Information Asymmetry in the Scientific Labor Market: The Role of Citation and Co-authorship Links

Pierre Boutros
Université Côte d'Azur, CNRS, GREDEG (UMR 7321), France
pierre.boutros@etu.univ-cotedazur.fr

Michele Pezzoni
Université Côte d'Azur, CNRS, GREDEG (UMR7321), France
Observatoire des Sciences et Techniques, HCERES, France
michele.pezzoni@univ-cotedazur.fr

Lionel Nesta
Université Côte d'Azur, CNRS, GREDEG (UMR 7321), France
OFCE Sciences Po, France
SKEMA Business School, France
lionel.nesta@univ-cotedazur.fr

Sonia Paty
Université Lumière Lyon 2, CNRS, GATE (UMR 5824), France
sonia.paty@cnrs.fr

Abstract: This paper investigates the factors predicting the destination choice of mobile researchers. To do so, we use a unique dataset on researcher mobility between labs within France's largest public research organization from 2012 to 2022. We find that citations to scientific works and co-authorship between mobile researchers and 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 cognitive distance between the researcher and the lab is high, whereas co-authorship plays a stronger role when 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.

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

JEL codes: D83, J61, J24

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 and 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 and 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 referring to 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 and 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 and 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.

The match between a researcher and a lab can be viewed as a process of knowledge combination between two parties (Kogut and 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 and 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 and 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 and Levinthal, 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 and 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 and Gibbs, 2014). This situation is commonly referred to in the labor literature as employee-employer matching under conditions of imperfect information (Burdett and 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 and Pissarides, 1994). 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 (Spence, 1978).

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 and 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 and 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 citations and co-authorship 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.

[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

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 CNRS researchers, university faculty members, postdoctoral fellows, PhD students, engineers, and technical staff. 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. CNRS allows for internal mobility between labs to ensure skill alignment, support career development, and increase job satisfaction among its employees. It is easy for CNRS researchers to transfer from one lab to another. CNRS allows for "*au fil de l'eau*" mobility, where researchers can initiate their own moves between different locations independently.[4]

Our empirical analysis relies on three data sources. First, we use administrative records from CNRS 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[5] 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-

[3] https://www.cnrs.fr/en/le-cnrs/organisation
[4] https://www.cnrs.fr/sites/default/files/pdf/DRH2020-11_LDG-mobilite.pdf
[5] 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.

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,[6] 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.

3.2 Model and variables

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

$$\begin{aligned}\Pr(Choice_{i,l}) = \Lambda(&\beta_0 + \beta_1 Citation\ link_{i,l} + \beta_2 Coauthorship\ link_{i,l} \\ &+ \beta_3 Citation\ link_{i,l}\ X\ Low\ cognitive\ distance_{i,l} \\ &+ \beta_4 Coauthorship\ link_{i,l}\ X\ Low\ cognitive\ distance_{i,l} \\ &+ \beta_5 Low\ cognitive\ distance_{i,l} + \beta_6 L_l + \beta_7 R_i + \beta_8 \lambda_i),\end{aligned}$$

Equation 1: Destination choice equation

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 explained by two main explanatory variables, i.e., *Citation link*$_{i,l}$ and *Co-authorship link*$_{i,l}$ between the researcher and the potential destination lab before mobility. Specifically, *Citation link*$_{i,l}$ is a dummy variable that equals one if at least one citation link exists between researcher *i* and all researchers affiliated with lab *l* during the five years prior to mobility, and zero otherwise. Co-*authorship link*$_{i,l}$ is also a dummy variable that equals one if researcher *i* has co-authored at least one paper with researchers affiliated with lab *l* during the five years prior to mobility, and zero otherwise. Both variables are interacted with a proxy of the cognitive distance between researcher *i* and all researchers affiliated with lab *l*. Specifically, we calculated the variable *Low cognitive distance* as a dummy variable equal to one if the cognitive distance is below the cognitive distance

[6] 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.

sample mean, and zero otherwise. The cognitive distance measure is based on the distance between the journals referenced in the publications of mobile researchers and the those referenced by researchers affiliated with the destination lab 5 years prior to the mobility. The journal distance is computed using a co-citation matrix that captures how often pairs of journals are cited together in the same article[7]. In this matrix, journals that are more frequently co-cited are considered cognitively closer, and vice versa.

The vectors L_l and R_i in Equation 1 are vectors of variables controlling for the lab and researcher characteristics, respectively. L_l include the *Geographical distance* that represents the distance between the origin and the potential destination lab *l* in kilometers.[8] 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[9] that held an excellence grant[10] in the year preceding mobility*,* and zero otherwise.

As researcher characteristics R_i, we control for the career rank using CNRS research fellow categories, *CR* (*Chargé.e de recherche*) and *DR* (*Directeur.trice de recherche*)[11], 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

[7] Appendix C provides additional details on the construction of this measure.
[8] 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).
[9] In France, a single university hosts many research labs across disciplines. Affiliation here refers to the lab's formal institutional link to the university.
[10] 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.
[11] 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.

heterogeneity across disciplines and years by including *Discipline* fixed effects corresponding to CNRS institutes[12] and *Year* fixed effects capturing the time of the move.

Studying destination choices might raise a sample selection problem (Dahl, 2002; Verginer and 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 distance* and *Geographical 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.

[12] 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).

Table 1 – Descriptive statistics

	Choice=0					Choice=1				
	N	Mean	sd	min	max	N	Mean	sd	Min	Max
Citation links	6,658	0.135	0.342	0	1	217	0.371	0.484	0	1
Co-authorship links	6,658	0.029	0.169	0	1	217	0.258	0.439	0	1
Low cognitive distance	6,658	0.401	0.490	0	1	217	0.516	0.501	0	1
Geographical distance	6,658	381.847	231.614	0	1,029.905	217	285.375	244.295	0	958.361
Difference in lab size	6,658	-3.459	105.790	-439	419	217	-3.127	101.711	-301	319
Difference in nb of publications	6,658	-22.740	429.258	-1,965	1,950	217	-38.761	399.727	-1,643	1,570
Difference in admin support	6,658	-0.236	6.182	-29	34	217	-0.028	5.774	-26	22
Difference in science support	6,658	0.378	15.768	-89	108	217	0.254	15.700	-79	72
Paris	6,658	0.227	0.419	0	1	217	0.235	0.425	0	1
Funding capacity	6,658	0.659	0.474	0	1	217	0.615	0.488	0	1
DR	6,658	0.122	0.327	0	1	217	0.127	0.333	0	1
Female researcher	6,658	0.340	0.474	0	1	217	0.371	0.484	0	1
Age at movement	6,658	37.691	4.399	31	55	217	38.019	4.624	31	55
Relative productivity	6,658	3.804	5.883	0	181.138	217	4.050	8.039	0	95.464
Movement year	6,658	2019.580	2.049	2014	2022	217	2019.610	2.001	2014	2022
λ	6,658	1.548	0.341	0.768	2.721	217	1.562	0.344	0.768	2.721

4. Results

Table 2 reports the average marginal effects[13] 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.[14]

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 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*

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

[14] 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.

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. *Geographical 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

Dependent variable: choice	(1)	(2)	(3)	(4)
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				
Geographical distance (ln)			-0.006***	-0.006***
			(0.001)	(0.001)
Difference in lab size (std)			0.004	0.005
			(0.004)	(0.004)
Difference in nb of publications (std)			-0.003	-0.006*
			(0.004)	(0.003)
Difference in admin support (std)			0.002	0.002
			(0.002)	(0.002)
Difference in science support (std)			-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				
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 (ln)			-0.006***	-0.003***
			(0.001)	(0.001)
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
Number of observations	6,875	6,875	6,875	6,875

Notes: 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-researcher citations instead of the bidirectional citation links. All four models include a correction term from the first-stage selection equation reported in Table E1 in Appendix E.

Variables labelled *ln* have been transformed using the natural logarithm, and variables labelled *std* have been standardized.

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

4.1 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 3 reports the average marginal effects of our main independent variables while including all control variables. Columns (1) and (2) are estimated using the same specification as in column (3) of Table 2, while the average marginal effects are computed separately for *Low cognitive distance* equal to 0 and 1, respectively. We find that the two information channels play different roles depending on the level of cognitive distance between the researcher and the potential destination lab. When *Low cognitive distance* equals 0, citation links are the only predictor of destination choice, while prior co-authorship links do not show any statistical significance. However, when *Low cognitive distance* equals 1, prior co-authorship becomes the only significant predictor, whereas citations no longer have a significant effect. This suggests that each information channel, i.e., citations and prior co-authorship, addresses a different type of information asymmetry depending on the level of cognitive distance. Specifically, when researchers work on distant topics (*Low cognitive distance* = 0), the main source of information asymmetry concerns whether the knowledge held by one party is relevant to that of the other. In this case, citation links are highly informative because they indicate not only mutual awareness of each other's work, but also that the knowledge produced by one party is useful for the research activities of the other, even when cognitive distance is high. A citation from the researcher to the potential destination lab indicates that the researcher draws on that lab's scientific output. A citation from the potential destination lab to the researcher indicates that the lab relies on that researcher's knowledge stock. This information appears to be sufficient for researchers to make an informed decision. While co-authorship can still occur in this context, it likely reflects complementarity on a specific project and does not predict destination choice. In other words, a researcher may collaborate with another working on a distant

topic simply because they need that researcher's specific expertise to complete the project. On the other hand, when researchers work on closely related topics (*Low cognitive distance* = 1), both parties already understand each other's scientific output, and the main source of information asymmetry concerns the tacit aspects of a collaboration, that is, whether the two parties can work together effectively. In this context, prior co-authorship provides sufficient information, while citation links no longer add any relevant information. Similar results are observed in columns (3) and (4), which report the average marginal effects estimated using the same specification as in column (4) of Table 2, evaluated separately for when *Low cognitive distance* equals 0 and 1, respectively.

Table 3 – Average marginal of the second-stage outcome equation, by low cognitive distance

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	0.052***	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)
Controls	Yes	Yes	Yes	Yes
Number of observations	6,875	6,875	6,875	6,875

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

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[15] and report the average marginal effects separately for cases where *Low cognitive distance* equals 0 and 1. Columns (1) and (2) of Table 4 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.

[15] 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), humanities and social sciences (INSHS), ecology & environment (INEE), nuclear & particle physics (IN2P3), earth & universe sciences (INSU), and mathematical sciences (INSMI).

Our results show that the mechanism discussed in Table 3 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 3. 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 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; Larivière et al., 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; Larivière et al., 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 4 – Average marginal of the second-stage outcome equation, by low cognitive distance and discipline

Dependent variable: choice	(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*	0.003	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)
Controls	Yes	Yes	Yes	Yes
Number of observations	1,197	1,197	5,678	5,678

Notes: INSHS stands for social sciences and humanities. 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

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.

Important implications for labor allocation emerge from this paper. Because destination choices determine where knowledge is being allocated, helping researchers to make informed decisions is crucial. Our results imply that improving allocation is unlikely to come from a "one-size-fits-all" policy. The same information can be highly informative in one context and less relevant in another. Organizations, therefore, should not assume that providing more information will automatically improve destination choices, while the right type of information matters more. Indeed, we show that information channels play different roles depending on the cognitive distance and discipline.

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.

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. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative science quarterly, 35(1), 128-152.

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.

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, 397-433.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: Journal of the econometric society, 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.

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

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.

Levin, S. G., & Stephan, P. E. (1999). Are the foreign born a source of strength for US science?. Science, 285(5431), 1213-1214.

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. (2000). Learning and innovation in organizations and economies. OUP Oxford.

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.

Spence, M. (1978). Job market signaling. In Uncertainty in economics (pp. 281-306). Academic Press.

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

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.

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

Zucker, L. G., Darby, M. R., & Brewer, M. B. (1998). Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. The American Economic Review, 88(1), 290–306. http://www.jstor.org/stable/116831.

Appendix

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 and A2 remain stable.

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

Dependent variable: choice	(1)	(2)	(3)	(4)
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				
Geographical distance (ln)			-0.006***	-0.006***
			(0.001)	(0.001)
Difference in lab size (std)			0.006	0.005
			(0.004)	(0.004)
Difference in nb of publications (std)			-0.007*	-0.006*
			(0.003)	(0.003)
Difference in admin support (std)			0.002	0.002
			(0.002)	(0.002)
Difference in science support (std)			-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				
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 (ln)			-0.004**	-0.003**
			(0.001)	(0.001)
Discipline	No	No	Yes	Yes
Movement year	No	No	Yes	Yes
Number of observations	6,875	6,875	6,875	6,875

Notes: 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-researcher citations instead of the bidirectional citation links.
Variables labelled *ln* have been transformed using the natural logarithm, and variables labelled *std* have been standardized.
Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table A2 – Average marginal 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)
Controls	Yes	Yes	Yes	Yes
Number of observations	6,875	6,875	6,875	6,875

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

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**

Note: CNRS has ten institutes: 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

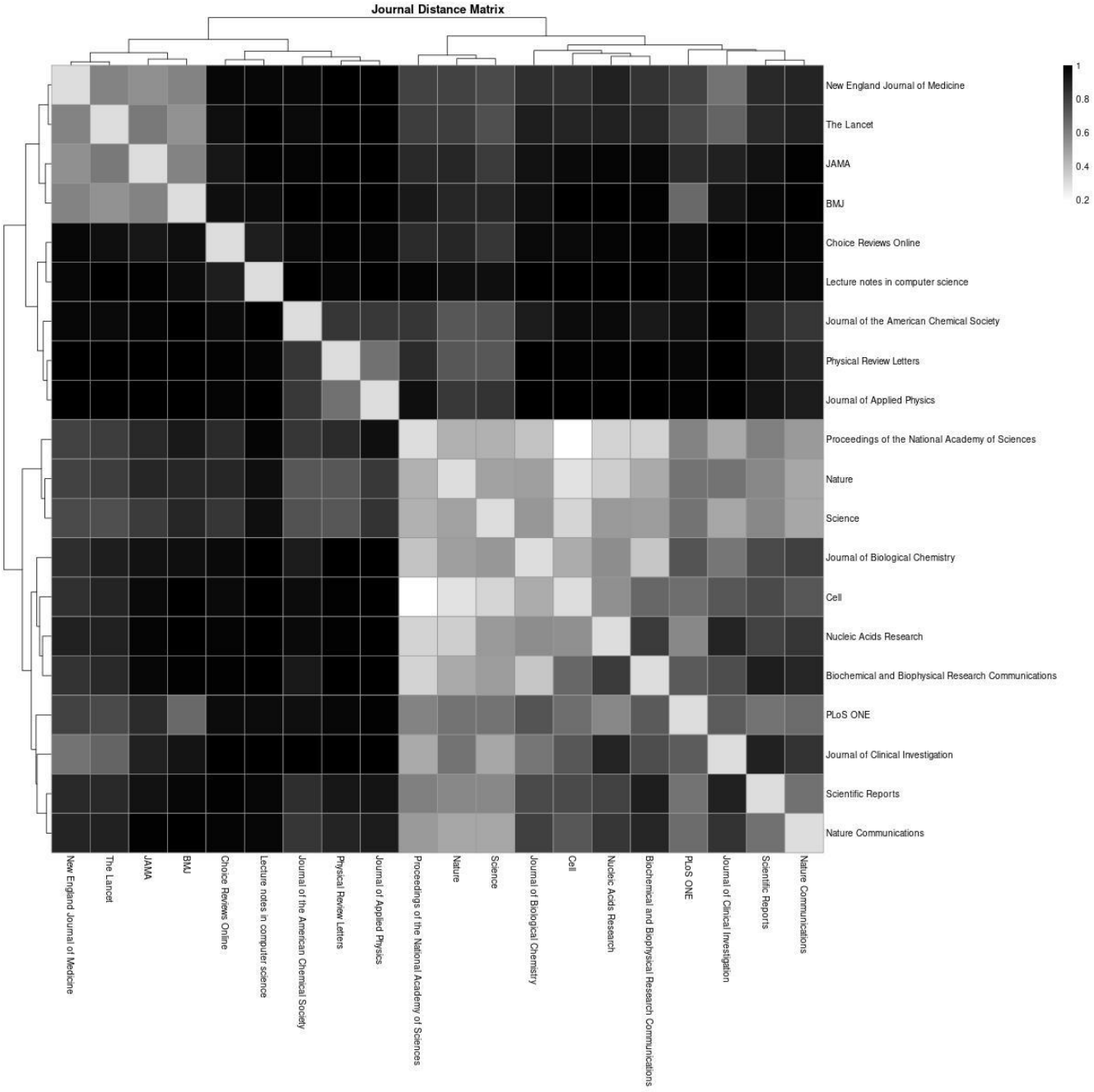
In Appendix C, we describe the construction of the journal distance matrix used to measure the cognitive distance between researchers and labs. The goal is to quantify how similar two journals are in terms of content. Figure C1 represents a heatmap of the distances between 20 journals. Darker colors indicate larger distances, while lighter colors indicate smaller distances. The figure is extracted from a larger matrix that includes the distances between all possible journal pairs formed by the 8,305 journals cited in the reference list of a random sample of 20,000 articles published by researchers affiliated with CNRS labs. 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 $freq_{A,B}$, 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:

$$Distance_{A,B} = 1 - \frac{freq_{A,B}}{\min\left(freq_A, freq_B\right)}$$

where $freq_A$ and $freq_B$ are the number of articles in which journals A and B appear individually. We divide by $\min\left(freq_A, freq_B\right)$ 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:

$$Distance_{A,B} = 1 - \frac{10}{\min(10, 100)} = 0$$

Figure C1 – Journal distance matrix, a graphical representation for 20 journals



Appendix D: The structure of CNRS

Founded in 1939, the French National Centre 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 – French: *Unité Mixte de Recherche*) 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.[16]

[16] 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

Dependent variable: choice	(1)	(2)	(3)	(4)
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				
Geographical distance (ln)			-0.208***	-0.209***
			(0.033)	(0.033)
Difference in lab size (std)			0.148	0.196
			(0.143)	(0.130)
Difference in nb of publications (std)			-0.118	-0.223*
			(0.130)	(0.120)
Difference in admin support (std)			0.085	0.082
			(0.088)	(0.087)
Difference in science support (std)			-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				
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 (ln)			-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)
Discipline	No	No	Yes	Yes
Movement year	No	No	Yes	Yes
Correction term	0.070	0.062	0.258**	0.188*
	(0.089)	(0.085)	(0.121)	(0.113)
Pseudo R^2	0.079	0.084	0.118	0.112
Log likelihood	-874.65	-869.66	-837.34	-843.34
Number of observations	6,875	6,875	6,875	6,875

Notes: 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-researcher citations instead of the bidirectional citation links.
Variables labelled *ln* have been transformed using the natural logarithm, and variables labelled *std* have been standardized.
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.

$$\Pr\left(Mobility_i^*\right) = \Phi(\alpha_0 + \alpha_1 R_i + \alpha_2 L_l + \alpha_3 Z_i)$$

Equation F1: First-stage selection equation

Let $Mobility_i^*$ denote researcher *i*'s latent propensity to undertake an internal move between CNRS labs in France. 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.

L_l 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

	N	Mean	sd	min	max
Mobility	2,033	0.102	0.302	0	1
Female researcher	2,033	0.348	0.477	0	1
Age at recruitment	2,033	34.186	4.501	25	61
DR	2,033	0.031	0.175	0	1
Productivity at recruitment	2,033	20.002	28.152	0	302
Researcher above average	2,033	0.857	0.350	0	1
Co-author origin	2,033	0.788	0.409	0	1
Cognitive distance	2,033	0.924	0.059	0.553	1
Lab above average	2,033	0.492	0.500	0	1
Lab size	2,033	117.090	91.991	2	462
Admin support	2,033	5.113	5.303	0	79
Science support	2,033	13.253	17.775	0	234
Funding capacity	2,033	0.698	0.459	0	1
Paris	2,033	0.203	0.402	0	1
Entry year	2,033	2016.770	2.559	2013	2021

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 offer to all eligible researchers regardless of their gender.[17]

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

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

the researcher's tie with her current lab, is negatively associated with mobility. In other words, researchers who have a collaboration 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.

Table F2 – 2

Dependent variable: mobility	(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 (ln)		-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 (std)		-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 (ln)			-0.017	0.007
			(0.011)	(0.011)
Admin support (ln)			0.000	-0.012
			(0.012)	(0.013)
Science support (ln)			-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)
Fixed effects				
Discipline	No	No	No	Yes
Entry year	No	No	No	Yes
Number of observations	2,033	2,033	2,033	2,033

Notes: Variables labelled *ln* have been transformed using the natural logarithm, and variables labelled *std* have been standardized.
Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

References for Appendix F

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

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: Journal of the econometric society, 153-161.

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

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 4. 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 4 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 discipline.

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

Dependent variable: choice	(1) INSHS	(2) STEM	(3) INSHS	(4) STEM
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				
Geographical distance (ln)	-0.008***	-0.006***	-0.007***	-0.006***
	(0.003)	(0.001)	(0.003)	(0.001)
Difference in lab size (std)	0.064**	0.003	0.057***	0.002
	(0.023)	(0.004)	(0.021)	(0.004)
Difference in nb of publications (std)	-0.051**	-0.005	-0.041*	-0.004
	(0.023)	(0.003)	(0.023)	(0.003)
Difference in admin support (std)	-0.014	0.004	-0.012	0.004
	(0.012)	(0.003)	(0.012)	(0.003)
Difference in science support (std)	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				
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 (ln)	-0.009*	-0.004**	-0.008	-0.004**
	(0.005)	(0.002)	(0.005)	(0.001)
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: Notes: INSHS stands for social sciences and humanities. STEM stands for science, technology, engineering, and mathematics.
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 E1 in Appendix E.
Variables labelled *ln* have been transformed using the natural logarithm, and variables labelled *std* have been standardized.
Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Acknowledgments: *We thank participants of the Workshop on the Organization, Economics, and Policy of Scientific Research (WOEPS), Milan, Italy. We are also grateful to CNRS for providing the dataset.*

Competing interest: *The authors disclose no relevant financial interests related to the research in this paper.*