yes
Number of observations	1,197	5,678	1,197	5,678

Notes: The dependent variable is *choice*, a dummy variable equal to 1 if researcher i chooses lab l as the destination lab and 0 otherwise. INSHS stands for humanities and social sciences. STEM stands for science, technology, engineering, and mathematics. Superscript a indicates standardized variables. Columns (1) and (2) report the average marginal effects based on the same specification as in column (3) of Table 2, estimated separately for humanities and social sciences and STEM disciplines, respectively. Columns (3) and (4) report the average marginal effects based on the same specification as in column (4) of Table 2, again for humanities and social sciences and STEM disciplines, respectively. All four models include a correction term from the first-stage selection equation reported in Table F2 in Appendix F. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix H. Placebo tests

This appendix evaluates whether our main results are driven by mechanical correlations between researcher and lab characteristics rather than by the informational content of citation and co-authorship links. To do so, we implement several placebo tests by randomizing the correlation structure between the destination choice, the relational variables, namely *Citation links, Co-authorship links* and *Low cognitive distance* and the remaining covariates in the model. The objective is to alter any meaningful informational content embedded in these links while preserving the overall structure of the data.

The first placebo test randomizes the destination choice while leaving the matrix of explanatory variables unchanged. For each researcher, we construct a placebo dependent variable by drawing one destination uniformly at random from her set of potential destination labs $\mathcal{C}_i$. This procedure preserves the researcher-specific choice set and researcher-level characteristics, while breaking any systematic relationship between the observed covariates and the realized destination. We then re-estimate Equation 1 on $B = 2000$ placebo samples, storing the coefficient estimates and associated marginal effects. The distribution of placebo estimates should be centered around zero.

$$H_0 : \mathbb{E}\left[\beta_k\right] = 0 \quad \text{for all explanatory variables } k.$$

The alternative hypothesis is that the baseline estimates are driven by systematic differences across researchers that are preserved within choice sets:

$$H_1 : \quad \mathbb{E}[\beta_k] \neq 0.$$

so that, even after randomizing the chosen destination within each researcher's set $\mathcal{C}_i$, the expected coefficients remain different from zero.

Under the null, and across a large number of replications, we expect

$$\frac{1}{B}\sum_{b=1}^{B}\widehat{\beta}_k^{(b)} \to 0 \text{ implying that } \frac{1}{B}\sum_{b=1}^{B}\left(\frac{\partial \Pr}{\partial X_k}\right)^{(b)} \to 0.$$

Failure to reject the null in the placebo exercise is therefore consistent with the interpretation that the baseline results capture meaningful relational links rather than spurious correlations driven by researcher-level heterogeneity.

Importantly, this placebo test is informative about the role of unobserved researcher heterogeneity: since the randomization occurs within each researcher's choice set $\mathcal{C}_i$, it preserves systematic differences across researchers. If the baseline results were driven solely by such differences, the estimated relationships should persist under randomization. By contrast, if they reflect the fact that chosen destinations are more strongly connected to researchers than non-chosen alternatives within a researcher's set, the placebo estimates should collapse toward zero. This placebo therefore mimics the logic of a within-researcher comparison, although it is not a substitute for including researcher fixed effects in the baseline specification.

Column (1) of Table H1 displays the average marginal effects of all three relational links' variables, the standard errors, and the 95% confidence intervals. We observe that none of the parameters is statistically different from zero, indicating that the model does not generate spurious correlations when the outcome is randomly assigned within each researcher's set of potential destination labs. This suggests that the baseline estimates are not driven by mechanical features of the choice-set structure, but instead reflect genuine associations between the explanatory variables and observed destination choices.

Table H1: Placebo marginal effects over $B = 2,000$ placebo replications

	(1)	(2)	(3)	(4)
Citation link	0.001	-0.000	-0.001	0.000
	(0.008)	(0.006)	(0.007)	(0.008)
	[-0.016;0.018]	[-0.013;0.012]	[-0.016;0.015]	[-0.016;0.016]
Co-authorship link	0.004	0.002	0.005	0.005
	(0.014)	(0.011)	(0.015)	(0.015)
	[-0.025;0.033]	[-0.021;0.024]	[-0.025;0.035]	[-0.024;0.034]
Low cognitive distance	0.003	0.003	0.002	0.003
	(0.005)	(0.004)	(0.004)	(0.005)
	[-0.006;0.012]	[-0.005;0.011]	[-0.005;0.010]	[-0.007;0.012]

Notes: Table H1 reports results from Equation 1 pertaining to the three variables of interest: citation links, co-authorship links, and low cognitive distance. The reported results are the average marginal effects, the standard errors in parentheses, and the 95% confidence intervals in square brackets. Column (1) randomizes destination choices within researcher-specific choice sets. Column (2) permutes relational links across potential destinations within choice sets. Column (3) jointly permutes relational links to preserve their correlation structure. Column (4) reassigns researchers across labs while preserving the number of realized matches per lab.

In columns (2) and (3), we implement alternative randomization procedures targeting the relational structure of the data. In column (2), we randomize relational links by permuting *Citation links, Co-authorship links* and *Low cognitive distance* across potential destination labs within each researcher's choice set. This procedure preserves the correlation between destination choices and all researcher and lab characteristics, as well as the marginal distributions of the relational variables, while breaking the alignment between relational links and the true researcher–lab pairs. As a result, any informational content conveyed by relational links is eliminated, although the structure of the choice sets is left unchanged.

Importantly, this placebo also disrupts the joint structure between relational links and cognitive distance, which is central to the interaction effects estimated in the baseline specification. In column (3), we instead preserve the joint correlation structure across the three relational tie variables by permuting them jointly within each researcher's choice set. This procedure maintains the dependence structure among *Citation links, Co-authorship links* and *Low cognitive distance*, as well as their marginal distributions and the composition of choice sets, while eliminating any systematic relationship between these variables and the realized destination choice, including their interaction with cognitive distance. Under the null hypothesis that relational links do not carry meaningful information for destination choices, the corresponding coefficients and marginal effects should be centered around zero in both placebo exercises.

In column (4), we implement a final placebo that reverses the assignment logic by randomizing researchers across labs. Specifically, for each lab, we randomly assign researchers by selecting the same number of realized matches as observed in the data. This procedure preserves the distribution of realized matches at the lab level, while breaking the correspondence between labs and the characteristics of the researchers matched to them.

This placebo should be interpreted as a stringent stress test rather than a direct counterpart to the previous exercises. In contrast to columns (1)–(3), it does not preserve the researcher-specific choice sets that underpin the baseline discrete choice framework, and therefore alters the underlying economic structure of the problem. As a result, the estimates obtained under this randomization are not directly comparable to the baseline specification. Instead, this exercise assesses whether the results rely on the joint matching structure between researchers and labs. In particular, if the baseline findings are driven by systematic alignment between researcher characteristics and destination labs, the estimated relationships should disappear under this reassignment. Accordingly, this placebo complements the previous tests by providing a more demanding benchmark, albeit at the cost of departing from the original choice-set structure.

Across columns (2) to (4), we consistently observe that all estimated marginal effects collapse toward zero and are statistically insignificant. We therefore fail to reject the null hypothesis that, once the alignment between relational variables and realized choices is disrupted, these variables do not systematically explain destination choices. This pattern holds across the different randomization schemes, despite their distinct ways of perturbing the underlying data structure.

Taken together, these placebo tests provide consistent evidence that the baseline results are not driven by mechanical features of the data or by spurious correlations induced by the choice-set structure. When the alignment between destination choices, relational links, and cognitive distance is disrupted, the estimated effects systematically collapse toward zero. This supports the interpretation that the predictive power of citation and co-authorship links reflects meaningful informational content rather than mechanical statistical artifacts.

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2026
GREDEG Working Papers Released in 2026