Build or Buy?
Human Capital and Corporate Diversification

Paul Beaumont Camille Hebert Victor Lyonnet*

June 6, 2022

Abstract

Why do some firms enter a new sector by building on their resources ("build") while others buy an existing company ("buy")? Using French administrative data, we propose a measure of human capital distance between a firm and a sector of entry. We show that firms build in close sectors and buy in distant sectors in terms of human capital. We establish causality using a shift-share instrument. Firms build by hiring new workers, which is more costly in distant sectors because it requires more organizational skills. Hence, firms buy in distant sectors to acquire already operational human capital.

Keywords: Diversification, M&As, Human capital, Labor market frictions.

JEL codes: G34, L25, J24.

*Paul Beaumont is with McGill University (paul.beaumont@mcgill.ca). Camille Hebert is with the University of Toronto (camille.hebert@rotman.utoronto.ca). Victor Lyonnet is with the Ohio State University (lyonnet.1@osu.edu). We thank Manuel Adelino, Pat Akey, Sébastien Betermier, Alice Bonaime, Claire Célérier, Gilles Chemla, Edith Ginglinger, Andrey Golubov, Denis Gromb, Johan Hombert, Kai Li, Julien Martin, Adrien Matray, Isabelle Méjean, Paige Ouimet, Gordon Phillips, Manju Puri, Amiyatosh Purnanandam, David Robinson, Christoph Schneider, Matthew Serfling, Merih Sevilir, Rui Silva, René Stulz, Geoffrey Tate, Vish Viswanathan, Ramona Westermann, Mike Wittry, Liu Yang, and Sheng-Jun Xu for valuable comments and helpful discussions. We also thank seminar participants at CREST Paris, the French Bureau of Statistics (INSEE), Université Paris-Dauphine, Tilburg University, HEC Paris, Duke University, the Ohio State University, University of Toronto, McGill University, PUC Chile, UQAM, the 5th ECGC workshop on Governance and Control (Best Student Paper Award), the UT Austin Corporate Finance forum, the Mitsui Labor and Finance symposium, NFA, SDU workshop on the Firm's Life Cycle, RCEA Growth, Innovation and Entrepreneurship conference, the Junior Entrepreneurial Finance lunch group, UNC Private Equity spring symposium, the CEPR-Bank of Italy workshop on Ownership, Management & Firm Performance, AFA, SFS Cavalcade, MFA and FIRS. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'avenir" program (reference: ANR-10-EQPX-17). We are responsible for all remaining errors.

1. Introduction

Theories of the firm, dating back to Penrose (1955) and Teece (1982), propose that successful product diversification is an engine of corporate growth. By expanding their scope, firms overcome the limits to growth from operating in a single sector (Maksimovic and Phillips, 2002; Giroud and Mueller, 2015; Tate and Yang, 2015). However, our understanding of how firms expand their scope of activities is limited. How do firms acquire the resources they need to operate in a new sector? Why do some firms expand organically by acquiring new resources to complement their existing ones ("build"), while others buy an existing company to acquire these resources ("buy")? In this paper, we study the tradeoff between these two ways of acquiring new resources to better understand the frictions that firms face in expanding their scope.

We focus on human capital as a key resource of the firm for two reasons. First, unlike physical capital, human capital cannot be bought "off-the-shelf" outside of an acquisition. Although firms can hire new workers, they face significant costs to integrate these workers into their existing workforce, assigning them to tasks, and training them to acquire firm-specific knowledge (Prescott and Visscher, 1980). As a result, the acquisition of sector-specific human capital is likely to be a major barrier of entry into new sectors. Second, the existing literature shows that human capital is a major determinant of the decision to diversify (Neffke and Henning, 2013) and of mergers and acquisitions (M&As) across industries (Tate and Yang, 2016; Lee, Mauer and Xu, 2018).[1] However, little is known about the role of human capital in the decision to build or buy.

Using French administrative datasets linked to commercial M&A datasets, this paper starts by documenting that 98% of entries into new sectors (85% when weighting by entry sales) are made by building on existing resources, suggesting that firms only buy when they cannot build.[2] Our main result is that firms build in close sectors and buy in more distant ones. We establish causality by exploiting variation in firms' distance to the sector of entry that are due to changes in the workforce composition of incumbent firms (shift-share design).

[1] An important strand of the M&A literature is based on the "property rights" theory (Grossman and Hart, 1986; Hart and Moore, 1990), according to which value is created by combining complementary assets under the control of a single firm. Rhodes-Kropf and Robinson (2008) show that the search for complementarities implies the existence of an assortative matching between acquirers and target firms, i.e.,"like buys like". Mergers are more likely to occur and create more value when merging firms sell similar products (Hoberg and Phillips, 2010), use similar technology (Bena and Li, 2014), or share similar corporate cultures (Li, Qiu and Shen, 2018). However, this literature has been mostly silent on how firms acquire new assets and resources beyond M&As.

[2] In the Capital IQ data, cross-industry M&As represent 55% of all M&A deals globally in 2021, for a total transaction value of $1.231 billion.

Our study is based on a novel dataset of build and buy entries in France from 2003 to 2014. We rely on administrative data that contain the detailed breakdown of firm sales by sector to identify the universe of entries into new sectors. We define entries at the business group level: a firm enters a new sector at time t if its sales in this sector across all subsidiaries are positive at t but equal to zero at $t-1$.[3] We then exploit data on M&A deals from SDC Platinum and Bureau van Dijk's Zephyr to determine whether the firm enters by acquiring an existing company. The resulting dataset consists of 61,000 build or buy entries.

We develop a new measure of human capital (HC) distance between a firm and a potential sector of entry. Specifically, we use the French matched employer-employee dataset that includes a detailed occupation code at the employee level and the exact employment composition of all firms in France. Following Lazear (2009), we define a firm's human capital as a combination of occupation-specific skills. This definition has the advantage of encompassing several components of the human capital of the entire workforce – including education, experience, and sector-specific skills. We first compute the share of each worker occupation in the wage bill of every firm. Then, we construct the HC distance measure at the firm level, based on the overlap between each firm's shares and those of incumbents in the sector of entry. HC distance ranges from zero (no overlap) to one (perfect overlap). In contrast to existing sector-level measures, ours uses the detailed composition of firms' workforce and is at the firm level.[4] Therefore, our HC distance measure allows us to test whether differences in human capital across firms explain why different firms from the same sector of origin choose to enter the same sector of entry in different ways.

Our main result is that firms build in close sectors and buy in more distant sectors in terms of HC distance. A one standard deviation increase in HC distance is associated with a 20% increase in the unconditional probability of buying. Our baseline specification compares firms that operate in the same sector of origin and enter the same sector of entry in the same year, so that the effect is not confounded by unobservable, time-varying factors that would affect firms' choice to build or buy within sector pairs. This specification includes other possible firm-level determinants of the decision to build or buy, such as firm size, profitability, capital intensity, and access to finance (Matvos, Seru and Silva, 2018; Berg, Streitz and Wedow, 2020).

[3]We define sectors at the 3-digit level of the French classification of industries, which corresponds to 272 different sectors, and check that our results remain unchanged when using a different level of industry classification. We require that sales in the new sector represent at least 1% of the firm's total sales the year before entry to ensure that the entry is economically meaningful from the firm's perspective. To avoid false positives, we also discard entries that might in fact correspond to changes in the coverage of the data.

[4]Other papers have proposed measures of human capital transferability and relatedness at the sector level (Farjoun, 1994; Neffke and Henning, 2013; Tate and Yang, 2016; Lee, Mauer and Xu, 2018).

We perform a battery of robustness checks to rule out alternative explanations. The relationship holds when we include firm fixed effects, which rules out alternative explanations based on unobservable differences between firms that build and firms that buy. Given that firms choose to build or buy *conditionally on* diversifying in a given sector, we examine and rule out several possible alternative mechanisms related to the existence of a selection bias. In addition, we show that our main finding is robust to using alternative definitions of HC distance, different entry thresholds in the new sector, and to the inclusion of additional firm controls (e.g., firm age, size, or degree of diversification). The results hold in different subsamples of firms (e.g., public versus private firms, cash-poor versus cash-rich firms) or sectors (e.g., degree of competition, growth of the sector), suggesting that the role of human capital in the decision to build or buy is not specific to certain firms but is widely relevant in the cross-section of firms and sectors. Finally, we show that firms' choice to build or buy is not solely explained by the minimization of physical capital adjustment costs, input-output linkages, product market distance, or geographical proximity between firms and the sector of entry (Hoberg and Phillips, 2010; Bloom, Schankerman and Van Reenen, 2013; Ma, Murfin and Pratt, 2021).

Although the granular nature of our data allows us to include a host of control variables and fixed effects, it is still possible that other time-varying unobservable factors impact both HC distance and the decision to build or buy. To address this concern, we adopt a shift-share design (Bartik, 1991; Goldsmith-Pinkham, Sorkin and Swift, 2020; Borusyak, Hull and Jaravel, 2022) and construct an instrumental variable, called HC Bartik, that captures the variation in a firm's HC distance due to changes in the workforce composition of firms in the sector of entry. For instance, the share of worker occupations in the wage bill of incumbent firms in the *Business services* sector evolved in favor of IT workers from the beginning to the end of our sample period (the "shock").[5] Our instrument predicts that firms with a high share of IT workers in their wage bill at the beginning of our sample (the "share") should see a decrease in their HC distance to the *Business services* sector. Consistent with our main results, we find that firms buy when their (instrumented) HC distance to the sector of entry increases over time. Assuming that long-term changes in the human capital of incumbents in the sector of entry do not correlate with unobserved determinants of firms' decision to build or buy, this approach identifies the

[5]The occupation *Engineers and managers in administration, maintenance, support and user services in IT* went from 0.7% of the wage bill in 2004 to 2.9% in 2012. This is the largest change in wage shares of the wage bill across all occupations in the *Business services* sector.

causal effect of HC distance on firms' on the decision to build or buy.[6]

Next, we analyze the tradeoff faced by firms when deciding whether to build or buy in a new sector. On the one hand, firms can build on their existing human capital by hiring new workers to complement their workforce. We show that the more distant firms are to the sector of entry, the more workers they hire to build in this sector. Firms build by hiring workers in the top 5 occupations for the sector of entry, allowing them to reduce their distance to the sector. Despite this reduction in HC distance, firms that build underperform when they are initially more distant do the sector of entry: sales and survival probability in the sector of entry are negatively correlated with HC distance up to three years after entry. These findings show that it is more costly for firms to build in more distant sectors, suggesting that buying becomes a more attractive option than building when HC distance is high enough.

On the other hand, firms can buy to acquire the human capital they cannot build organically. We show that firms that buy adjust their target's workforce by retaining workers in the top 5 occupations for the sector of entry and downsizing employment in other occupations. This finding is consistent with the idea that firms buy to acquire and retain the human capital they lack (Ouimet and Zarutskie, 2020; Chen, Gao and Ma, 2021) and let go of other types of workers (Dessaint, Golubov and Volpin, 2017; Lagaras, 2020). Although an acquisition entails important acquisition costs, we show that buying allows firms to perform well in the sector of entry regardless of their initial HC distance. Therefore, firms that are distant enough to the sector of entry prefer to buy to avoid the high cost of building by hiring many workers.

We confirm the importance of human capital in the choice to build or buy by showing that HC distance correlates more strongly with the decision to build or buy when hiring new workers is more costly. Specifically, we use occupation-level data obtained from the French unemployment agency to create a measure of local labor market tightness based on the availability of the key occupations the firm needs to enter the new sector.[7] We show that the relationship between the decision to build or buy and HC distance only holds when firms enter sectors in which key occupations are in short supply in their local labor market. This finding confirms that the cost

[6] In line with our example, these results are particularly strong for sectors in which the share of the wage bill that goes to IT workers increased the most over the sample period. This source of identifying variation is consistent with existing evidence on the increasing importance of IT in the employment composition of French firms during our sample period (Harrigan, Reshef and Toubal, 2016; Lashkari, Bauer and Boussard, 2018).

[7] Occupations are flagged as being in "short supply" if they meet two conditions: (i) job postings for this occupation exceed job applications, and (ii) surveyed employers anticipate that they will not fill a position in this occupation. The data report occupations in short supply across 350 local labor markets.

of hiring new workers plays a key role in a firm's decision to build or buy. Firms buy when they cannot build the human capital they need.

We explore the reasons why distant firms prefer to buy an existing company rather than hiring workers to build in the new sector. The organizational economics literature suggests that building on existing human capital requires more than hiring new workers (see, e.g., Prescott and Visscher, 1980; Dessein and Prat, 2022): Firms also need organizational skills to train and integrate these new workers into their existing processes. To test whether a firm's organizational skills explain why it prefers to buy rather than build when it is distant to the sector of entry, we construct three measures of organizational skills: (i) the fraction of a firm's wage bill that goes to workers in the "top layers" of the hierarchy (Caliendo, Monte and Rossi-Hansberg, 2015), (ii) to workers associated with Selling, General and Administrative (SG&A) expenses (Eisfeldt and Papanikolaou, 2013; Li, Qiu and Shen, 2018), (iii) and a dummy equal to one if a firm employs any worker in human resources (i.e., the firm has an HR service). With each of these measures, we find that the relationship between HC distance and the decision to build or buy only holds when firms have low organizational skills. These results suggest that building by hiring new workers is costly because it requires organizational skills. The more distant the sector of entry, the more organizational skills are required. Therefore, firms buy in distant sectors to avoid the high cost of building, thereby acquiring already operational human capital from their target.

Finally, we study how the decision to build or buy affect firms' subsequent entry decisions. Because the decision to build or buy in the new sector leads firms to adjust their human capital differently upon entry, this decision has implications for firms' ability to seize subsequent growth opportunities. Both in the cross-section (comparing firms that buy to firms that build) and in the time series (within firms that buy), we find that firms build in more distant sectors after a buy entry than after a build entry, so that the initial HC distance becomes less predictive of firms' subsequent entry decisions after a buy entry. This finding supports the idea that buying allows firms to not only enter new sectors they could not enter by building but also to subsequently seize growth opportunities that would not have been available without buying first.

This paper adds to the growing literature on the role of human capital in M&As. The novelty of our approach is to consider build entries as an alternative to M&As. Our contribution is threefold. First, we find that firms build more often than they buy, suggesting that firms only buy when they cannot build. Second, our analysis shows that firms buy in sectors that are not

too distant in terms of HC distance, but more distant than the sectors firms build in. Thus, our paper reconciles the existing evidence that diversifying M&As are more likely to occur between firms that share similar human capital (Tate and Yang, 2016; Lee, Mauer and Xu, 2018) and other evidence that firms buy to acquire the human capital they lack (Ouimet and Zarutskie, 2020; Chen, Gao and Ma, 2021). Third, our results shed light on a new source of value creation in M&As beyond rent extraction (Lagaras, 2021; He and le Maire, 2022) or the exploitation of synergies between workforces (Tate and Yang, 2015; Dessaint, Golubov and Volpin, 2017; Ma, Ouimet and Simintzi, 2018; Lagaras, 2020). We show that organizational skills are required for firms to build on their existing human capital and that buying allows firms to acquire already operational human capital when they lack these skills.

Our paper also relates to the literature on the determinants of the decision to build or buy across sectors. Few papers have jointly studied these alternative approaches to diversification.[8] McCardle and Viswanathan (1994) theoretically show that firms buy in a new sector when barriers to entry are high. Phillips and Zhdanov (2013) explore the tradeoff between investing in R&D and buying an existing innovative firm. Bernard, Redding and Schott (2010) document that multiproduct firms often change their product portfolio but infrequently do so through an acquisition (only 7% of cases). We contribute to this literature by studying the choice to build or buy using an exhaustive dataset covering all French firms. We document that buy entries represent only 2% of entries in new sectors (15% when weighted by sales) and that the ability to acquire key resources (i.e., human capital) to operate in the sector of entry is a major determinant of this decision.

Finally, our paper relates to the body of work exploring the determinants of the firm's scope of activity, starting with Lucas (1978). Research in corporate finance (see, e.g., Maksimovic and Phillips, 2002; Schoar, 2002) and international trade (see, e.g., Bernard, Redding and Schott, 2010; Eckel and Neary, 2010) shows that the number of sectors (or products) of the firm is determined by its ability to be productive in a wide range of activities.[9] Hoberg and Phillips (2018, 2020) and Boehm, Dhingra and Morrow (2019) show that firms operate in sectors that

[8] The literature on corporate diversification primarily focuses on the choice of sector in the decision to diversify (see Maksimovic and Phillips, 2013, for a review of the literature). The strategy literature has studied this question more extensively (see, e.g., Yip, 1982; Lee and Lieberman, 2010; Capron and Mitchell, 2012). In the international economics literature, Nocke and Yeaple (2007) study the choice of entering a new country via foreign direct investment versus acquiring an existing company.

[9] A related strand of the literature studies the creation of value by conglomerate firms through the reallocation of capital (Shin and Stulz, 1998; Schoar, 2002; Giroud and Mueller, 2015, 2019) or labor (Tate and Yang, 2015; Cestone et al., 2018) in response to demand shocks.

require similar inputs. We also find that firms are more likely to enter sectors that are not too distant in terms of HC distance. Our contribution is to show that firms build in close sectors and buy in more distant ones, and that the choice to build or buy is driven by the cost for a firm to acquire new resources required for the sector of entry.

2. Data and Descriptive Statistics

Our dataset consists of a combination of commercial M&A data with administrative data available at the French Bureau of Statistics (Insee).

2.1. Data sources and main variables

2.1.1 M&A deals

We collect M&A deals occurring between January 2003 and December 2014 from SDC Platinum and Bureau van Dijk's Zephyr. We exclude leveraged buyouts and private equity deals from the sample. We focus on majority deals in which the acquirer owns strictly less than 50% of the target shares before the acquisition and at least 50% thereafter. We retain deals that involve a French acquirer and a French target to be able to match them with the French administrative data. SDC Platinum and Bureau van Dijk's Zephyr do not provide the standardized French firm identifiers. To retrieve firm identifiers, we build a Python webcrawler that searches for identifiers based on (acquiring and target) firms' names and addresses.[10] The webcrawler allows us to identify firms involved in M&A deals at the business group level. Our final sample includes 7,165 deals from 2003 to 2014 involving 4,139 acquiring firms.

[10]First, we use tickers (available only for publicly traded firms) and the Bureau van Dijk identifiers (available only for deals in Zephyr) to retrieve the French firm standard identifiers (siren). Second, we build a Python webcrawler that retrieves firms' standard identifiers by searching for firms' names and addresses on two specialized websites: www.bodacc.fr (Bulletin Officiel des Annonces Civiles et Commerciales), which is a governmental website that reports official notifications since 2003, and www.societe.com, which is a commercial website that aggregates information about French companies from various sources (mostly from the French Bureau of Statistics and Bodacc.fr). Both websites are supposed to cover the universe of French firms. Third, after running the webcrawler, we drop companies for which the address, city, and zip code are missing because we cannot identify with certainty the corresponding company identifier among several matches. We retain only observations for which the Jaró-Winkler string distance between the original name and the retrieved name is above 0.8. Fourth, we manually check the resulting matches. The next step consists of finding the business group's standard identifier (sirtg) using the ownership link dataset (LIFI), as we run our analysis at this level. Indeed, it is difficult to precisely identify with a company name which entity of a business group is involved in the M&A deal. Business groups' entities often share a common company name with their parent company. In addition, note that we cannot use changes in ownership links to identify M&As deals in our data, as we cannot distinguish between newly reported ownership links in the database and changes in existing ownership links. Finally, we discard LBO deals from the sample.

2.1.2 Firm-level data

Ownership link dataset. Our analysis is at the business group level (thereafter "firm"). A firm includes the parent company and majority-owned subsidiaries connected through ownership links.[11] We place the analysis at the firm-group level, as decisions to diversify are made at this level and may involve resource reallocation within the group. We use the ownership links dataset (*Enquête sur les Liaisons financières entre sociétés*, LIFI) to retrieve the structure of firms and identify all entities linked to the diversifying firm and to identify all the subsidiaries of companies targeted in M&A deals. We treat these subsidiaries as if they were themselves acquired along with their parent company.[12]

Tax files. The tax files (*Bénéfices Industriels et Commerciaux*) provide detailed yearly accounting information (balance sheet and income statements) at the subsidiary level.[13] We retrieve cash holdings, total sales, total assets, tangible assets, and the firm's main sector of activity from the tax files. Sectors are defined by the French classification of sectors (*Nomenclature des activités Francaises, NAF*). We define sectors at the 3-digit industry level, which includes 272 different sectors. We use the firm registry (*SIRENE*) to obtain the geographic location of firms' establishments. Consistent with the approach on firms involved in M&A deals, we consolidate all variables at the business group level.

Breakdown of sales by sector. To identify entries in new sectors, we use the subsidiaries' breakdown of sales by sector (*Enquête Annuelle des Entreprises, EAE* from 2003 to 2007 and then *Ventilation des Ventes par Activité, VAC* from 2008 to 2014). This dataset records the firms' comprehensive amounts of sales realized in every sector every year. The data are exhaustive for firms with at least 20 employees and randomly includes smaller firms. The survey covers 85% of aggregate sales and 96% of sales in the manufacturing sector. For smaller subsidiaries for which the sales breakdown across sectors is not available, we assume that these firms sell

[11] French administrative data include three levels of entities: the business group or firm-level, identified with a *sirtg* identifier, the subsidiary level, identified with a *siren* identifier, and the establishment level, identified with a *siret* identifier.

[12] In addition, we check that the target of an M&A deal in our sample becomes one of the acquirer's subsidiaries after the date of the M&A deal as registered in Zephyr/SDC Platinum. We replace the date of the M&A deal with the effective link's date observed in the ownership link dataset. If the gap between the event date and the link date is larger than two years, we discard the M&A deal and the involved firms from the sample.

[13] We exclude from our sample firms in the financial, agricultural and public sectors because they use different accounting standards. We focus on firms using the standard tax regime (*Régime normal*), which represent on average 94% of total value-added. We discard firms that use the simplified regime (*Régime simplifié*) because some important information is not reported by these firms.

only in their main sector of activity, which is observed directly from the tax files.

We define entries at the business group level, such that a firm enters a new sector in a given year when its sales in this sector are positive for the first time in that year across all of its subsidiaries. We observe whether this entry is due to the acquisition of an existing company (buy) or to the organic entry of one or several of its existing subsidiaries (build). We impose an entry threshold to ensure that the entry is economically meaningful from the firm's perspective. We require that sales in the new sector represent at least 1% of the firm's total sales the year before the entry. We also ensure that buy entries are not the byproduct of horizontal acquisitions by removing buy entries in which the acquirer and target realized more than 50% of their sales in the same sectors before the acquisition.[14] We check the robustness of our results to these definitions in Appendix Table A3.

2.1.3 Matched employer-employee dataset

We use the French matched employer-employee dataset (*Déclarations Annuelles des Données Sociales*, DADS) to observe firms' workforce composition. When filing yearly payroll taxes, employers are required to report detailed information about their employees: gross and net wages, the number of hours worked, the type of contract, and a detailed occupation code for each employee. Occupations are reported as 4-digit code occupations (*Nomenclatures des professions et catégories socio-professionnelles des emplois salariés des employeurs privés et publics*, PCS-ESE). There are 414 different occupations at this level, including, for instance, 28 different types of engineers (e.g., logistics, IT, electrical, or mechanical).[15]

We use this information to construct a firm-level measure of HC distance to the sector of entry:

$$\text{HC distance}_{g,n} = 1 - \frac{\sum_i s_{g,i} \cdot s_{n,i}}{\sqrt{\sum_i s_{g,i}^2}\sqrt{\sum_i s_{n,i}^2}}, \tag{1}$$

where $s_{g,i}$ is the share of the wage bill of firm g that goes to workers employed in occupation i and $s_{n,i}$ the share of the wage bill in sector n that goes to workers employed in occupation

[14] To avoid false positives, we also discard entries that might in fact correspond to increased coverage of the LIFI or VAC datasets. First, we discard entries due to changes in ownership links in LIFI for which we do not find a corresponding M&A transaction in SDC Platinum or Zephyr. Second, we discard entries due to a subsidiary entering the VAC survey.

[15] Note that reporting the occupation code is required for firms with at least 20 employees and is optional for firms below this threshold. Therefore, we restrict our sample to firms with at least 20 employees.

i.[16] We define HC distance as one minus the cosine similarity between vectors s_g and s_n. The vector s_n represents the workforce composition of a "representative firm" resulting from the consolidation of all single-sector firms in sector n.[17]

The variable HC distance$_{g,n}$ ranges from 0 to 1, with 1 being the maximum HC distance between firm g and sector n. HC distance is close to one when the firm's wage bill is tilted towards those worker occupations that are heavily represented in the sector of entry and close to zero when it is not. Intuitively, HC distance measures to what extent firm g's production function (as reflected by its workforce composition) is close to the production function of a representative firm in sector n. Another way to understand HC distance is that it yields a unique Hotelling-like human capital space between firms and sectors (see Hoberg and Phillips, 2016, 2020, for examples of similar measures).[18]

2.1.4 Local labor market tightness

We use information collected by the French national unemployment agency (*Pôle emploi*) to assess the role of LLM tightness in the build or buy decision. The data, available from 2010 to 2014, identify 350 different LLMs, which correspond to commuting zones. An occupation is flagged as being "in short supply" if it meets two conditions: (i) job postings for this occupation exceed job applications, and (ii) surveyed employers anticipate that they will not fill a position in this occupation. We use the list of occupations "in short supply" by LLM to construct a sector- and LLM-specific measure of LLM tightness as follows:

$$\text{LLM tightness}_{z,n,t} = \sum_i 1(\text{Occupation in high demand})_{i,z,t} \cdot s_{n,i,t}$$

[16] For simplicity, we omit the t subscripts, although HC distance varies between years. Note that because we do not observe the exact allocation of employees across the firm's different sectors of activity, we only retain firms that operate only in sector n to calculate $s_{n,i}$.

[17] An alternative method would be to compute the distance between firm g and each incumbent, single-sector firm in sector n and to take the (weighted or simple) average distance. We replicate our main results using this alternative method in Appendix Table A2. In robustness tests, we also define a sector-level measure of HC distance to replicate our results at the sector level. Instead of using the share $s_{g,i}$ of wages that goes to workers employed in occupation i for firm g, we use the share of wages $s_{o,i}$ in a given sector of origin o that goes to occupation i. The measure becomes HC Distance$_{o,n} = 1 - \frac{\sum_i s_{o,i} \cdot s_{n,i}}{\sqrt{\sum_i s_{o,i}^2}\sqrt{\sum_i s_{n,i}^2}}$.

[18] As an illustration, suppose that in aggregating all the firms operating in the civil engineering sector, (1) we find that the aggregate workforce in the sector is evenly split across three occupations: engineers (E), salesmen (S), and managers (M) ($s_{1,E} = s_{1,S} = s_{1,M} = 1/3$). Now, consider two firms A and B, for which employees are both evenly split across two occupations. Firm A's workforce is such that $s_{A,E} = s_{A,S} = 1/2$, and firm B's workforce is such that $s_{B,E} = 1$. In that case, firm A's HC distance to the civil engineering sector (HC Distance$_{A,1} = 0.24$) is smaller than firm B's (HC Distance$_{B,1} = 0.43$). Instead, if firm B's workforce were such that $s_{B,E} = s_{B,M} = 1/2$, then the two firms' human capital would be equidistant to the civil engineering sector.

where $1(\text{Occupation in high demand})_{i,z,t}$ is a dummy variable equal to one if occupation i is in high demand in LLM z and $s_{n,i}$ is the share of employees employed in occupation i in sector n. A high value of LLM $\text{tightness}_{z,n,t}$ implies that firms in LLM z are likely to face difficulties finding workers in occupation i that are key to operating in sector n. For each entry in a new sector n, we use the value of LLM $\text{tightness}_{z,n}$ for the LLM z where the subsidiary that enters is located.[19]

2.2. Descriptive statistics

Build or buy? Panel A of Table 1 presents the evolution of the proportion of build and buy entries between 2005 and 2014. While at the beginning of the period, approximately 1.3% of entries are made by acquisition, this figure increases over the sample period to reach 2.25% of total entries in 2014. Buy entries are, on average, larger than build entries. When weighting by entry sales, buy entries represent 15.68% of total entries between 2005 and 2014. In our baseline analysis, we identify entries into new sectors at the 3-digit level. The proportion of buy entries remains stable at 1.6-1.8% when using the other French industry classification levels (Panel B). On average, buy entries represent approximately 2% of entries, so 98% are build entries.

[Insert Table 1 here]

HC distance. We find that the average HC distance between a diversifying firm and the sector of entry is 0.79, suggesting that the average firm lacks employees in occupations that are key to operating in the sector of entry. In Appendix Table A2, we test our results' robustness to alternative measures of HC distance. Our baseline measure of HC distance is based on the overlap between the shares of the diversifying firms' wage bill going to each worker occupation and those of incumbents in the sector of entry. The distribution of HC distance remains stable when we use the shares of the firms' workers and the shares of the number of hours worked in the firm as alternative specifications. This finding suggests that the distribution of the number of workers and hours worked by occupations is similar to the distribution of wages by occupation. We also calculate the average HC distance between the diversifying firm and incumbents in the sector of entry. The (simple or weighted) average HC distances are larger than the baseline

[19] In the case of multiplant firms located in several LLMs, we assume that the entry is made in the LLM in which the subsidiary has the most employees.

measure, suggesting that incumbents in the sector of entry exhibit significant heterogeneity in their human capital composition.

In the Appendix Figure A1, we report the HC distance between sectors at a one-digit sector classification in 2014, HC Distance$^{\text{Sector}}_{\text{o,n,t-1}}$, defined by equation (B). Each cell of the heatmap corresponds to the HC distance between the two sectors located on the cell's corresponding row and column. For instance, the *Administration and Support Services sector*, on row one, has a human capital distance to the *Food and Beverages* sector of 0.93. Darker cells indicate more distant sectors, whereas lighter cells indicate closer sectors. Interestingly, some sectors are remote from all other sectors because they employ specialized worker occupations that are not found elsewhere (e.g., the *Accommodation and Food Services* sector).[20]

Other variables. Table 2 reports summary statistics on the variables used in our analysis. The average firm in our sample employs 45 workers ($= e^{3.81}$), hired in 16.7 different occupations ($= 45 \times 0.37$). It produces approximately €60,000 in value-added per worker, owns €70,000 in tangible assets per employee, and holds €30,000 in cash per worker. The average firm invests €85,000 at the time of the entry. Of the diversifying firms, 55% were already diversified before the entry into the new sector. Only 2% of them include a listed subsidiary in the business group.

Firms report on average €0.5 million in sales in the sector of entry in the first year, with a large dispersion around the mean. Moreover, 16% of the entries identified at the 3-digit of the French SIC happen to be in the same 2-digit industry, and 33% happen to be in the same 1-digit industry. In addition, we find that 23% of the entries occur in an upstream sector while 16% occur in a downstream sector. Only 3% of the entries occur in a different *région* (a coarser administrative area), and 6% occur in a different *département* (a finer administrative area). These stylized facts suggest that the average firm diversifies in a distant sector from its sector of origin but in the same geographic area.

[Insert Table 2 here]

[20] As an illustration, Table A1 in the Appendix shows the top occupations for three sectors in our data, including the *Accommodation and Food Services* sector.

3. Empirical Strategy

3.1. Human capital and the decision to build or buy

The literature shows that M&As are more likely to occur between firms with similar human capital (e.g., Tate and Yang, 2016; Lee, Mauer and Xu, 2018). However, firms need not enter a new sector through an acquisition. Instead, firms have a choice between building and buying, and the evidence in Section 2.2 shows that firms choose to build in the vast majority of cases (98% of entries). Motivated by this evidence, we ask whether human capital can explain why some firms enter a new sector by building on their existing human capital ("build") while others buy a company already operating in the sector ("buy").

In our main specification, we investigate whether the mode of entry in a new sector is systematically related to the HC distance between the diversifying firm and the sector of entry (see Section 2.1.3):

$$\mathbb{1}(\text{Buy})_{g,n,t} = \lambda_{n,o,t} + \delta \text{HC Distance}_{g,n,t-1} + \beta X_{g,n,t-1} + \varepsilon_{g,n,t}, \tag{2}$$

where the dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable equal to one if firm g buys an incumbent company in the sector of entry n in year t and equal to zero if the firm builds in sector n (see Section 2.1.2). The main independent variable, HC Distance$_{g,n,t-1}$, is our measure of the HC distance between firm g and the sector of entry n in year $t-1$ (see Section 2.1.3). The vector of control variables $X_{g,n,t-1}$ includes other firm characteristics that may influence the choice to build or buy. Control variables include the firm size measured as the total (log) number of workers, the value-added, number of occupations, total cash holdings, tangible assets, and total wages. We scale all control variables (except firm size) by the number of employees. Appendix B contains the definitions of the variables.

We use a set of sector of origin $\times$ entry $\times$ year fixed effects ($\lambda_{n,o,t}$) to identify the effect of HC Distance$_{g,n,t-1}$ on the decision to build or buy across firms that enter the same sector of entry n in the same year t and operate in the same sector of origin o. These fixed effects capture unobservable synergies between sectors, as well as unobservable factors related to the sector of entry (e.g., fixed costs and barriers to entry) or the sector of origin (e.g., demand shocks) that could vary over time. We double-cluster standard errors at the sector of origin and sector

of entry levels to account for correlations of errors across firms within the sector of origin and within the sector of entry, respectively.

3.2. Bartik instrument

A potential concern with specification (2) is that there might be unobservable variables associated with both the distance to the sector of entry and the mode of entry in the new sector. For instance, firms may anticipate the mode of entry and hire employees in key occupations before entering the new sector.

To establish a causal link between firms' human capital and their decision to build or buy, we need to identify changes in HC distance that are not related to the firm's decision to build or buy in the new sector. Given that HC distance is based on the overlap of occupations between diversifying firms and incumbents in the sector of entry, we need a source of exogenous variation either in the workforce composition of new entrants or in the workforce composition of incumbents in the sector of entry.

Our approach focuses on incumbents in the sector of entry. The workforce composition of French firms experienced significant changes during our sample period, specifically because of the increasing importance of information technologies (Harrigan, Reshef and Toubal, 2016; Lashkari, Bauer and Boussard, 2018). Holding the diversifying firms' workforce fixed at their composition at the beginning of the sample period, in 2003, we isolate variation in HC distance that are due to changes in incumbent firms' workforce composition over time. We define the Bartik instrument as follows:

$$\text{HC Bartik}_{g,n} = \sum_i \hat{s}_{g,i,03} \cdot \Delta\hat{s}_{n,i,03,11}, \tag{3}$$

where $\Delta\hat{s}_{n,i,03,11} = \hat{s}_{n,i,11} - \hat{s}_{n,i,03}$ is the change in the (normalized) share of occupation i in sector n from 2003 to 2011. The normalized share $\hat{s}_i$ is $s_i/\sqrt{\sum_j s_j^2}$. We replace $\text{HC Distance}_{g,n,t-1}$ with $\text{HC Bartik}_{g,n}$ in equation (2) and estimate the equation on the sample of entries occurring between 2011 and 2014, after the time period we use to estimate long-term changes in human capital. We can rewrite the definition of HC Bartik as follows:

$$\text{HC Bartik}_{g,n} = -(\text{HC distance}^{03}_{g,n,11} - \text{HC distance}^{03}_{g,n,03}), \tag{4}$$

where HC distance$^{t_0}_{g,n,t} = 1 - \sum_i \hat{s}_{g,i,t_0}$, and $\hat{s}_{n,i,t}$ measures the HC distance of firm g to sector n taking firm g's occupation shares $\hat{s}_{g,i,t_0}$ at their value at t_0. This formulation shows that HC Bartik corresponds to the opposite of the change in HC distance that is due to the long-term sector-level changes in workforce composition, holding the occupation shares of g at their 2003 value. Therefore, a high value of HC Bartik means that sector n has become less distant from firm g.

The identifying assumption is that there is no sorting of firms into occupations such that there are firm-level unobservable characteristics correlated with both long-term changes in the human capital of incumbents in the sector of entry and the mode of entry in the new sector (Borusyak, Hull and Jaravel, 2022).

3.3. Other empirical concerns

To mitigate the role of omitted variables in the decision to build or buy, we estimate equation (2) on a matched control group of firms that build or buy. We construct the matched sample of firms based on the sector of origin, the sector of entry, and quartiles of firm size before entry, measured by realized sales in the sector of origin the year before the entry. This matching allows us to compare build and buy entries within the same (sector of origin, entry, size quartile) triplet. In an alternative specification, we add firm fixed effects in equation (2) to exploit variation in HC distance within firms across sectors.

Another potential concern is the presence of unobservable latent variables that may drive the choice of the sector of entry *and* the decision to build or buy. For example, firms may make the decision to build or buy *jointly* with the decision to enter a given sector. We test whether there is a systematic link between our measure of HC distance and the choice of the sector of entry in Appendix Table A9. Specifically, we estimate

$$\mathbb{1}(\text{Diversify})_{g,n,t} = \lambda_{g,t} + \lambda_{o,n,t} + \rho \cdot \text{HC Distance}_{g,n,t\text{-}1} + \mu X_{g,n,o,t-1} + \eta_{g,n,o,t} \tag{5}$$

where $\mathbb{1}(\text{Diversify})_{g,n,t}$ is a dummy variable equal to one if firm g operating in the sector of origin o enters sector n in year t, zero otherwise. $\lambda_{o,n,t}$ are sector of origin $\times$ entry $\times$ year fixed effects. In some specifications, we include firm $\times$ year fixed effects $\lambda_{g,t}$ to capture unobservable firm-level time-varying determinants of the decision to diversify. We show in Section 7.2 that

our results still hold when selection mechanisms are unlikely to be at play.

4. Main Results

4.1. Human capital and the decision to build or buy

We begin the analysis by investigating what drives firms' decision to build or buy. Using our model described in equation (2), we test whether firms are more likely to buy when their human capital is distant from the sector of entry and to build when their human capital is close. Table 3 reports the results with different specifications and sets of fixed effects.

The baseline specification in column 1 includes sector of origin $\times$ sector of entry $\times$ year fixed effects. This specification allows us to compare firms operating in the same sector of origin o, entering the same new sector n in the same year t while controlling for unobservable synergies between the sector of origin and that of entry. We show that firms' HC distance to the sector of entry is positively correlated with the probability of acquiring an incumbent in the sector of entry. A one standard deviation increase in HC distance is associated with a 0.39 percentage-point increase in the probability of buying (column 1). This relationship is sizable, equal to approximately 20% of the unconditional probability of buying, and significant at the 5% level. Regarding the control variables, we find that large firms, firms with a more diverse mix of occupations and less financially constrained firms are more likely to buy.

In columns 2 and 3, we test the robustness of our main finding to different combinations of fixed effects. In column 2, we include interacted sector of origin $\times$ entry $\times$ quartile of firm size fixed effects to compare firms of similar size in the sector of origin that enter the same sector. In column 3, we compare firms that initially operate in the same *main* sector of origin and same *second* sector of origin and enter the same sector. This specification accounts for unobservable synergies between the firms' main and secondary sectors of activity and the sector of entry. The HC distance coefficient is robust to these alternative combinations of fixed effects, and its value remains very similar in these specifications.

In column 4, we add firm and year fixed effects, in addition to the interacted sector fixed effects. The HC distance coefficient remains positive and significant at the 5% level, suggesting that within the same firm, we observe build entries in close sectors and buy entries in distant ones. This result shows that the positive relationship between buy entries and HC distance

cannot be fully explained by the presence of firm-level omitted variables.[21]

[Insert Table 3 here]

4.2. Robustness tests

Alternative measures of HC distance. In Appendix table A2, we test the robustness of our main findings to alternative measures of HC distance. Our baseline measure of HC distance is based on the overlap between the shares of the diversifying firms' wage bill going to each worker occupation and those of incumbents in the sector of entry. This measure may overestimate the weight of a few high-paid individuals (e.g., CEOs) and underestimate the weights of numerically important occupations (e.g., part-time workers). In Appendix Table A2 panel A, columns 2 and 3, we define a firm's HC distance as the cosine distance between the vectors of the number of workers and number of hours worked by occupation. In panel B, instead of considering the distance to the aggregate distribution of firms' wage bill in the sector of entry, we define HC distance as the average distance between the diversifying firm and incumbents in the sector of entry. We weight incumbents in different ways (equally weighted, total sales, number of employees) to account for their importance in the sector of entry. Consistent with our main results, we find that firms build in close sectors and buy in more distant sectors.

Alternative thresholds in the sector of entry. In Appendix table A3, we test the robustness of our results to the definition of an entry in the new sector. In column 1, we do not impose any minimal entry size. Thus, entries corresponding to sales that amount to €1 are included in the sample. In columns 2 to 4, we impose an entry size threshold equal to 1% (baseline), 5%, and 10% , respectively. The magnitude of the effect of HC distance on buy entries is economically and statistically stronger as we impose a higher threshold of entry sales, suggesting that the decision to buy or build is more sensitive to HC distance when the entry in the new sector is meaningful. When the entry is small, possibly driven by experimental motives, buying an incumbent firm is a less favored option, so that the workforce composition of firms is not as crucial. We confirm this intuition in column 5 by running a weighted least squares (WLS) regression, where observations are weighted by entry sales.

[21] Note that the relationship between the firm's size and the decision to buy becomes negative when we add firm fixed effects. This suggests that firms that diversify several times over the course of their life-cycle tend to buy at a relatively earlier stage when they are relatively smaller and less profitable than when they build (e.g., declining firms).

Sector-level analysis. Our main specification compares firms from the same sector of origin that enter the same sector in the same year to control for unobservable time-varying factors that could affect the mode of entry of firms within sector pairs. Our estimations are silent on the magnitude of entry flows across sectors. In Appendix Table A4, we exploit cross-industry variation to test how aggregate buy and build flows relate to the HC distance between sectors. For this test, we compute the total number of build and buy entries by pair of sector of origin and sector of entry and calculate a sector-level HC distance between the sector of entry and the sector of origin. The empirical specification includes sector of origin, entry, and year fixed effects, as well as sector-level control variables.[22] Column 1 shows that entries are more common between sectors with similar human capital. Columns 2 and 3 show that both build and buy entries are less likely between sectors with a high HC distance and that this effect is stronger for build entries than for buy entries. As a result, we find in column 4 that the fraction of buy entries over build entries increases with sector-level HC distance.

HC distance within firms, across subsidiaries. In the baseline tests, our definition of the firm includes all the subsidiaries under the ownership and control of the same parent company. In Appendix Table A5, we change the unit of observation and calculate HC distance at the subsidiary level. We test whether, within the firm, the entry in the new sector is made by subsidiaries that are close to the sector of entry in terms of human capital. We expect the entry in the new sector to be made through the subsidiary that employs the right set of occupations to minimize reallocation costs in the internal labor market.[23] Our results support our hypothesis that within the firm, entry is more likely to be made through subsidiaries with a low HC distance to the sector of entry. This finding not only validates that our measure of HC distance explains firms' entry decisions but also suggests the presence of labor reallocation frictions in internal labor markets.

[22] The control variables of the sector-level estimations include a dummy equal to one if the two sectors belong to the same two-digit industry of the French classification of industries, the intensity of downstream and upstream links between the two sectors, and a product market distance between the sector of origin and the sector of entry.

[23] We test this prediction for build entries only. It is unclear whether the human capital of the subsidiaries of firms that buy should predict which subsidiary would acquire and take control of the target. We estimate the following equation for build entries at the subsidiary level:

$$\mathbb{1}(\text{Build})_{f,n,t} = \lambda_{g,n,t} + \beta \cdot \text{HC distance}_{f,n,t\text{-}1} + \gamma \cdot \text{X}_{f,n,t\text{-}1} + \epsilon_{f,n,t}$$

The dependent variable is a dummy variable that takes value one if the entry into the new sector is made through subsidiary f, zero otherwise. We include business group $\times$ sector of entry $\times$ year fixed effects to compare the different subsidiaries' HC distance to the sector of entry within the business group g. This specification excludes standalone firms because it relies on the variation in HC distance between subsidiaries.

4.3. Bartik instrument

As explained in Section 3.2, we exploit long-term changes in the human capital of incumbent firms in the sector of entry to establish a causal link between firms' human capital and their decision to build or buy. We construct an HC Bartik instrument that captures the variation in HC distance between a firm and a sector of entry that is due to changes in incumbent firms' human capital over time. A positive value of HC Bartik means that the wage bill composition of incumbents in the sector of entry became more similar to that of the diversifying firm in 2003 between 2003 and 2011. Note that the sample size is smaller than the main specification because we focus on entries made after 2011. The results are contained in Table 4.

Column 1 shows that firms are more likely to build in sectors that have become closer over time. This result holds if we replace sector of origin $\times$ entry fixed effects with sector of origin $\times$ entry $\times$ year fixed effects as in the main specification (column 2), although the coefficient is only significant at the 10% level. In column 3, we check that the mode of entry cannot solely be predicted by the firm's past human capital as of 2003. For instance, this would be the case if the incumbent firms' human capital was highly persistent over time. We show that controlling for HC distance as of 2003 does not affect the coefficient of interest, suggesting that the variation over time in incumbent firms' human capital is critical for its identification.

Following Harrigan, Reshef and Toubal (2016), who find that the workforce composition of French firms experienced significant changes due to the rise of information technologies, we find that the HC Bartik coefficient is statistically significant only in sectors in which the share of the wage bill of IT workers increased more than the sample median between 2003 and 2011 (columns 4 and 5). This finding implies that the variation in human capital on which we rely for identification are linked to the rise of IT at the beginning of the 2000s (see also Lashkari, Bauer and Boussard, 2018).

A potential concern with the interpretation of the results is that firms that already employed IT workers in 2003 differ on unobservable dimensions (e.g., level of innovation). Under this alternative explanation, the negative HC Bartik coefficient would not reflect the fact that firms build more in sectors that became closer to them but instead that firms with IT workers in 2003 build more in sectors that happened to have experienced a rise in IT during the sample period. We rule out that concern in column 6 by showing that the coefficient of interest remains unchanged even after controlling for a dummy equal to one if the firm was employing IT

workers in 2003. This additional test lends credence to the identifying assumption of our Bartik instrument.

[Insert Table 4 here]

5. Evidence of the Role of Human Capital

In this section, we provide several pieces of evidence to show that firms buy in the sector of entry when building on their existing human capital is too costly. First, we show that firms build on their existing human capital by hiring new workers to complement their workforce and that firms buy to acquire and retain their target's human capital. Second, we show that firms that build underperform when they are initially more distant to the sector of entry, whereas firms that buy perform well in the sector of entry regardless of their initial HC distance. Third, we show that HC distance is only correlated with the decision to build or buy when firms enter sectors in which key occupations are in short supply in their local labor markets. Fourth, we show that the relationship between HC distance and the decision to build or buy only holds when firms have low organizational skills, consistent with the idea that firms buy to acquire the resources they cannot build organically.

5.1. Post-entry labor adjustments

How do firms adjust their workforce when entering a new sector? The graphs above Table 5 capture the employment dynamics of high HC distance firms compared to low HC distance firms four years before to four years after entry. We identify "high HC distance firms" by computing a dummy equal to one if the firm's HC distance to the sector of entry is above the median in each sector of origin × entry × year triplet. We study build and buy entries separately because we expect post-entry employment dynamics to differ depending on whether firms build or buy.[24] We focus on the employment dynamics of the subsidiary that enters in the new sector. We use

[24]This graph plots the result of the regression

$$log(Employment)_{f,t} = \alpha_f + \beta_t + \lambda_{o,n,t_0} + \sum_j \delta_j 1(High\ HC\ distance)_{f,t} \cdot 1(t = j) + \epsilon_{f,t}$$

where f denotes the subsidiary of group g that enters the new sector n at time t_0, $log(Employment)_{f,n,t}$ is the logarithm of the number of workers employed by subsidiary f at time t, α_f denotes subsidiary fixed effects, β_t denotes year to entry fixed effects, λ_{o,n,t_0} denotes sector of origin × entry × year of entry fixed effects, and $1(High\ HC\ Distance)$ is a dummy equal to one if the firm is above the median of HC distance taken by sector of origin × entry × year triplet. The graph plots the δ_j coefficients.

HC distance calculated at the firm level as the main independent variable.[25]

The graphs show that high and low HC distance firms have similar employment dynamics before entry, for both build and buy entries. This finding implies that firms do not significantly restructure their workforce in anticipation of entry. After a build entry, the employment of high HC distance firms grows by 2 to 3% more than for low HC distance firms. By contrast, we do not observe any significant post-entry differences between low and high HC distance firms after a buy entry.

[Insert Table 5 here]

We further analyze these post-entry labor dynamics in Table 5.[26] In column 1 of Panel A, we regress the employment growth from year $t-1$ to $t+3$ on HC distance for firms that build. We compare firms that operate in the same sector of origin and build in the same sector of entry in the same year. We find that a one standard deviation increase in pre-entry HC distance is associated with a 1.6 p.p. increase in employment growth around the entry in the new sector.

In columns 2 and 3, we decompose the change in the employment level of firms that build. We break down this change between the 5 most important occupations for the sector of entry (in terms of shares of the wage bill) and the other occupations.[27] The results show that high HC distance firms grow significantly more in the top 5 occupations relative to low HC distance firms. Finally, in column 4, we show that a one standard deviation increase in HC distance is associated with a 7 p.p. decrease in HC distance three years after entry. Overall, Panel B shows that firms build by hiring new workers to reduce their distance to the sector of entry. The more distant firms are to the sector of entry, the more workers they hire to build in this sector.

In Table 5, Panel B, we report the post-entry labor dynamics of firms that buy. The evidence shows no statistically significant relationship between the acquirers' HC distance and their targets' employment growth. We then decompose the employment growth between occupations.

[25]Two main reasons motivate the choice of using the firm's HC distance and studying the employment dynamics at the subsidiary level. First, the decision to build or buy is likely to be made at the firm level. Second, the subsidiary that enters the new sector is arguably the one that needs to adjust its human capital. As a result, we should expect most of the workforce changes to happen at the subsidiary level. Third, focusing on the entering subsidiary avoids contamination from changes in the human capital of other subsidiaries that are not directly associated with the entry in the new sector. In addition, note that in the case of buy entries, the subsidiary through which the entry is performed is the target company.

[26]Note that the number of observations in Table 5 is lower than in other estimation tables because we focus on firms surviving at least three years after the entry to compute the three-year-employment growth.

[27]Note that $\Delta\text{Employment}_f = \Delta\text{Top5 occupations}_f + \Delta\text{Other occupations}_f$, such that the coefficients in columns 2 and 3 sum to the coefficients in column 1. This equality stems from the fact that $\text{Employment}_f = \text{Top5 occupations}_f + \text{Other occupations}_f$ and that the denominator for all three variables is $\text{Employment}_{f,t-1}$.

We find that the employment size of target companies acquired by high HC distance firms grow relatively more in the top 5 occupations (column 2). However, targets' employment growth in other occupations is negatively related to acquirers' HC distance. (column 3). In column 4, we find that the of targets' human capital becomes closer to that of incumbents in the sector of entry. This change in targets' employment composition is consistent with the idea that firms buy to acquire and retain the human capital they lack and let go of other types of workers.

5.2. Post-entry performance

In this section, we investigate whether having the right human capital to operate in the sector of entry affects post-entry performance, depending on whether firms build or buy. We measure performance using sales in the sector of entry and the probability of surviving in this sector after three years. The graphs above Table 6 plot the dynamics of post-entry sales of high HC distance firms compared to low HC distance firms. We again study build and buy entries separately because we expect post-entry firm performance differ depending on whether firms build or buy.[28]

[Insert Table 6 here]

Figure a shows that high HC distance firms realize approximately 40% lower sales in the sector of entry relative to low HC distance firms. Panel A of Table 6 shows that a one standard deviation increase in the firm's HC distance in the year before entry is associated with 22% lower sales at the end of the year of entry and a 2.8 p.p. lower probability of survival in the sector of entry after 3 years (column 2). This performance gap is persistent but decreases over time, consistent with the idea that distant firms adjust their workforce to reduce their HC distance to the sector of entry (Table 5). These results suggest that building on existing human capital is costly and takes time.

Figure b and Panel B of Table 6 compare the performance in the sector of entry of high and

[28]This graph plots the result of the regression

$$log(Sales)_{f,n,t} = \beta_t + \lambda_{o,n,t_0} + \sum_j \delta_j 1(\textit{High HC Distance})_{f,t} \cdot 1(t = j) + \epsilon_{f,t}$$

where f denotes the subsidiary of group g that enters the new sector n at time t_0, $log(Sales)_{f,n,t}$ is the logarithm of sales realized by subsidiary f at time t in sector n, β_t denotes year to entry fixed effects, λ_{o,n,t_0} denotes sector of origin $\times$ entry $\times$ year of entry fixed effects, and $1(\textit{High HC Distance})$ is a dummy equal to one if the firm is above the median of HC distance taken by sector of origin $\times$ entry $\times$ year triplet. The graph plots the δ_j coefficients. Adding subsidiary fixed effects yields consistent results in terms of dynamics of difference in sales between low- and high-distance firms but does not allow us to quantify the baseline level of the gap.

low HC distance firms that buy. We find that targets' post-entry performance is not significantly correlated with their acquirer's pre-entry human capital. This finding is consistent with the idea that firms that buy acquire an already operational workforce from their target, so that their post-entry performance is independent of their existing human capital.

5.3. Labor market tightness

Our main findings show that firms are more likely to buy when they do not already employ workers in the key occupations for the sector of entry. To confirm the importance of human capital in the choice to build or buy, we ask whether the composition of the existing workforce matters more when hiring new workers is more costly. This is the case if, for instance, key occupations are in short supply in the local labor market (LLM) in which the firm is operating.

To test this hypothesis, we construct a measure of LLM tightness based on occupations in high demand in the location of the firm's entry (see Section 2.1.4).[29] We estimate our main equation (2) on subsamples corresponding to terciles of LLM tightness.[30] We add LLM fixed effects to control for fixed geographical variation in the likelihood of building or buying. The first tercile corresponds to LLMs where firms can easily find workers in key occupations for the sector of entry. The third tercile corresponds to tight LLMs where key occupations for the sector of entry are in short supply. If human capital is a key driver of the decision to build or buy, its effect should be driven by entries in tight LLMs.

[Insert Table 7 here]

Table 7 reports the results. In columns 1 and 2, we regress the choice to build or buy on firms' HC distance in the first and second terciles of LLM tightness. The relationship between the probability of buying and HC distance is positive but not significantly different from zero, suggesting that human capital does not play a significant role in this context. By contrast, consistent with our prediction, column 3 shows a positive and significant relationship between the probability of buying and HC distance. This finding supports the idea that firms buy when

[29] Appendix figure A2 maps the geographic heterogeneity in the number of occupations in short supply across LLMs in 2013. Darker shades of blue indicate a higher number of occupations in high demand in the LLM. Interestingly, we see that LLM tightness is not systematically related to population density, as tight LLMs can be observed in both urban and rural areas.

[30] Note that the variable LLM tightness$_{z,n,t-1}$ requires the use of additional data obtained from the French unemployment agency available only from 2010 to 2014, which explains why these tests are conducted on a smaller sample.

building on their existing human capital is too costly.

5.4. The role of organizational skills

Why is it too costly for distant firms to hire enough workers to build in the new sector such that they instead prefer to buy an existing company? To answer this question, we draw from the organizational economics literature that suggests that building on existing human capital requires more than hiring new workers: Firms also need organizational skills to train and integrate these new workers into their existing processes (Prescott and Visscher, 1980). To evaluate this hypothesis, we test whether the relationship between HC distance and the decision to build or buy is driven by firms with low organizational skills.

We test this hypothesis using three different measures of organizational skills at the firm level. The first measure is based on the "knowledge-based hierarchy" literature (Garicano, 2000; Caliendo, Monte and Rossi-Hansberg, 2015) and proxies for a firm's organizational skills using the share of the wage bill that goes to workers in the top layers of the hierarchy. The second measure is based on SG&A expenses. While we do not directly observe this expense item in the French accounting system, SG&A expenses typically include the salaries paid to the staff of the accounting, information technology, marketing, and human resources departments. Therefore, our second measure of organizational skills is the share of the wage bill that goes to workers in occupations linked to these functions.[31] Third, we proxy for organizational skills by defining a dummy variable 1(HR Workers) equal to one if the firm employs at least one HR worker.[32]

Table 8 reports the results for the three different measures that proxy for organizational skills. Columns 1 to 3 report the results on the sample split by terciles of our first measure. The results show that the relationship between HC distance and the decision to build or buy is driven by firms in which top managers obtain a smaller fraction of the wage bill (column 1). Similarly, in columns 4 to 6, we find that the relationship between HC distance and the decision to build or buy is driven by firms with a small fraction of the wage bill dedicated to SG&A occupations. Finally, columns 7 and 8 show that the effect of human capital on the decision to build or buy is driven by firms that do not employ any HR workers. Our findings validate the view that firms buy in distant sectors when their lack of organizational skills implies a high cost

[31] We provide the list of occupations in Appendix B.

[32] We use a dummy instead of a continuous variable because 77% of firms in our sample do not have any HR workers in house.

of building, allowing them to acquire already operational human capital from their target.

[Insert Table 8 here]

6. Implications for Firms' Subsequent Entries

In this section, we explore the implications of the choice to build or buy for firms' subsequent entry decisions. We show in the previous sections that firms buy in sectors that are distant in terms of human capital and acquire their target's human capital to successfully operate in the sector of entry. Instead, firms build in close sectors and adjust their human capital for the sector of entry only incrementally, by hiring new workers. Therefore, buy entries lead to more dramatic changes in firms' human capital than build entries, so that firms that buy are in a position to subsequently enter new sectors that were initially too distant.

We test whether after an acquisition, firms enter more distant sectors. Our hypothesis is that the human capital of firms that buy matters less for the subsequent entry than that of firms that build. The dependent variable, $\mathbb{1}(\text{Diversify})_{g,n,t}$, is a dummy variable equal to one if firm g enters the sector of entry n in year t and zero if it enters another sector.[33] The main independent variable, HC distance$_{g,n,t_0}$, is the distance between the firm and a potential sector of entry, computed at t_0, which corresponds to the first year the firm appears in our sample. On average, firms appear in our sample 5.9 years before their first entry in a new sector.[34] Table 9 reports the results.

Column 1 shows a negative correlation between the choice to enter a given sector n and the initial HC distance to the sector, suggesting that firms enter sectors that are not too distant in terms of human capital. The coefficient implies that a one standard deviation increase in the firm's HC distance to the sector of entry is associated with a 0.22 p.p. decrease in the probability of diversifying in a given sector. Column 2 tests whether firms enter sectors that are more distant at time t when their first entry at t_1 was a buy entry. We interact HC distance$_{g,n,t_0}$ with two dummy variables: $1(t > t_1)$, for entries occurring after the first entry at t_1, and $1(\text{First entry by buying})_{g,t_1}$, equal to one if the first entry at t_1 is a buy entry. We expect the coefficient

[33] In this section, we are not attempting to explain whether the subsequent entry is a build or buy entry. Instead, we ask whether firms that buy are more likely to subsequently seize growth opportunities that would not have been available without buying first.

[34] Note that this specification multiplies the number of observations by the number of potential sectors that the firm can potentially enter. We restrict the sample to firms that diversify at least once in the sample period and compute the firm's HC distance to all sectors, every year.

of the triple interaction to be positive. In columns 3 and 4, we focus on the subsample of firms that buy at least once during the sample period. t_{Buy} is the first year a firm buys. We include firm fixed effects to test whether after a buy entry, firms enter sectors that are more distant to their initial human capital.

The two sets of tests yield a consistent picture. Both in the cross-section (comparing firms that buy to firms that build in column 2) and in the time series (within firms that buy in column 4), we find that HC distance becomes less predictive of firms' subsequent entry decisions after a buy entry than after a build entry. This finding supports the idea that buying allows firms to not only enter new sectors they could not enter by building but also to subsequently seize growth opportunities that would not have been available without buying first.

[Insert Table 9 here]

7. Alternative Determinants of the Decision to Build or Buy

7.1. Observable determinants

The decision to build or buy may depend on other considerations than human capital alone. This section considers other determinants of this decision and shows that HC distance plays a significant role in the decision to build or buy above and beyond these other determinants.

Physical capital. HC distance may capture differences in the physical capital investment needed to enter a given sector. Although the evidence on labor adjustments suggests that our results cannot fully be explained by the role of physical capital, in Appendix Table A7, we control for investment in physical capital. In column 1, we rank firms into ten deciles of physical capital investment in the year of entry. We run our baseline regressions with interacted sector of origin $\times$ entry $\times$ investment decile fixed effects, as well as year fixed effects.[35] This specification allows us to compare firms that operate in the same sector of origin and enter the same sector in the same year and also invest similar amounts when entering the new sector. In column 2, we include the (log) amount of capital expenditures at the time of entry. The point estimates are smaller but not significantly different from the baseline specification. In columns 3 to 5,

[35]Note that for firms that build, we measure investment using capital expenditures in the year of the entry. For buying firms, investment is measured using the target company's tangible assets.

we rerun our main specification in three subsamples split by sector-level capital intensity. We show that even when entering sectors with low capital intensity (for which physical capital is less likely to matter), firms are more likely to build in close sectors and buy in more distant sectors in terms of human capital.

Size and access to financing. In Appendix Table A8, we interact the firms' employment size, an indicator of whether a firm is already diversified, its total cash holdings per employee, and an indicator of whether the firm is publicly listed with HC distance. In columns 1 and 2, the coefficients on the interactions between employment size and firms' cash holdings and HC distance, respectively, are negative. These findings suggest that larger firms and cash-rich firms can build in more distant sectors. The evidence also shows that diversified firms build in more distant sectors. Finally, in column 4, we find that listed firms are 10% more likely to buy. The interaction coefficient shows that listed firms are not significantly more likely to build in more distant sectors.

Overall, the evidence shows that the decision to build or buy is more sensitive to the firm's human capital for smaller, financially constrained, or single-industry firms. Larger and diversified firms are more likely to have developed organizational skills, and hence to find it easier to build on their existing human capital (see section 5.4).

Geographical distance. Firms that enter a geographically distant sector may lack the specific resources to enter a sector by building (e.g., information or local customer base) and may prefer to buy instead. In Appendix Table A9, we investigate the role of geographical distance between firms' initial location and the location of the new activities. We construct two dummy variables, *County* and *Region*, equal to one if the firm diversifies in a new county (*département*) and in a new region (*région*), respectively. The results show that firms are 16.4% and 22.3% more likely to buy when the entry is in a new county or a new region, respectively (columns 1 and 2). However, the coefficients of the interaction terms between HC distance and the geographical distances are not significantly different from zero. In addition, the sensitivity of the decision to build or buy to HC distance remains the same when we control for entry in a new geographic area. Therefore, the evidence suggests that the role of geography in the firm's decision to build or buy does not confound the role of human capital.

Industry and product market distance. In Appendix Table A9, we consider the role of industry and product market distance between the firm's sector of origin and the sector of entry. First, we use the French sector classification to construct a simple measure of industry distance. The dummy variable $1(SameIndustry)_{o,n}$ is equal to one if the sector of origin and the sector of entry belong to the same one-(two-)digit industry code. To estimate the effect of industry distance on the decision to build or buy, we modify the set of fixed effects and use a combination of sector of origin $\times$ year and sector of entry $\times$ year fixed effects. In columns 3 and 4 of Appendix Table A9, we find that firms are significantly more likely to buy in a same one-(two-)digit industry. The interaction between HC distance and the industry dummy variable is negative and significantly different from zero, suggesting that build or buy decisions in related industries are primarily driven by non-human capital factors (e.g., market consolidation).

Next, we construct a product market distance adapted from Bloom, Schankerman and Van Reenen (2013).[36] The product market distance is equal to one when the sales portfolio of the diversifying firm is orthogonal to the sales portfolio of firms operating in the sector of entry and equal to zero when the two vectors are collinear. In columns 5 and 6, we show that the firms are less likely to buy in distant product markets. The HC distance coefficient is still significant and positive. The estimate is stronger in magnitude, reflecting the fact that the human capital and product market distance are positively correlated.

Firms may enter upstream sectors by acquiring suppliers or downstream sectors by acquiring customers to transfer goods along the supply chain, irrespective of human capital considerations. In columns 7 to 10, we test the effects of upstream and downstream vertical links on the decision to build or buy. We identify vertical links using the input-output (I/O) matrix made available by the French Bureau of Statistics.[37] We define vertical links between a firm and the sector of entry when more than 5% (10%) of the inputs of the firm's sector of origin originate from the sector of entry. We do not find that vertical integration considerations intervene in the decision to build or buy in a new sector. Moreover, the role of HC distance is not significantly affected

[36]The product market distance is defined as follows:

$$\text{Product Market Distance}_{g,n} = 1 - \frac{\sum_i v_{g,i} \cdot v_{n,i}}{\sqrt{\sum_i v_{g,i}^2}\sqrt{\sum_i v_{n,i}^2}}$$

where $v_{g,i}$ denotes the sales that firm g realizes in sector i, and $v_{n,i}$ denotes the share of sales that all firms with sector n as main sector of activity realize in sector i.

[37]The 2017 I/O matrix provides input-output links between 139 sectors. The I/O matrix is based on different industry classifications. We exclude from the regression sample observations whose sector of origin and that of entry belongs to the sample industry in I/O matrix classification.

when controlling for industry vertical links.

7.2. Build or buy and selection in the sector of entry

In the previous sections, we study the decision to build or buy independently from the choice of the sector of entry. However, if latent unobservable factors drive both the choice of the sector and the decision to build or buy, our estimates may erroneously lead us to conclude that firms' HC distance to the sector of entry explains the choice to build or buy.

The regression specified in equation (2) is subject to selection bias if the residuals ϵ are correlated with the residuals η of the selection equation (5). Assuming that η and ϵ, the residuals of the selection and the build or buy equations, are jointly normal with a correlation term γ, then the estimated coefficient $\hat{\delta}$ obtained from the estimation of equation (2) is given by the standard Heckman (1979) formula:

$$\hat{\delta} = \delta - \gamma \cdot \Gamma(X) \cdot \rho \tag{6}$$

where ρ is the HC distance coefficient in equation (5) and $\Gamma(X)$ is equal to minus the partial derivative of the inverse Mills ratio with respect to HC distance. $\Gamma(X)$ can be shown to be positive (Wooldridge, 2010).

Appendix Table A6 shows that firms are less likely to diversify in distant sectors in terms of human capital ($\rho < 0$)[38]. Therefore, self-selection bias may lead us to overestimate δ if the correlation γ between the error terms is positive. Intuitively, we overestimate δ if latent factors that lead firms to enter more distant sectors also lead them to buy rather than build.

In theory, self-selection bias could be addressed using a Heckman selection model (Li and Prabhala, 2007). However, this method requires an instrument that affects the type of entry only through the choice of sector. Instead, we identify three specific situations where the self-selection bias is likely to bias our estimate, i.e., situations where $\gamma > 0$. We then change our specification or focus on subsamples to limit the role of the selection mechanisms in our estimations and

[38] As in the main specification, we include sector of origin × entry × year fixed effects to account for unobservable complementarities between sectors that could drive the entry in a given sector. In column 3, we add firm × year fixed effects to examine the firm's choice to enter a sector n rather than another sector in a given year. This specification accounts for all unobservable firm-level determinants of the choice to diversify in a given sector. In addition, because the coefficient is estimated on firms that diversify at least once in a given year, this specification allows us to focus on the role of HC distance on the choice of sector of entry rather than on the choice of diversifying or not.

verify that our conclusions are still valid.

Shifting firms. First, we identify firms that operate in a declining sector and shift a substantial part of their activities to the sector of entry. Because diversification is usually associated with good pre-entry performance, firms operating in a declining sector are less likely to diversify ($\eta < 0$). However, conditional on diversifying, these firms would attempt to shift their activity toward sectors in which their workers would be productive, that is, sectors that are close in terms of HC distance. This mechanism would lead firms to build in close sectors irrespective of their workforce composition ($\epsilon < 0$ and hence $\gamma > 0$). In this situation, the effect of HC distance (δ) may be overestimated.

To address this issue, we remove firms that shift a substantial part of their activity from the sectors of origin to the sector of entry. Table 10 reports the results. In columns 1 to 3, we exclude firms that shift 100%, 50%, or 25% of their activity, respectively, from their sectors of origin to the sector of entry. The coefficients δ estimated without shifting firms are not statistically and economically different from those reported in Table 3. Thus, our results cannot be explained by firms building in close sectors to shift their activity away from declining sectors.

Serial acquirers. Second, we focus on serial acquirers. These firms represent a disproportionate share of buy entries (40.2%) and may buy companies for reasons that are disconnected from human capital considerations (Golubov, Yawson and Zhang, 2015; Malmendier and Tate, 2008). As a result, serial acquirers may choose to enter more distant sectors ($\eta > 0$) and do so by buying incumbents in these sectors ($\epsilon > 0$). We identify serial acquirers as firms that buy at least two firms during our sample period. We exclude these serial acquirers from our estimation sample and rerun our main build or buy specification (equation (2)). In column 4 of Table 10, we show that excluding serial acquirers does not significantly affect our estimate of the δ coefficient.

Life cycle. Third, the life cycle of firms may affect both their decision to diversify and their decision to build or buy in new sectors. When young, firms may be less inclined to diversify and may favor close sectors to search for activities that are good matches for their capabilities (Matsusaka, 2001). When old, firms are more likely to diversify by buying in the absence of internal growth options (Hoberg and Maksimovic, 2021). To control for the firm's life cycle, we

compare firms that build or buy from the same sector of origin, enter the same sector of entry, and operate in the same number of sectors of origin before entry (column 5) or in the same age decile (column 6). Our baseline estimates of δ are unchanged. Overall, we conclude that the role of selection is likely to be limited.

[Insert Table 10 here]

8. Conclusion

In this paper, we ask why some firms build by acquiring new resources to complement their existing ones ("build"), while others buy an existing company to acquire these resources ("buy"). We study the tradeoff between these two ways of acquiring new resources to better understand the frictions that firms face in expanding their scope.

Using French administrative datasets linked to commercial M&A datasets, we start by documenting that 98% of entries into new sectors (85% when weighting by entry sales) are made by building on existing resources, suggesting that firms only buy when they cannot build. Our main result is that firms build in close sectors and buy in more distant ones. We establish causality by exploiting variation in firms' distance to the sector of entry that are due to changes in the workforce composition of incumbent firms (shift-share design).

Next, we analyze the tradeoff faced by firms when deciding whether to build or buy in a new sector. On the one hand, firms can build on their existing human capital by hiring new workers to complement their workforce. We show that it is more costly for firms to build in more distant sectors, suggesting that buying becomes a more attractive option than building when HC distance is high enough. On the other hand, firms can buy to acquire the human capital they cannot build organically. We show that buying allows firms to perform well in the sector of entry regardless of their initial HC distance. Therefore, firms that are distant enough to the sector of entry prefer to buy to avoid the high cost of building by hiring many workers.

We explore the reasons why distant firms prefer to buy an existing company rather than hiring workers to build in the new sector. Consistent with the organizational economics literature, our results suggest that building by hiring new workers is costly because it requires organizational skills. The more distant the sector of entry, the more organizational skills are required. Therefore, firms buy in distant sectors to avoid the high cost of building, thereby

acquiring already operational human capital from their target.

Finally, we study how the decision to build or buy affect firms' subsequent entry decisions. Both in the cross-section (comparing firms that buy to firms that build) and in the time series (within firms that buy), we find that firms build in more distant sectors after a buy entry than after a build entry, so that the initial HC distance becomes less predictive of firms' subsequent entry decisions after a buy entry. This finding supports the idea that buying allows firms to not only enter new sectors they could not enter by building but also to subsequently seize growth opportunities that would not have been available without buying first.

References

Bartik, Timothy J. 1991. "Who benefits from state and local economic development policies?"

Bena, Jan, and Kai Li. 2014. "Corporate innovations and mergers and acquisitions." *Journal of Finance*, 69(5): 1923–1960.

Berg, Tobias, Daniel Streitz, and Michael Wedow. 2020. "Credit supply and (in)organic growth." *Working Paper*.

Bernard, Andrew B, Stephen J Redding, and Peter K Schott. 2010. "Multiple-product firms and product switching." *American Economic Review*, 100(1): 70–97.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying technology spillovers and product market rivalry." *Econometrica*, 81(4): 1347–1393.

Boehm, Johannes, Swati Dhingra, and John Morrow. 2019. "The comparative advantage of firms." *Working Paper*.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *Review of Economic Studies*, 89(1): 181–213.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The anatomy of french production hierarchies." *Journal of Political Economy*, 123(4).

Capron, Laurence, and Will Mitchell. 2012. *Build, borrow, or buy: Solving the growth dilemma.* Harvard Business Press.

Cestone, Giacinta, Chiara Fumagalli, Francis Kramarz, and Giovanni Pica. 2018. "Insurance between firms: The role of internal labor markets." *Working Paper*.

Chen, Deqiu, Huasheng Gao, and Yujing Ma. 2021. "Human capital-driven acquisition: Evidence from the Inevitable Disclosure Doctrine." *Management Science*, 67(8): 4643–4664.

Dessaint, Olivier, Andrey Golubov, and Paolo Volpin. 2017. "Employment protection and takeovers." *Journal of Financial Economics*, 125(2): 369–388.

Dessein, Wouter, and Andrea Prat. 2022. "Organizational capital, corporate leadership, and firm dynamics." *Journal of Political Economy*, 130(6): 000–000.

Eckel, Carsten, and J Peter Neary. 2010. "Multi-product firms and flexible manufacturing in the global economy." *Review of Economic Studies*, 77(1): 188–217.

Eisfeldt, Andrea L, and Dimitris Papanikolaou. 2013. "Organization capital and the cross-section of expected returns." *Journal of Finance*, 68(4): 1365–1406.

Fan, Joseph P. H., and Vidhan Goyal. 2006. "On the patterns and wealth effects of vertical mergers." *Journal of Business*, 79(2): 877–902.

Farjoun, Moshe. 1994. "Beyond industry boundaries: Human expertise, diversification and resource-related industry groups." *Organization Science*, 5(2): 185–199.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, 108(5): 874–904.

Giroud, Xavier, and Holger M Mueller. 2015. "Capital and labor reallocation within firms." *Journal of Finance*, 70(4): 1767–1804.

Giroud, Xavier, and Holger M Mueller. 2019. "Firms' internal networks and local economic shocks." *American Economic Review*, 109(10): 3617–49.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review*, 110(8): 2586–2624.

Golubov, Andrey, Alfred Yawson, and Huizhong Zhang. 2015. "Extraordinary acquirers." *Journal of Financial Economics*, 116(2): 314–330.

Grossman, Sanford J, and Oliver D Hart. 1986. "The costs and benefits of ownership: a theory of vertical and lateral integration." *Journal of Political Economy*, 94(4): 691–719.

Harrigan, James, Ariell Reshef, and Farid Toubal. 2016. "The march of the techies: Technology, trade, and job polarization in France, 1994-2007." Working Paper.

Hart, Oliver, and John Moore. 1990. "Property rights and the nature of the firm." *Journal of Political Economy*, 98(6): 1119–1158.

He, Alex Xi, and Daniel le Maire. 2022. "Mergers and managers: Manager-specific wage premiums and rent extraction in M&As." *Working Paper*.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica*, 47(1): 153–161.

Hoberg, Gerard, and Gordon Phillips. 2010. "Product market synergies and competition in mergers and acquisitions: A text-based analysis." *Review of Financial Studies*, 23(10): 3773–3811.

Hoberg, Gerard, and Gordon Phillips. 2016. "Text-based network industries and endogenous product differentiation." *Journal of Political Economy*, 124(5): 1423–1465.

Hoberg, Gerard, and Gordon Phillips. 2018. "Conglomerate industry choice and product language." *Management Science*, 64(8): 3735–3755.

Hoberg, Gerard, and Gordon Phillips. 2020. "Scope, Scale and Competition: The 21st Century Firm." *Working Paper*.

Hoberg, Gerard, and Vojislav Maksimovic. 2021. "Product life cycles in corporate finance." *Working Paper*.

Lagaras, Spyridon. 2020. "Corporate takeovers and labor restructuring." *Working Paper*.

Lagaras, Spyridon. 2021. "M&as, employee costs and labor reallocation." *Working Paper*.

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard. 2018. "Information technology and returns to scale." *Working Paper*.

Lazear, Edward P. 2009. "Firm-specific human capital: A skill-weights approach." *Journal of Political Economy*, 117(5): 914–940.

Lee, Gwendolyn K, and Marvin B Lieberman. 2010. "Acquisition vs. internal development as modes of market entry." *Strategic Management Journal*, 31(2): 140–158.

Lee, Kyeong Hun, David C. Mauer, and Emma Qianying Xu. 2018. "Human capital relatedness and mergers and acquisitions." *Journal of Financial Economics*, 129(1): 111–135.

Li, Kai, and Nagpurnanand R Prabhala. 2007. "Self-selection models in corporate finance." *Handbook of Empirical Corporate Finance*, 37–86.

Li, Kai, Buhui Qiu, and Rui Shen. 2018. "Organization capital and mergers and acquisitions." *Journal of Financial and Quantitative Analysis*, 53(4): 1871–1909.

Lucas, Robert. 1978. "On the size distribution of business firms." *Bell Journal of Economics*, 508–523.

Maksimovic, Vojislav, and Gordon M Phillips. 2013. "Conglomerate firms, internal capital markets, and the theory of the firm." *Annual Review of Financial Economics*, 5(1): 225–244.

Maksimovic, Vojislav, and Gordon Phillips. 2002. "Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence." *Journal of Finance*, 57(2): 721–767.

Malmendier, Ulrike, and Geoffrey Tate. 2008. "Who makes acquisitions? CEO overconfidence and the market's reaction." *Journal of Financial Economics*, 89(1): 20–43.

Ma, Song, Justin Murfin, and Ryan Pratt. 2021. "Young firms, old capital." *Journal of Financial Economics*.

Matsusaka, John G. 2001. "Corporate diversification, value maximization, and organizational capabilities." *Journal of Business*, 74(3): 409–431.

Matvos, Gregor, Amit Seru, and Rui C Silva. 2018. "Financial market frictions and diversification." *Journal of Financial Economics*, 127(1): 21–50.

Ma, Wenting, Paige Ouimet, and Elena Simintzi. 2018. "Mergers and acquisitions, technological change and inequality." *Working Paper*.

McCardle, Kevin F, and S Viswanathan. 1994. "The direct entry versus takeover decision and stock price performance around takeovers." *Journal of Business*, 1–43.

Neffke, Frank, and Martin Henning. 2013. "Skill relatedness and firm diversification." *Strategic Management Journal*, 34(3): 297–316.

Nocke, Volker, and Stephen Yeaple. 2007. "Cross-border mergers and acquisitions vs. greenfield foreign direct investment: The role of firm heterogeneity." *Journal of International Economics*, 72(2): 336–365.

Ouimet, Paige, and Rebecca Zarutskie. 2020. "Acquiring labor." *Quarterly Journal of Finance*, 10(03): 2050011.

Penrose, Edith. 1955. "Limits to the growth and size of firms." *American Economic Review*, 45(2): 531–543.

Phillips, Gordon M, and Alexei Zhdanov. 2013. "R&D and the Incentives from Merger and Acquisition Activity." *Review of Financial Studies*, 26(1): 34–78.

Prescott, Edward C, and Michael Visscher. 1980. "Organization capital." *Journal of Political Economy*, 88(3): 446–461.

Rhodes-Kropf, Matthew, and David T. Robinson. 2008. "The market for mergers and the boundaries of the firm." *Journal of Finance*, 63(3): 1169–1211.

Schoar, Antoinette. 2002. "Effects of corporate diversification on productivity." *Journal of Finance*, 57(6): 2379–2403.

Shin, Hyun-Han, and René M Stulz. 1998. "Are internal capital markets efficient?" *Quarterly Journal of Economics*, 113(2): 531–552.

Tate, Geoffrey A, and Liu Yang. 2016. "The human factor in acquisitions: Cross-industry labor mobility and corporate diversification." *Working Paper*.

Tate, Geoffrey, and Liu Yang. 2015. "The bright side of corporate diversification: Evidence from internal labor markets." *Review of Financial Studies*, 28(8): 2203–2249.

Teece, David J. 1982. "Towards an economic theory of the multiproduct firm." *Journal of Economic Behavior & Organization*, 3(1): 39–63.

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

Yip, George S. 1982. "Diversification entry: Internal development versus acquisition." *Strategic Management Journal*, 3(4): 331–345.

Figures and Tables

Table 1. Evolution of Build and Buy Entries

This table reports the number and relative frequency of buy entries and build entries. A firm "buys" when it enters a new sector through an acquisition. A firm "builds" when it enters a new sector organically through one of its subsidiaries. In panel A, sectors refer to an industry at the 3-digit level of the French SIC. *Sources:* SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset. *Sample:* Firms that enter a new sector during the 2005-2014 period.

Panel A. Build and buy entries by year

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Build (number)	12,548	10,576	11,460	15,548	12,952	18,019	11,973	11,544	11,320	11,245	127,185
Buy (number)	165	246	216	236	165	200	282	446	200	259	2,415
Buy (%, equally-weighted)	1.30	2.27	1.85	1.50	1.26	1.10	2.30	3.72	1.74	2.25	1.86
Buy (%, entry sales-weighted)	11.73	19.81	15.11	9.73	6.05	7.04	19.46	36.96	19.74	20.80	15.68

Panel B. Build and buy entries by industry classification level

Industry level:	5 digits	4 digits	3 digits	2 digits	1 digit
Build (number)	157,792	144,730	127,185	96,750	75,528
Buy (number)	2,650	2,533	2,415	1,811	1,088
Buy (%, equally-weighted)	1.65	1.72	1.86	1.84	1.42
Buy (%, entry sales-weighted)	10.79	12.78	15.68	21.02	20.75

Table 2. Summary Statistics

This table reports descriptive statistics (number of observations, mean, standard deviation, 5th, 25th, 50th, 75th, and 95th percentiles) for the variables used in the analysis. Variables are defined in Appendix Table A. *Sources:* SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, and matched employer-employee dataset. *Sample:* Firms that enter a new sector in the 2005-2014 period. Sectors refer to an industry at the 3-digit level of the French SIC.

	N	Mean	SD	P5	P25	P50	P75	P95
Dependent variables								
$1(\text{Buy})_{g,n,t}$	61,228	0.02	0.13	0.00	0.00	0.00	0.00	0.00
$1(\text{Diversification})_{g,n,t-1}$	86,661,511	0.00	0.04	0.00	0.00	0.00	0.00	0.00
$1(\text{Diversifying Subsidiary})_{f,n,t}$	77,838	0.24	0.43	0.00	0.00	0.00	0.00	1.00
Human capital distance								
$\text{HC distance}_{g,n,t-1}$	61,228	0.79	0.20	0.35	0.69	0.85	0.94	0.99
$\text{HC distance}_{g,n,t-1}$ (# workers)	61,228	0.79	0.21	0.34	0.71	0.86	0.94	0.99
$\text{HC distance}_{g,n,t-1}$ (hours)	61,224	0.79	0.21	0.34	0.70	0.86	0.94	0.99
$\text{HC distance}_{g,n,t-1}$ (simple average)	61,088	0.88	0.12	0.60	0.84	0.91	0.96	0.99
$\text{HC distance}_{g,n,t-1}$ (weight: workers)	61,088	0.88	0.12	0.58	0.86	0.93	0.97	0.99
$\text{HC distance}_{g,n,t-1}$ (weight: sales)	61,088	0.89	0.12	0.59	0.86	0.93	0.97	0.99
$\text{HC distance}_{o,n,t-1}$ (sector-level)	23,452	0.77	0.21	0.28	0.68	0.83	0.92	0.98
$\text{HC Bartik}_{g,n,\Delta t=03,11}$	16,347	-0.06	0.44	-0.73	-0.19	-0.02	0.06	0.57
Other firm-level independent variables								
$\log(\text{N. Employees})_{g,t-1}$	61,228	3.81	1.29	1.61	3.14	3.74	4.51	6.04
$\text{N. Occupations/N. Employees}_{g,t-1}$	61,228	0.37	0.26	0.07	0.18	0.31	0.49	1.00
$\text{Value added/N. Employees}_{g,t-1}$	61,228	0.06	0.05	0.02	0.04	0.05	0.07	0.14
$\text{Tangible Assets/N. Employees}_{g,t-1}$	61,228	0.07	0.17	0.00	0.01	0.03	0.06	0.20
$\text{Total wages/N. Employees}_{g,t-1}$	61,228	0.03	0.01	0.02	0.02	0.03	0.03	0.05
$\text{Cash/N. Employees}_{g,t-1}$	61,228	0.03	0.07	0.00	0.00	0.01	0.03	0.11
$\text{Sales shift (100\%)}_{g,t-1,t}$	61,228	0.01	0.12	0.00	0.00	0.00	0.00	0.00
$1(\text{Public})_{g,t-1}$	61,228	0.02	0.14	0.00	0.00	0.00	0.00	0.00
$\text{Serial acquirer}_{g}$	61,228	0.01	0.10	0.00	0.00	0.00	0.00	0.00
$1(\text{Diversified})_{g,t-1}$	61,228	0.55	0.50	0.00	0.00	1.00	1.00	1.00
$\text{Age}_{g,t-1}$	59,878	24.04	17.73	4.00	12.00	21.00	32.00	53.00
$\log(1+\text{Investment})_{g,t}$	48,908	4.46	2.17	0.51	3.09	4.47	5.85	8.02
$\text{Top layers}_{g,t-1}$	61,228	0.26	0.23	0.00	0.10	0.20	0.36	0.78
$\text{SG\&A}_{g,t-1}$	61,228	0.25	0.23	0.00	0.09	0.20	0.35	0.76
$1(\text{HR Workers})_{g,t-1}$	61,228	0.23	0.42	0.00	0.00	0.00	0.00	1.00
$1(\text{IT workers})_{g,03}$	16,347	0.28	0.45	0.00	0.00	0.00	1.00	1.00
Performance variables								
$\log(\text{Employment})_{f,t}$	34,415	3.53	1.02	1.79	3.09	3.56	4.06	5.19
$\Delta\ \text{Employment}_{f,t-1,t+3}$	31,280	-0.34	0.75	-1.00	-1.00	-0.17	0.04	0.42
$\Delta\ \text{Top 5 occupations}_{f,t-1,t+3}$	31,280	-0.07	0.34	-0.64	-0.10	0.00	0.00	0.19
$\Delta\ \text{Other occupations}_{f,t-1,t+3}$	31,280	-0.26	0.68	-1.00	-0.83	-0.14	0.06	0.44
$\Delta\ \text{HC distance}_{f,t,t+3}$	31,280	-0.34	0.63	-1.00	-1.00	-0.04	0.00	0.15
$\log(\text{Sales})_{g,n,t}$	34,802	6.20	1.50	3.69	5.15	6.26	7.31	8.59
$\text{Survival}_{g,n,t+3}$	31,280	0.18	0.39	0.00	0.00	0.00	0.00	1.00
Sector-level analysis								
$\log(1+\text{Entries})_{o,n,t}$	23,452	0.68	0.73	0.00	0.00	0.69	1.10	2.08
$\log(1+\text{Build})_{o,n,t}$	23,452	0.66	0.73	0.00	0.00	0.69	1.10	2.08
$\log(1+\text{Buy})_{o,n,t}$	23,452	0.03	0.15	0.00	0.00	0.00	0.00	0.00
$\text{Buy frequency}_{o,n,t}$	438	0.58	0.36	0.08	0.25	0.50	1.00	1.00
Other distances and sector-level independent variables								
$\text{LLM Tightness}_{z,n,t-1}$	16,094	0.18	0.18	0.01	0.05	0.12	0.26	0.57
$\text{Capital intensity}_{n,t}$	61,228	0.13	0.37	0.02	0.03	0.04	0.08	0.35

Continued next page

Summary statistics (continued)

	N	Mean	SD	P5	P25	P50	P75	P95
1(New county)$_{g,t}$	61,035	0.06	0.24	0.00	0.00	0.00	0.00	1.00
1(New region)$_{g,t}$	61,035	0.03	0.18	0.00	0.00	0.00	0.00	0.00
Product market Distance$_{o,n,t\text{-}1}$	61,228	0.07	0.26	0.00	0.00	0.00	0.00	0.99
1(Same 1-digit Industry)$_{o,n}$	61,228	0.33	0.47	0.00	0.00	0.00	1.00	1.00
1(Same 2-digit Industry)$_{o,n}$	61,228	0.16	0.37	0.00	0.00	0.00	0.00	1.00
1(Upstream link >5%)$_{o,n}$	61,228	0.23	0.42	0.00	0.00	0.00	0.00	1.00
1(Upstream link >10%)$_{o,n}$	61,228	0.14	0.35	0.00	0.00	0.00	0.00	1.00
1(Downstream link >5%)$_{o,n}$	61,228	0.16	0.37	0.00	0.00	0.00	0.00	1.00
1(Downstream link >10%)$_{o,n}$	61,228	0.11	0.31	0.00	0.00	0.00	0.00	1.00
N. New firms$_{n,t\text{-}1}$ (thousands)	61,225	2.92	3.81	0.10	0.65	1.44	3.09	12.43
New firm survival$_{n,t\text{-}1}$	61,228	0.74	0.12	0.46	0.67	0.76	0.83	0.88
Sales growth$_{n,t\text{-}1}$	61,228	0.06	0.14	-0.11	-0.01	0.04	0.10	0.29
Herfindahl index$_{n,t\text{-}1}$	61,228	0.02	0.04	0.00	0.00	0.01	0.02	0.06

Table 3. Human Capital and Build or Buy

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, and matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and analyzes the effect of human capital on the decision to build or buy. The dependent variable, $1(\text{Buy})_{g,n,t}$, is a dummy variable that takes value one if firm g buys in the sector of entry n in year t and zero if it builds. Entries in the new sector n are identified with sales reported in the new sector n. Sectors refer to an industry at the 3-digit level of the French SIC. The main independent variable, HC distance$_{g,n,t-1}$, measures the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The sector of origin o denotes the sector in which the firm realizes most of its sales at $t-1$. Sector of origin (second) denotes the second sector (in terms of sales) in which the firm realizes most of its sales at $t-1$. Sector of entry n denotes a new sector n in which firm g starts reporting sales. t is the year of entry. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Buy})_{g,n,t}$			
	(1)	(2)	(3)	(4)
HC Distance$_{g,n,t-1}$	0.019**	0.017**	0.015**	0.012**
	(0.009)	(0.007)	(0.007)	(0.005)
N. Occupations/N. Employees$_{g,t-1}$	0.016**	0.018***	0.014***	-0.020
	(0.006)	(0.006)	(0.004)	(0.015)
log(N. Employees)$_{g,t-1}$	0.008***	0.010***	0.005***	-0.036***
	(0.002)	(0.002)	(0.001)	(0.010)
Tangible Assets/N. Employees$_{g,t-1}$	-0.016**	-0.021**	-0.001	-0.026
	(0.008)	(0.008)	(0.006)	(0.021)
Cash/N. Employees$_{g,t-1}$	0.061**	0.053**	0.035	0.032
	(0.024)	(0.025)	(0.024)	(0.043)
Value added/N. Employees$_{g,t-1}$	-0.005	0.022	-0.009	-0.115**
	(0.035)	(0.031)	(0.025)	(0.046)
Total wages/N. Employees$_{g,t-1}$	0.923***	0.872***	0.615***	0.598*
	(0.354)	(0.304)	(0.215)	(0.341)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	No	No	No
Sector of Origin × Size × Entry FE	No	Yes	No	No
Sector of Origin (main) × Origin (second) × Entry FE	No	No	Yes	No
Sector of Origin × Entry FE	No	No	No	Yes
Firm FE	No	No	No	Yes
Year FE	No	Yes	Yes	Yes
Adjusted R^2	0.206	0.155	0.263	0.629
Observations	61,228	59,346	50,083	43,875

Table 4. Build or Buy and Changes in Incumbent Firms' Human Capital

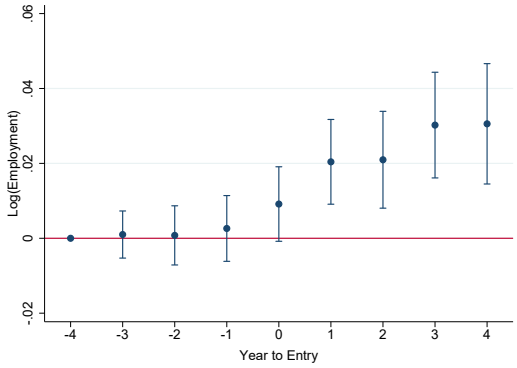
Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, and matched employer-employee dataset. *Sample:* Firms that are present in 2003 in the employer-employee dataset and that enter a new sector in the 2011-2014 period. The table reports OLS estimates and shows how changes in incumbent firms' human capital affect diversifying firms' build or buy decisions. The dependent variable, $1(\text{Buy})_{g,n,t}$, is a dummy variable that takes value one if firm g buys in the sector of entry n in year t, and zero if it builds. Entries in the new sector n are identified with sales reported in the new sector n. Sectors refer to an industry at the 3-digit level of the French SIC. The main independent variable, HC $\text{Bartik}_{g,n,03-11}$ corresponds to the opposite of the change of HC distance that is due to the long-term changes in the workforce composition of incumbent firms, holding the occupation shares of g at their 2003 value. A high value of HC Bartik means that sector n has become less distant from firm g in terms of its workforce composition between 2003 and 2011. HC $\text{distance}_{g,n,03-11}$ measures the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in 2003. Share IT $\text{workers}_{g,n,03-11}$ measures the share of wage bill of firm g in 2003 that goes to workers in IT-related occupations. In columns (4) and (5), we retain sectors for which the change between 2003 and 2011 in the share of IT-related occupations in the incumbent firms' wage bill composition is below the median (above in column 5). Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Buy})_{g,n,t}$					
				Change in IT share_n		
				Low	High	
	(1)	(2)	(3)	(4)	(5)	(6)
HC $\text{Bartik}_{g,n,03\text{-}11}$	-0.006**	-0.006*	-0.006**	-0.008	-0.004**	-0.006**
	(0.002)	(0.003)	(0.003)	(0.006)	(0.002)	(0.002)
HC $\text{Distance}_{g,n,03}$			-0.000			
			(0.002)			
$1(\text{IT workers})_{g,03}$						-0.001
						(0.003)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector of Origin × Entry FE	Yes	No	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	No	Yes	No	No	No	No
Adjusted R^2	0.194	0.194	0.194	0.169	0.224	0.194
Observations	16,347	16,347	16,347	8,289	8,058	16,347

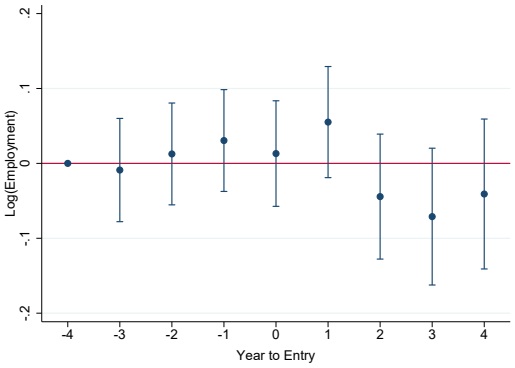
Table 5. Human Capital and Post-Entry Labor Adjustments

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector, realize at least 1% of their sales in the new sector in the 2005-2014 period. The graphs display the results of the OLS regression of the (log) of the number of employees on subsidiary fixed effects, sector of origin × entry × year of entry, and time to entry fixed effects. The high HC distance dummy variable is interacted with time to entry dummies (plotted on the graph; the first year is set as the baseline). We report build and buy entries separately. We identify "high HC distance firms" by computing a dummy equal to one if the HC distance of the firm to the sector of entry is above the median taken for each sector of origin × entry × year triplet. The table reports OLS estimates and shows the labor dynamic adjustments to HC distance after build entries (panel A) and after buy entries (panel B). The samples include only firms that survive three years after entry. The dependent variable in column (1) is the growth rate of the number of workers between $t-1$ and $t+3$ employed by the subsidiary f associated with the entry in sector n. In columns (2) and (3), we decompose $\Delta\text{Employment}_f$ into employment growth in the 5 most important occupations for the sector of entry in terms of worker share and the remaining occupations. We have that $\Delta\text{Employment}_f = \Delta\text{Top5 occupations}_f + \Delta\text{Other occupations}_f$, so that the coefficients in columns (2) and (3) sum to the coefficients in column (1). The dependent variable in column (4) is the subsidiary's HC distance variation between t and $t+3$. The main independent variable, $\text{HC distance}_{g,n,t-1}$, measures the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

(a) Build entries



(b) Buy entries



Panel A: Build entries

Dependent variable:	$\Delta\text{Employment}_f$	$\Delta\text{Top5 occupations}_f$	$\Delta\text{Other occupations}_f$	$\Delta\text{HC distance}_f$
Time interval:	$[t-1, t+3]$	$[t-1, t+3]$	$[t-1, t+3]$	$[t, t+3]$
	(1)	(2)	(3)	(4)
$\text{HC Distance}_{g,n,t-1}$	0.080***	0.551***	-0.470***	-0.351***
	(0.028)	(0.043)	(0.046)	(0.033)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.483	0.360	0.441	0.654
Observations	34,287	34,287	34,287	30,698

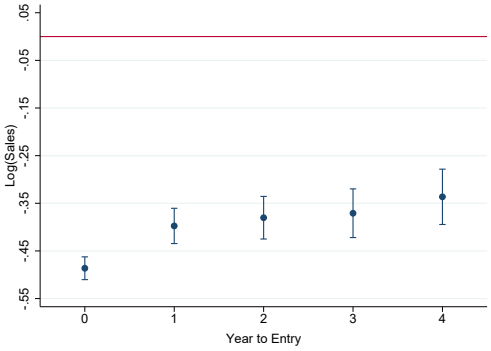
Panel B: Buy entries

Dependent variable:	$\Delta\text{Employment}_f$	$\Delta\text{Top5 occupations}_f$	$\Delta\text{Other occupations}_f$	$\Delta\text{HC distance}_f$
Time interval:	$[t-1, t+3]$	$[t-1, t+3]$	$[t-1, t+3]$	$[t, t+3]$
	(1)	(2)	(3)	(4)
$\text{HC Distance}_{g,n,t-1}$	-0.402	0.423**	-0.824***	-0.527**
	(0.259)	(0.164)	(0.157)	(0.220)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.478	0.540	0.359	0.658
Observations	287	287	287	240

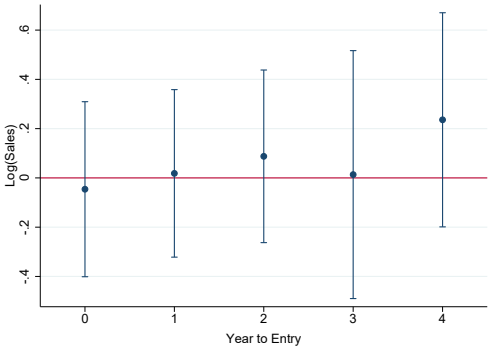
Table 6. Human Capital and Post-Entry Performance

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector, realize at least 1% of their sales in the new sector in the 2005-2014 period. The graphs display the results of the OLS regression of the (log) of sales on a sector of origin × entry × year of entry, a time to entry fixed effect and the high distance dummy coefficient interacted with time to entry dummies (plotted). We identify "high HC distance firms" by computing a dummy equal to one if the HC distance of the firm to the sector of entry is above the median taken for each sector of origin × entry × year triplet. The table reports OLS estimates and investigates the post-entry performance associated with HC distance after build entries (panel A) and after buy entries (panel B). The dependent variable in column (1) is firm g's logarithm of realized sales in sector of entry n and year of the entry t. The dependent variable in column (2) is a dummy variable that is equal to one if the firm still reports sales in sector of entry n three years after entry, zero otherwise. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

(a) Build entries



(b) Buy entries



Panel A: Build entries

Dependent variable:	$\text{Log(Sales)}_{g,n}$	$\text{Survival}_{g,n}$
Time interval:	t	$t+3$
	(1)	(2)
HC Distance$_{g,n,t-1}$	-1.119***	-0.141***
	(0.128)	(0.023)
Controls	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes
Adjusted R^2	0.471	0.164
Observations	34,415	34,415

Panel B: Buy entries

Dependent variable:	$\text{Log(Sales)}_{g,n}$	$\text{Survival}_{g,n}$
Time interval:	t	$t+3$
	(1)	(2)
HC Distance$_{g,n,t-1}$	-2.212	-0.285
	(2.201)	(0.201)
Controls	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes
Adjusted R^2	0.335	0.241
Observations	294	294

Table 7. Build or Buy and Local Labor Market Tightness

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset, unemployment agency dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2010-2014 period. The table reports OLS estimates and investigates how local labor market tightness shapes the decision to build or buy. The dependent variable $1(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in sector of entry n in year t and zero if it builds. Entries in new sector n are identified with sales reported in new sector n. Sectors refer to an industry at the 3-digit level of the French SIC. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Models are estimated on subsamples of firms corresponding to terciles of LLM tightness$_{z,n,t}$. LLM tightness$_{z,n,t}$ measures labor scarcity by local labor market (LLM) and corresponds to the number of occupations in high demand in a given local labor market. Occupations in high demand are identified by the French unemployment agency as occupations for which (i) the ratio of job offers over job applications is high and (ii) surveyed employers forecast that it will be difficult to fill posted offers. There are 348 different local labor markets. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects and local labor market fixed effects. Standard errors are triple-clustered at the sector of origin, sector of entry and local labor market level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Buy})_{g,n,t}$		
Subsamples of LLM Tightness:	Tercile 1	Tercile 2	Tercile 3
	(1)	(2)	(3)
HC Distance$_{g,n,t-1}$	0.001	-0.013	0.023**
	(0.020)	(0.022)	(0.011)
Controls	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	Yes
Local labor market FE	Yes	Yes	Yes
Adjusted R^2	0.404	0.517	0.291
Observations	3,825	3,673	3,815

Table 8. Build or Buy and Organizational Skills

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and investigates how organizational skills shape human capital and the decision to build or buy. The dependent variable $1(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in the sector of entry n in year t and zero if it builds. Entries in new sector n are identified with sales reported in new sector n. Sectors refer to an industry at the 3-digit level of the French SIC. The main independent variable, HC $\text{distance}_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Models are estimated on subsamples of firms sorted by three measures of organizational skills: (i) the fraction of a firm's wage bill going to workers in the top layers of the hierarchy (columns 1 to 3), (ii) that going to workers associated with Selling, General and Administrative (SG&A) expenses (columns 4 to 6), (iii) and a dummy equal to one if a firm employs any worker in human resources (i.e., the firm has HR services, columns 7 and 8). Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Buy})_{g,n,t}$							
	Top layers$_{g,t-1}$			SG&A$_{g,t-1}$			1(HR workers)$_{g,t-1}$	
	Low	Medium	High	Low	Medium	High	None	Some
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HC Distance$_{g,n,t-1}$	0.013***	0.004	0.018	0.012***	0.001	0.025	0.013*	-0.016
	(0.004)	(0.012)	(0.015)	(0.005)	(0.008)	(0.016)	(0.007)	(0.017)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.154	0.170	0.199	0.182	0.174	0.180	0.233	0.133
Observations	17,702	17,234	17,653	17,912	17,431	18,007	45,061	11,133

Table 9. Build or Buy, and Firms' Subsequent Entries

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, and matched employer-employee dataset. *Sample:* Firms that enter a new sector in the 2005-2014 period. The table reports OLS estimates and analyzes the effects of the first build or buy decision on the subsequent entry decisions. The dependent variable $1(\text{Diversify})_{g,n,t}$ is a dummy variable that takes value one if firm g enters sector n in year t, zero otherwise. Entries in new sector n are identified with sales reported in new sector n. Sectors refer to an industry at the 3-digit level of the French SIC. The main independent variable, HC $\text{distance}_{g,n,t0}$, measures the distance of firm g's wage bill composition across occupations to incumbent firms' wage bill composition in sector n in year t_0, which is the first year that firm g appears in our sample. In columns (1) and (2), we include all firms and compare firms' entries depending on whether they built or bought the first time they enter a new sector in the sample period (denoted by t_1). $1(\text{First entry by buying})_g$ is a dummy equal to one if firm g buys at time t_1. $1(t > t_1)$ is a dummy equal to one after time t_1. In columns (3) and (4), we only include firms that buy and compare entries within-firm before and after the first buy entry. The time of the first buy entry is denoted by t_{Buy}, and $1(t > t_{Buy})$ is a dummy equal to one after time t_{Buy}. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. All columns include sector of origin o-sector of entry n-year $t-1$ fixed effects and time to entry fixed effects. Columns (3) and (4) also include firm fixed effects. Time to entry is defined as $t - t_1$ in columns (1) and (2) and as $t - t_{Buy}$ in columns (3) and (4). Standard errors are triple-clustered at the sector of origin, sector of entry and firm level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Diversify})_{g,n,t}$			
Sample:	All firms		Buying firms	
	(1)	(2)	(3)	(4)
HC $\text{Distance}_{g,n,t_0}$	-0.011***	-0.011***	-0.019***	-0.049***
	(0.002)	(0.002)	(0.004)	(0.007)
$1(t > t_1)\times$ $1(\text{First entry by buying})_g$		-0.006**		
		(0.003)		
$1(\text{First entry by buying})_g \times$ HC $\text{Distance}_{g,n,t_0}$		0.001***		
		(0.001)		
$1(t > t_1)\times$ $1(\text{First entry by buying})_g \times$ HC $\text{Distance}_{g,n,t_0}$		0.006*		
		(0.003)		
$1(t > t_{Buy})\times$ HC $\text{Distance}_{g,n,t_0}$				0.054***
				(0.008)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes
Time to Entry FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.035	0.035	0.015	0.016
Observations	78,473,812	78,473,812	1,657,050	1,657,050

Table 10. Build or Buy, and Selection in the Sector of Entry

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, and matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and tests three scenarios where selection for build or buy could drive the main results. The dependent variable, $1(\text{Buy})_{g,n,t}$, takes value one if firm g buys in sector of entry n in year t and zero if it builds. The main independent variable HC distance$_{g,n,t-1}$, measures the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. In columns (1) to (3), we estimate the main specification on subsamples of firms that shift a substantial part of their activity to other sectors. Firms that shift decrease their activity in one of their existing sectors of activity around the time they enter a new sector n. We compute the firm's growth rate of sales between $t-1$ and t for each sector where the firm operates at $t-1$. The firm-level shift rate is the minimum of sectoral growth rates. Columns (1) to (3) exclude firms for which the minimum sales growth rate is negative and greater than 100%, 50% and 25% in absolute value. In column (4), Serial Acquirers are excluded from the sample. Serial acquirers are firms that enter more than one sector by acquisition during our sample time period. In column (5), we interact the number of sectors of origin o where the firm operates at $t-1$ with the sector of origin-entry fixed effects, plus year fixed effects. In column (6), we interact the firm's age split into deciles with sector of origin-entry fixed effects, plus year fixed effects, and we estimate our main model. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models in columns (1), (2), (3) and (4) include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Buy})_{g,n,t}$					
	Sales Shift$_{g,n,t,t-1}$			No Serial Acquirer	Life Cycle	
	100%	50%	25%		N. Sectors$_{g,t-1}$	Age$_{g,t-1}$ decile
	(1)	(2)	(3)	(4)	(5)	(6)
HC Distance$_{g,n,t-1}$	0.030**	0.039***	0.049***	0.015*	0.034***	0.027**
	(0.012)	(0.014)	(0.017)	(0.008)	(0.009)	(0.010)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	Yes	Yes	No	Yes
Sector of Origin × Entry × N. of sectors FE	No	No	No	No	Yes	No
Sector of Origin × Entry × Age decile FE	No	No	No	No	No	Yes
Year	No	No	No	No	Yes	Yes
Adjusted R^2	0.217	0.220	0.228	0.126	0.263	0.181
Observations	33,463	25,563	18,510	33,692	36,881	28,731

Appendix A Description of Variables

Variables	Description
Dependent variables	
$1(\text{Buy})_{g,n,t}$	Dummy variable that takes value one if firm g buys in the sector of entry n in year t and zero if it builds. *Sources:* SDC Platinum, Zephyr, Ownership links dataset, Sales breakdown dataset.
$1(\text{Diversify})_{g,n,t}$	Dummy variable that takes value one if firm g enters the sector n in year t, zero otherwise. *Sources:* SDC Platinum, Zephyr, Ownership links dataset, Sales breakdown dataset.
$1(\text{Diversifying Subsidiary})_{f,n,t}$	Dummy variable that takes value one if the entry in sector n at time t is made through subsidiary f and zero if it is made through another subsidiary within firm g. *Sources:* SDC Platinum, Zephyr, Ownership links dataset, Sales breakdown dataset.
Human Capital distance measures	
$\text{HC distance}_{g,n,t}$	Human capital distance is defined by equation (1): $$\text{HC distance}_{g,n,t} = 1 - \frac{\sum_i s_{g,i,t} \cdot s_{n,i,t}}{\sqrt{\sum_i s_{g,i,t}^2}\sqrt{\sum_i s_{n,i,t}^2}},$$ where $s_{g,i,t}$ is the share of firm g's total wage bill that goes to employees in occupation i, and $s_{n,i,t}$ is the share of incumbent firms in sector n's total wage bill that goes to employees in occupation i. *Source:* Matched employer-employee dataset.
$\text{HC distance}_{g,n,t}$ (number of workers)	Human capital distance is defined by (1), where $s_{g,i,t}$ is the share of employees in firm g employed in occupation i, and $s_{n,i,t}$ is the share of employees of incumbent firms in sector n employed in occupation i. *Source:* Matched employer-employee dataset.
$\text{HC distance}_{g,n,t}$ (hours)	Human capital distance is defined by (1), where $s_{g,i,t}$ is the share of the total number of hours worked in firm g by employees in occupation i, and $s_{n,i,t}$ is the share of the total number of hours worked by employees of incumbent firms in sector n in occupation i. *Source:* Matched employer-employee dataset.
$\text{HC distance}_{g,n,t_0}$	HC distance of firm g to sector n (in terms of wage shares) the first year firm g appears in the sample. *Source:* Matched employer-employee dataset.
$\text{HC distance}_{g,n,t}$ (simple average)	Simple average of the HC distance of firm g to firms operating in the sector n (in terms of wage shares). *Source:* Matched employer-employee dataset.
$\text{HC distance}_{g,n,t}$ (weighted average: employment)	Average of the HC distance of firm g to incumbent firms in sector n (in terms of wage shares) weighted by the number of workers of incumbent firms. *Source:* Matched employer-employee dataset.
$\text{HC distance}_{g,n,t}$ (weighted average: sales)	Average of the HC distance of firm g to incumbent firms in sector n (in terms of wage shares) weighted by sales of incumbent firms. *Source:* Matched employer-employee dataset.
$\text{HC distance}_{o,n,t}$	Sector-level human capital distance between the sector of origin o and the sector of entry n given by equation: $$\text{HC distance}_{o,n,t} = 1 - \frac{\sum_i s_{o,i,t} \cdot s_{n,i,t}}{\sqrt{\sum_i s_{o,i,t}^2}\sqrt{\sum_i s_{n,i,t}^2}}.$$ where $s_{o,i,t}$ is the share of the wage bill that goes to workers employed in occupation i in sector of origin o at time t, and $s_{n,i,t}$ is the share of the wage bill that goes to workers employed in occupation i in sector of entry n at time t. *Source:* Matched employer-employee dataset.
$\text{HC distance}_{f,n,t}$	Human capital distance at the subsidiary f level adapted from (1), where $s_{f,i,t}$ is the share of the wage bill in subsidiary f that goes to workers employed in occupation i, and $s_{n,i,t}$ is the share of the wage bill of incumbent firms in sector n that goes to workers employed in occupation i. *Source:* Matched employer-employee dataset, Ownership link dataset.

Continued next page

Description of Variables (continued)

Variables	Description
HC Bartik$_{g,n,\Delta t=03,11}$	The human capital Bartik-like instrument is defined as follows: $$\text{HC Bartik}_{g,n} = \sum_i \hat{s}_{g,i,03} \cdot \Delta\hat{s}_{n,i,03,11},$$ where $\Delta\hat{s}_{n,i,03,11} = \hat{s}_{n,i,11} - \hat{s}_{n,i,03}$ is the change in the (normalized) share of occupation i in the workforce of incumbent firms in sector n from 2003 to 2011. The normalized share $\hat{s}_i$ is $s_i/\sqrt{\sum_j s_j^2}$. HC Bartik corresponds to the opposite of the change in HC distance that is due to the long-term sector-level changes in workforce composition, holding the occupation shares of g at their 2003 value. A high value of HC Bartik means that sector n has become less distant from firm g. *Source:* Matched employer-employee dataset.
Other firm-level independent variables	
log(N. Employees)$_{g,t}$	Logarithm of the number of employees in firm g in year t. *Source:* Matched employer-employee dataset.
N. Occupations/N. Employees$_{g,t}$	Number of occupations in firm g in year t, scaled by the number of workers in the firm. It represents the firm's occupational diversity. *Sources:* Matched employer-employee dataset.
Value added/N.workers$_{g,t}$	Total value added generated by firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Matched employer-employee dataset.
Tangible Assets/N. Workers$_{g,t}$	Total value of tangible assets held by firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Matched employer-employee dataset.
Total wages/N. Workers$_{g,t}$	Total wages of firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Matched employer-employee dataset.
Cash/N. Workers$_{g,t}$	Total cash holdings of firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Matched employer-employee dataset.
Entry threshold$_{g,n,t}$	Ratio of firm's realized sales in the sector of entry n in year t divided by the firm's total sales realized in sectors of origin in $t-1$. We consider four entry thresholds: None, 1%, 5% (baseline specification), and 10%. *Sources:* Sales breakdown dataset, Ownership links dataset.
Sales shift$_{g,t\text{-}1,t}$	A firm that shifts is a firm that decreases its activity in sector of origin o when it enters sector of entry n. Firms that shift are identified by computing the growth rate of sales between $t-1$ and t for each sector where the firm operates at $t-1$. We take the firm-level minimum of sector growth rates. Sales growth rates are negative and greater than 100%, 50%, and 25% in absolute value. *Sources:* Sales breakdown dataset, Tax files.
Public firm$_g$	Dummy variable equal to one if at least one subsidiary f within firm g is a publicly listed company, zero otherwise. *Source:* Bureau van Dijk's Amadeus.
Serial acquirer$_g$	A serial acquirer is a firm that enters a new sector by acquisition at least twice during the time period. *Sources:* SDC Platinum, BvD Zephyr, Ownership links dataset, Sales breakdown dataset.
1(Diversified)$_{g,t}$	Dummy variable equal to one if firm g is diversified in year t, i.e., the firm operates in more than one sector. *Sources:* Ownership links dataset, Sales breakdown dataset.
Age$_{g,t}$	Age is defined as the maximum age of subsidiaries f that are under the ownership of the firm g. *Sources:* Ownership links dataset, Tax files.
Log(1+Investment)$_{g,t}$	Logarithm of one plus the investment realized by firm g when entering the new sector. Investment is defined as the capital expenditures of the entering subsidiary when the firm builds and the amount of the target company's tangible assets when the firm buys. *Sources:* Tax files.
Top layers$_{g,t\text{-}1}$	Share of the wage bill that goes to workers in the top layers of the French hierarchy of occupations in $t-1$. Top layers correspond to occupations codes that start with "2" or "3". *Sources:* Matched employer-employee dataset
SG&A$_{g,t\text{-}1}$	Share of the wage bill that goes to workers in the accounting, information technology, marketing, and human resources departments in $t-1$. The list of occupations is reported in Appendix B. *Sources:* Matched employer-employee dataset
1(HR Workers)$_{g,t\text{-}1}$	Dummy variable equal to one if the firm has at least one HR worker in its workforce in $t-1$. The list of HR occupations is reported in Appendix B. *Sources:* Matched employer-employee dataset

Continued next page

Description of Variables (continued)

Variables	Description
$1(\text{IT workers})_{g,03}$	Dummy variable equal to one if the firm has at least one IT worker in its workforce in 2003. The list of IT occupations is reported in Appendix B. *Sources:* Matched employer-employee dataset
Performance measures	
$\text{Log(Employment}_{f,t})$	Logarithm of the number of employees employed by the subsidiary f associated with the entry of firm g in sector of entry n. *Sources:* Matched employer-employee dataset.
Δ $\text{Employment}_{f,t-1,t+3}$	Growth between years $t-1$ and $t+3$ in the number of workers employed by the subsidiary f associated with the entry of firm g in sector of entry n. *Sources:* Matched employer-employee dataset.
Δ Top 5 $\text{occupations}_{f,t-1,t+3}$	Growth between years $t-1$ and $t+3$ in the number of workers employed in the five most important occupations for sector of entry n by the subsidiary f associated with the entry of firm g in sector of entry n. It is defined as: $$\Delta\text{Top 5 occupations}_{f,t-1,t+3} = \frac{\text{Top 5 occup.}_{f,t+3} - \text{Top 5 occup.}_{f,t-1}}{\text{Employment}_{f,t-1}}$$ where $\text{Top 5 occupation}_{f,t}$ is the number of workers employed by the subsidiary f associated with the entry of firm g in sector of entry n and employed in the 5 most important occupations for sector of entry n. *Sources:* Matched employer-employee dataset.
Δ Other $\text{occupations}_{f,t-1,t+3}$	Defined as $$\Delta\text{Other occupations}_{f,t-1,t+3} = \frac{\text{Other occup.}_{f,t+3} - \text{Other occup.}_{f,t-1}}{\text{Employment}_{f,t-1}}$$ where $\text{Other occupation}_{f,t+3}$ is the number of workers employed by the subsidiary f associated with the entry of firm g in sector of entry n and not employed in the 5 most important occupations for sector of entry n. *Sources:* Matched employer-employee dataset.
Δ $\text{HC distance}_{f,t,t+3}$	Growth in the HC distance of the subsidiary f associated with the entry of firm g in sector of entry n between years t and $t+3$. *Sources:* Matched employer-employee dataset.
$\text{Log(Sales}_{g,n,t})$	Logarithm of sales realized by firm g at time t in sector n. *Sources:* Tax files, Sales breakdown dataset.
$\text{Survival}_{g,n,t+3}$	Dummy variable that takes value one if firm g survives at least 3 years in the sector n that it enters at time t, zero otherwise. *Sources:* Tax files.
Subsequent entries variables	
$1(t > t_1)$	Dummy variable that takes value one if the next entry takes place after the first entry at t_1 and zero if before. *Sources:* Tax files, Sales breakdown dataset, SDC Platinum, Bureau Van Dijk's Zephyr.
$1(t > t_{Buy})$	Dummy variable that takes value one if the next entry takes place after a buy entry at t_{Buy} and zero if before. *Sources:* Tax files, Sales breakdown dataset, SDC Platinum, Bureau Van Dijk's Zephyr.
$1(\text{First entry by buying})_g$	Dummy variable that takes value one if the first entry at t_1 was a buy entry and zero if it was a build entry. *Sources:* Tax files, Sales breakdown dataset, SDC Platinum, Bureau Van Dijk's Zephyr.
Sector-level variables	
$\log(1+\text{Entries})_{o,n,t}$	Logarithm of the total number of entries in sector n at time t coming from firms operating in sector o. *Sources:* Sales breakdown dataset, SDC Platinum, Bureau Van Dijk's Zephyr.
$\log(1+\text{Build})_{o,n,t}$	Logarithm of the total number of build entries in sector n at time t coming from firms operating in sector o. *Sources:* Sales breakdown dataset, SDC Platinum, Bureau Van Dijk's Zephyr.
$\log(1+\text{Buy})_{o,n,t}$	Logarithm of the total number of buy entries in sector n at time t coming from firms operating in sector o. *Sources:* Sales breakdown dataset, SDC Platinum, Bureau Van Dijk's Zephyr.

Continued next page

Description of Variables (continued)

Variables	Description
Buy frequency$_{o,n,t}$	Ratio of the number of buy entries to build entries in sector n at time t coming from firms operating in sector o. The table is only defined for sectors in which we observe both build and buy entries. *Sources:* Sales breakdown dataset, SDC Platinum, Bureau Van Dijk's Zephyr.
Other distances and sector-level independent variables	
Capital intensity$_{n,t}$	Ratio of aggregate tangible assets divided by the total number of employees employed by firms in sector n at time t. *Sources:* Tax files.
1(New county)$_{g,z,t}$	Dummy variable that indicates whether firm g enters a new county (*département*) z in year t. France is divided into 101 counties. A firm enters a new county if it either (i) opens a new subsidiary f located in the county in year t, (ii) opens a new plant in the county in year t, or (iii) acquires an existing firm in year t that operates in a county where the firm was not present in $t-1$. *Sources:* Firm registry.
1(New region)$_{g,z,t}$	Dummy variable that indicates whether firm g enters a new region (*région*) z in year t. France is divided into 25 regions. A firm enters a new region if it either (i) opens a new subsidiary f located in the region in year t, (ii) opens a new plant in the region in year t, or (iii) acquires an existing firm in year t that operates in a region where the firm was not present in $t-1$. *Sources:* Firm registry.
LLM tightness$_{z,n,t}$	Local labor market tightness is defined as: $$\text{LLM Tightness}_{z,n} = \sum_i 1(\text{Occupation in high demand})_{i,z} \cdot s_{n,i},$$ where $1(\text{Occupation in high demand})_{i,z}$ is a dummy variable equal to one if occupation i is in high demand in z, and $s_{n,i}$ is the share of employees in sector n employed in occupation i. A high value of LLM Tightness$_{z,n}$ means that firms in area z are likely to face difficulties finding workers in an occupation that are key to operate in sector n. An occupation is flagged as "in high demand" in a local labor market if (i) the number of job ads posted on the French unemployment agency's website for that occupation by local employers exceeds the number of unemployed workers qualified for that occupation residing in the area or (ii) local employers declare in a yearly survey that they anticipate facing hiring difficulties for that occupation. *Sources:* Matched employer-employee dataset, French unemployment agency dataset.
Product market distance$_{g,n,t}$	The product market distance between firm g and sector n measures whether firm g's sectoral portfolio of sales is similar to that of incumbents in the sector of entry. The product market distance is adapted from Bloom, Schankerman and Van Reenen (2013). It is defined as: $$\text{Product Market Distance}_{g,n,t} = 1 - \frac{\sum_p v_{g,p,t} \cdot v_{n,p,t}}{\sqrt{\sum_p v_{g,p,t}^2}\sqrt{\sum_p v_{n,p,t}^2}}$$ where $v_{g,q,t}$ is the share of sales firm g realizes in sector q at time t, and $v_{n,q,t}$ is the share of sales firms operating mainly in sector of entry n realize in sector q at time t. The distance ranges from 0 to 1. *Sources:* Tax files, Sales breakdown dataset.
Industry classification (1-digit)$_{o,n}$	Dummy variable equal to one if the sector of entry n and the sector of origin o, defined at the French 3-digit SIC level, belong to the same French 1-digit SIC sector. *Sources:* Tax files, Sales breakdown dataset.
Sector classification (2-digits)$_{o,n}$	Dummy variable equal to one if the sector of entry n and the sector of origin o, defined at the French 3-digit SIC level, belong to the same French 2-digit SIC sector. *Sources:* Tax files, Sales breakdown dataset.
Upstream link$_{o,n} \geq 5\%$	Dummy variable equal to one if firms in sector of origin o (as measured in the 2017 French input-output matrix) source more than 5% of their inputs from sector of entry n. We define sectors with the I-O matrix classification and then map them to the French 3-digit SIC. The variable is not defined when the sector of origin and the sector of entry belong to the same industry of the I-O classification. The measure is based on Fan and Goyal (2006)'s vertical relatedness measure. *Source:* Tax files, 2017 French Input-Output matrix.

Continued next page

Description of Variables (continued)

Variables	Description
Upstream link$_{o,n} \geq 10\%$	Dummy variable equal to one of firms in sector of origin o (as measured in the 2017 French input-output matrix) source more than 10% of their inputs from sector of entry n. *Source:* Tax files, 2017 French Input-Output matrix.
Downstream link$_{o,n} \geq 5\%$	Dummy variable equal to one if firms in the sector of origin o (as measured in the 2017 French input-output matrix) realize more than 5% of their sales in sector of entry n. *Source:* Tax files, 2017 French Input-Output matrix.
Downstream link$_{o,n} \geq 10\%$	Dummy variable equal to one if firms in sector of origin o (as measured in the 2017 French input-output matrix) realize more than 10% of their sales in sector of entry n. *Source:* Tax files, 2017 French Input-Output matrix.
N. New firms$_{n,t}$	Number of firms newly created in the sector of entry in year t. *Sources:* firm registry
New firm survival$_{n,t}$	Probability that firms in sector of entry n survive at least 3 years after creation. *Sources:* Firm registry, Tax files.
Sales growth$_{n,t}$	Growth of total sales realized by incumbent firms in sector of entry n between years t and t. *Sources:* Tax files.
Herfindahl index$_{n,t}$	Herfindahl index defined as the sum of the squares of the market shares of the firms within a 3-digit French SIC sector. The index ranges between 0 and 1. An index close to 1 indicates a concentrated sector, and an index close to 0 indicates a competitive sector. *Sources:* Tax files.

Appendix for Online Publication:

Build or Buy?

Human Capital and Corporate Diversification

Paul Beaumont Camille Hebert Victor Lyonnet

6th June, 2022

This internet appendix presents additional results to accompany the paper "Build or Buy? Human Capital and Corporate Diversification". The contents are as follows:

Appendix A presents additional analysis to accompany our main empirical results.

Figure A1 plots the HC distance between French SIC-1 sectors.

Figure A2 maps local labor market tightness in France.

Table A1 lists the top occupations in three different sectors.

Table A2 reruns our main build or buy specification with alternative HC distances.

Table A3 tests the robustness of our main build or buy specification to different sales threshold of entry in the new sector.

Table A4 replicates our build or buy analysis at the sector level.

Table A5 tests whether the subsidiary responsible for the entry in the new sector has a low HC distance relative to other subsidiaries within the same firm.

Table A6 tests whether firms choose to enter close sectors in terms of HC distance.

Table A7 tests the effects of physical capital and capital intensity on the decision to build or buy.

Table A8 tests the effects of financial constraints and access to finance on the decision to build or buy.

Table A9 tests the effects of product market distances and geographical distances on the decision to build or buy.

Table A10 tests the effects of sector profitability on the decision to build or buy.

Appendix B reports the list of occupations used to measure firms' organizational skills and use of IT services.

Appendix A Additional Tables for Internet Appendix

Examples of top worker occupations for 3 sectors

Table A1. Examples of Top Occupations in Three Sectors

Source: Matched employer-employee dataset. *Sample:* Single-sector firms with more than 20 employees (2013).

Machinery & transport equipment		**Plastics, minerals & metals**		**Accommodation & food services**	
Top occupations (4-digits)	Share	Top occupations (4-digits)	Share	Top occupations (4-digits)	Share
673c: Unskilled workers in metal working	7.6%	**673a:** Unskilled workers in metal removal	5.7%	**561d:** Kitchen helpers, kitchen apprentices	25.2%
624c: Qualified assemblers of mechanical equipment working in large production lines	6.5%	**626b:** Skilled workers in metallurgy and glass production	4.5%	**561c:** Skilled waiters, restaurant clerks	19.0%
474c: Manufacturing and quality control technicians in metalworking	5.1%	**623a:** Skilled workers in metal forging	4.0%	**561b:** Unskilled waiters, restaurant clerks	15.1%
628g: Skilled workers in heavy industry	3.7%	**674a:** Unskilled workers: chemistry, pharmacy, plastics processing	4.0%	**636d:** Cooks	13.1%
623e: Welders	2.9%	**623g:** Skilled metalworkers on other machinery (except mould makers)	3.8%	**561f:** General-purpose hotel employees	7.1%

Figure A1. Heatmap: Human Capital Distance Across Sectors

The heatmap shows the pairwise HC distance between sectors in 2013. The sector-level HC measures the distance between the vectors giving the share of workers by occupation in the aggregate sectoral workforce (see equation (B)). The measure ranges from 0 to 1. It is equal to zero when workers in two sectors are present in similar proportions across worker occupations and close to one workers are mostly present across worker occupations that do not overlap in the two sectors. Darker shades of grey indicate a greater human capital distance between sectors. *Source:* Matched employer-employee dataset. *Sample:* Single-sector firms with at least 20 employees.

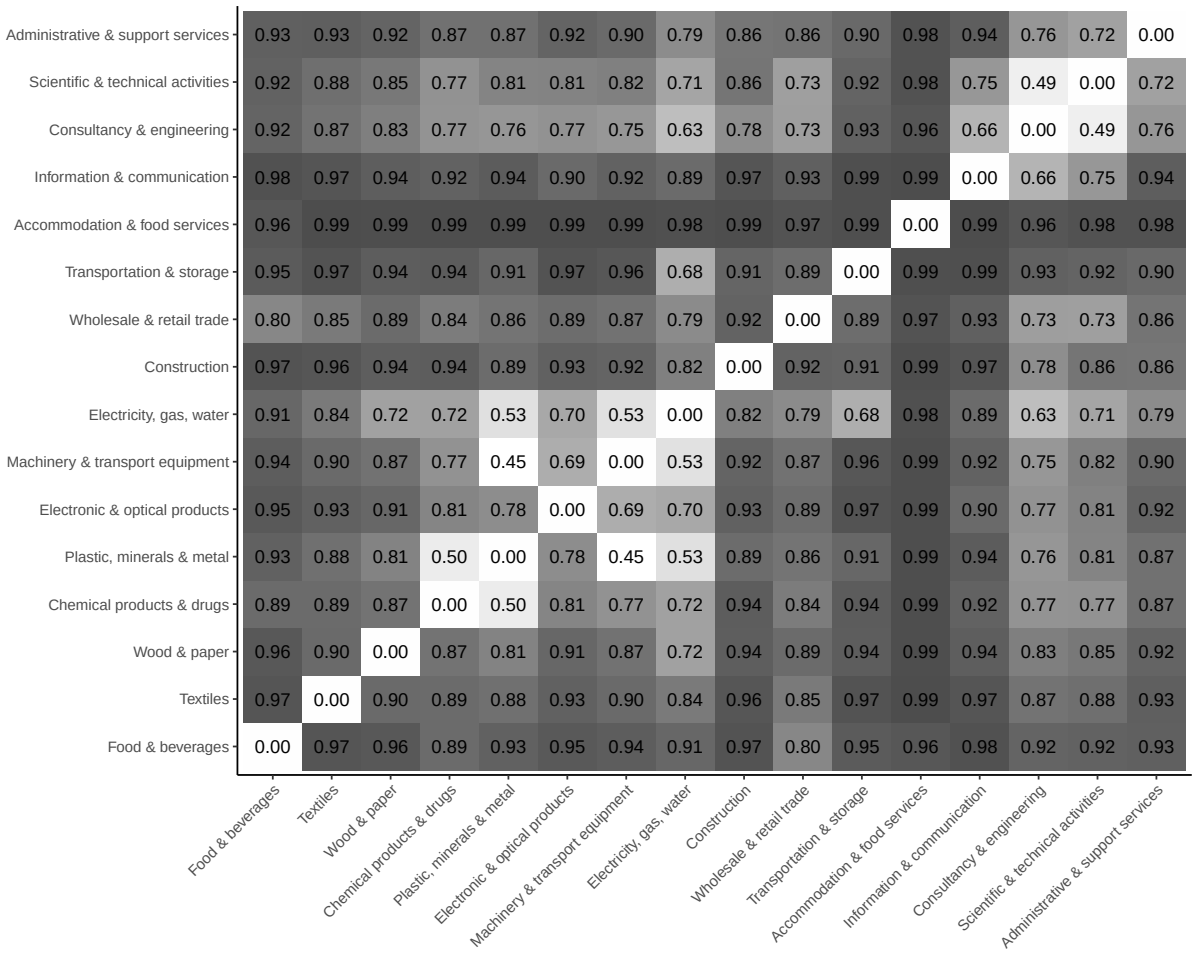


Figure A2. Local Labor Market Tightness in France

Source: French unemployment agency (Pole Emploi). This figure plots the distribution of local labor market tightness in 2013. LLM tightness$_{z,n,t}$ measures labor scarcity by local labor market (LLM) and corresponds to the number of occupations in high demand in a given local labor market. Occupations in high demand are identified by the French unemployment agency as occupations for which (i) the ratio of job offers over job applications is high and (ii) surveyed employers forecast that it will be difficult to fill posted offers. There are 348 different local labor markets. Darker shades of blue indicate a higher degree of LLM tightness$_{z,n,t}$.

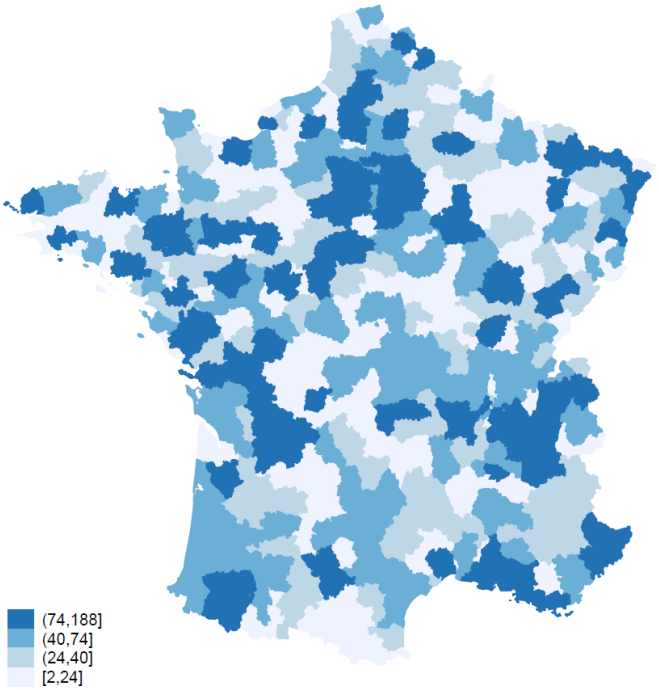


Table A2. Robustness Checks: Alternative Measures of HC Distance

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and tests the robustness of the different HC distances on the decision to build or buy. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in the sector of entry n in year t and zero if it builds. In column (1), the dependent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. In columns (2) and (3), we compute the distance based on the workforce composition in terms of the share of the total wage bill and in the number of hours worked per occupation. In column (4), we calculate the simple average of HC distance to firms operating in the sector of entry. In columns (5) and (6), we calculate the average of HC distance to firms operating in the sector of entry, weighted by incumbent firms' employment size and by incumbent firms' realized sales in the sector of entry, respectively. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
HC Distance$_{g,n,t-1}$ (wages - baseline)	0.019**					
	(0.009)					
HC Distance$_{g,n,t-1}$ (# workers)		0.021**				
		(0.008)				
HC Distance$_{g,n,t-1}$ (hours)			0.020**			
			(0.009)			
HC Distance$_{g,n,t-1}$ (simple average)				0.024*		
				(0.013)		
HC Distance$_{g,n,t-1}$ (weighted average: employment)					0.024*	
					(0.013)	
HC Distance$_{g,n,t-1}$ (weighted average: sales)						0.022*
						(0.013)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.206	0.206	0.206	0.205	0.205	0.205
Observations	61,228	61,228	61,224	61,064	61,064	61,064

Table A3. Robustness Checks: Alternative Entry Thresholds

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector in the 2005-2014 period. The table reports OLS estimates and tests the robustness of the relationship between HC distance and the decision to build or buy relative to the definition of sector entry thresholds. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in sector of entry n in year t and zero if it builds. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. The entry threshold is defined as the ratio of sales realized in the sector of entry n divided by the firm g's total sales in year $t-1$. In columns (1) to (4), we consider four thresholds: 0%, 1%, 5%, and 10%. In column (5), we estimate a WLS model weighted the entry sales. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
	Entry threshold$_{g,n,t,t-1}$				
	None	1% (baseline)	5%	10%	WLS
	(1)	(2)	(3)	(4)	(5)
HC Distance$_{g,n,t-1}$	0.011	0.019**	0.032***	0.039**	0.088***
	(0.007)	(0.009)	(0.012)	(0.015)	(0.030)
Controls	Yes	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.200	0.206	0.217	0.219	0.639
Observations	100,362	61,239	34,056	22,534	61,239

Table A4. Robustness Checks: Sector-level Analysis

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, and matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and replicates our analysis at the sector level. The dependent variables in columns (1) to (3) are the log of one plus the number of total entries, build entries, and buy entries, respectively, coming from firms in sector of origin o and entering sector n during year t. In column (4), the dependent variable is the ratio of buy entries divided by the total number of entries in sector n and year t. The model in column (4) includes only sectors that include both build and buy entries. The main independent variable, HC distance$_{o,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Control variables include the product market distance, vertical links upstream and downstream, and dummy that indicates whether sectors o and n belong to the same French 2-digit SIC. We include sector of origin-year and sector of entry-year fixed effects. Standard errors are clustered at the sector of origin level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	log(1+Entries)$_{o,n,t}$	log(1+Build)$_{o,n,t}$	log(1+Buy)$_{o,n,t}$	Buy frequency$_{o,n,t}$
	(1)	(2)	(3)	(4)
HC Distance$_{o,n,t-1}$	-1.057***	-1.047***	-0.069***	0.462***
	(0.104)	(0.105)	(0.017)	(0.157)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Year FE	Yes	Yes	Yes	Yes
Sector of Entry × Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.296	0.290	0.116	0.330
Observations	23,452	23,452	23,452	438

Table A5. Build or Buy: Within-group Analysis

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector in the 2005-2014 period. The table reports OLS estimates and tests whether the subsidiary that enters the new sector has a low HC distance to the sector of entry relative to other subsidiaries within the same firm. The dependent variable is a dummy variable that takes value one if the entry in sector n at time t is made through subsidiary f and zero if it is made through another subsidiary within firm g. Entries are identified with sales reported at the 3-digit level of the French SIC. The main independent variable is the subsidiary-level measure of HC distance$_{f,n,t-1}$, namely, the distance of the subsidiary f's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Columns (2) and (3) include the following control variables: the number of workers, the subsidiary's cash holdings, amount of tangible assets and value added, with the last three variables being scaled by the number of workers in the subsidiary. All models include firm × sector of entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}$(Diversifying Subsidiary)$_{f,n,t}$		
	(1)	(2)	(3)
HC Distance$_{f,n,t-1}$	-0.893***	-0.516***	-0.516***
	(0.016)	(0.014)	(0.015)
Controls	No	Yes	Yes
Firm × Entry × Year FE	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	No	No	Yes
Adjusted R^2	-0.039	0.177	-0.006
Observations	78,801	77,838	77,838

Table A6. Build or Buy, and the Choice of Sector

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, and matched employer-employee dataset. *Sample:* Firms that enter a new sector in the 2005-2014 period. The table reports OLS estimates and analyzes the effect of human capital on the choice of the sector of entry. The dependent variable $\mathbb{1}(\text{Diversify})_{g,n,t}$ is a dummy variable that takes value one if firm g enters sector n in year t and zero otherwise. Entries in new sector n are identified with sales reported in new sector n. Sectors refer to an industry at the 3-digit level of the French SIC. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models in columns (1) to (5) include various combinations of fixed effects. The sector of origin o denotes the sector in which the firm realizes most of its sales in year $t-1$. Sector of entry n denotes a new sector in which firm g starts reporting sales. t is the year of entry. Standard errors are triple-clustered at the sector of origin, sector of entry and firm level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Diversify})_{g,n,t}$		
	(1)	(2)	(3)
HC Distance$_{g,n,t-1}$	-0.014***	-0.007***	-0.009***
	(0.000)	(0.001)	(0.001)
Controls	Yes	Yes	No
Sector of Origin-Entry-Year FE	No	Yes	Yes
Firm-Year FE	No	No	Yes
Adjusted R^2	0.002	0.035	0.037
Observations	77,272,601	77,272,601	77,272,601

Table A7. The Role of Physical Capital

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and tests the robustness of the role of human capital in the decision to build or buy relative to the role of physical capital. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in sector of entry n in year t and zero if it builds. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. Column (1) includes sector of origin × sector of entry × year × investment deciles fixed effects. Investment is defined as the capital expenditures made by the firm when it enters by building, or as the amount of tangible capital of the target company when the firm buys. In column (2), we add the log of 1 plus capital expenditures to our baseline build or buy specification. In columns (3) to (5), we estimate our baseline specifications on subsamples of sectors of entry ranked by level of capital intensity. Column (2) to (5) include sector of origin-entry-year fixed effects. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
			Capital intensity$_{n,t-1}$		
			Low	Medium	High
	(1)	(2)	(3)	(4)	(5)
HC Distance$_{g,n,t-1}$	0.011**	0.013*	0.019*	0.009	0.036**
	(0.005)	(0.007)	(0.010)	(0.011)	(0.015)
log(1+Investment)$_{g,t}$		0.006***			
		(0.002)			
Controls	Yes	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	No	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Investment decile FE	Yes	No	No	No	No
Year FE	Yes	No	No	No	No
Adjusted R^2	0.370	0.198	0.175	0.194	0.253
Observations	47,715	47,488	20,701	20,150	20,377

Table A8. The Role of Firm Size and Access to Finance

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and tests the effect of human capital interacted with measures of firm size and access to internal finance on the decision to build or buy. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in sector of entry n in year t and zero if it builds. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. In columns (1) and (2), it is interacted with terciles of the firm's employment size and terciles of the firm's total cash holdings per employee, respectively. In column (3), it is interacted with a dummy variable that takes value one if the firm is already diversified before the entry and zero otherwise. In column (4), it is interacted with a dummy variable that takes value one if the firm is listed and zero otherwise. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
Measure of financial constraint:	N. Workers$_{g,t-1}$	Cash$_{g,t-1}$	Diversified$_{g,t-1}$	Public$_{g,t-1}$
	(1)	(2)	(3)	(4)
HC Distance$_{g,n,t-1}$	0.028**	0.032***	0.033***	0.024***
	(0.012)	(0.008)	(0.011)	(0.009)
2nd tercile of N. Workers$_{g,t-1}$	0.009			
	(0.006)			
3rd tercile of N. Workers$_{g,t-1}$	0.024***			
	(0.008)			
2nd t. N. Workers$_{g,t-1}\times$ HC Distance$_{g,n,t-1}$	-0.015*			
	(0.008)			
3rd t. N. Workers$_{g,t-1}\times$ HC Distance$_{g,n,t-1}$	-0.019*			
	(0.011)			
2nd tercile of Cash/N. Workers$_{g,t-1}$		0.010*		
		(0.005)		
3rd tercile of Cash/N. Workers$_{g,t-1}$		0.014*		
		(0.008)		
2nd t. Cash/N. Workers$_{g,t-1}\times$ HC Distance$_{g,n,t-1}$		-0.011*		
		(0.006)		
3rd t. Cash/N. Workers$_{g,t-1}\times$ HC Distance$_{g,n,t-1}$		-0.014*		
		(0.008)		
1(Diversified)$_{g,t-1}$			0.017**	
			(0.007)	
1(Diversified)$_{g,t-1}\times$ HC Distance$_{g,n,t-1}$			-0.021***	
			(0.007)	
1(Public)$_{g,t-1}$				0.102***
				(0.039)
1(Public)$_{g,t-1}\times$ HC Distance$_{g,n,t-1}$				-0.049
				(0.054)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.138	0.139	0.140	0.145
Observations	61,228	61,228	61,228	61,228

Table A9. The Role of Product Market and Geographic Distances

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and tests the effect of human capital interacted with measures of product market distances and geographical distances on the decision to build or buy. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in sector of entry n in year t and zero if it builds. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. In columns (1) and (2), HC distance$_{g,n,t-1}$ is interacted with a dummy variable that indicates whether the firm enters a new geographical area at time t, a department or a region, respectively. In columns (3) and (4), HC distance$_{g,n,t-1}$ is interacted with a dummy variable that indicates if the sector of entry is in the same 1-digit or 2-digit French SIC as the sector of origin. In columns (5) and (6), HC distance$_{g,n,t-1}$ is interacted with a product market distance based on Bloom et al. (2013). In columns (7) and (8), HC distance$_{g,n,t-1}$ is interacted with a dummy variable that indicates if the sector of entry has an upstream vertical link with the sector of origin. In columns (9) and (10), HC distance$_{g,n,t-1}$ is interacted with a dummy variable that indicates if the sector of entry has a downstream vertical link with the sector of origin. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin-entry-year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$									
Alternative distance:	Geographical distance		Sector Classification		Product distance		Upstream link		Downstream link	
	Department	Region	1-digit	2-digits			> 5%	> 10%	> 5%	> 10%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
HC Distance$_{g,n,t-1}$	0.021***	0.019***	0.048***	0.039***	0.023***	0.396***	0.043***	0.042***	0.035***	0.041***
	(0.007)	(0.007)	(0.016)	(0.013)	(0.006)	(0.130)	(0.015)	(0.014)	(0.012)	(0.014)
1(New department)$_{g,t}$	0.164***	0.223***								
	(0.038)	(0.051)								
New area$_{g,t}$× HC Distance$_{g,n,t-1}$	-0.064	-0.067								
	(0.042)	(0.061)								
1(Same Industry)$_{o,n}$			0.031**	0.025**						
			(0.013)	(0.011)						
Same Ind.$_{o,n}$× HC Distance$_{g,n,t-1}$			-0.040***	-0.032**						
			(0.015)	(0.013)						
Product market distance$_{g,n,t-1}$					-0.225***	0.029				
					(0.042)	(0.077)				
PM Distance$_{g,n,t-1}$× HC Distance$_{g,n,t-1}$						-0.382***				
						(0.130)				
1(Upstream link)$_{o,n}$							0.016	0.024		
							(0.019)	(0.017)		
Upstream link$_{o,n}$× HC Distance$_{g,n,t-1}$							-0.024	-0.034*		
							(0.022)	(0.019)		
1(Downstream link)$_{o,n}$									-0.025	0.008
									(0.017)	(0.021)
Downstream link$_{o,n}$× HC Distance$_{g,n,t-1}$									0.035	0.005
									(0.024)	(0.034)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	No	No	Yes	Yes	No	No	No	No
Sector of Origin × Year FE	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Entry × Year FE	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Adjusted R^2	0.250	0.259	0.119	0.119	0.120	0.122	0.128	0.128	0.129	0.128
Observations	61,013	61,013	71,051	71,051	67,871	67,871	60,468	60,468	60,468	60,468

Table A10. The Role of Sectoral Profitability

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, sales breakdown dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 1% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and tests the effect of human capital interacted with measures of sectoral profitability on the decision to build or buy. The dependent variable $1(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if firm g buys in sector of entry n in year t and zero if it builds. The main independent variable, HC distance$_{g,n,t-1}$, is the distance of firm g's wage bill composition across occupations to the incumbent firms' wage bill composition in sector n in year $t-1$. In column (1), HC distance$_{g,n,t-1}$ is interacted with tercile dummies for the number of firms created in the sector of entry in $t-1$. In column (2), HC distance$_{g,n,t-1}$ is interacted with tercile dummies for the 3-year survival probability of firms created in the sector of entry in $t-1$. In column (3), HC distance$_{g,n,t-1}$ is interacted with tercile dummies for the growth of sectoral sales between $t-2$ and $t-1$. In column (4), HC distance$_{g,n,t-1}$ is interacted with tercile dummies for the Herfindahl index of the sector of entry. Control variables include the firm's total number of workers in logarithms, the value added, number of occupations, total cash holdings, tangible assets, and total wages, all five scaled by the number of workers employed by the firm. The models include sector of origin, sector of entry, and year fixed effects. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Buy})_{g,n,t}$			
Sector of entry characteristics:	Firm creation	New firm survival	Sales growth	Competition
	(1)	(2)	(3)	(4)
HC Distance$_{g,n,t-1}$	0.048***	0.044***	0.025*	0.024**
	(0.007)	(0.006)	(0.012)	(0.010)
2nd tercile of New firms$_{n,t-1}$	0.008			
	(0.008)			
3rd tercile of New firms$_{n,t-1}$	0.016*			
	(0.008)			
2nd t. New firms$_{n,t-1}$ × HC Distance$_{g,n,t-1}$	-0.013			
	(0.009)			
3rd t. New firms$_{n,t-1}$ × HC Distance$_{g,n,t-1}$	-0.027***			
	(0.008)			
2nd tercile of 3-yr Survival probability$_{n,t-1}$		0.012*		
		(0.006)		
3rd tercile of 3-yr Survival probability$_{n,t-1}$		0.011		
		(0.007)		
2nd t. Survival prob.$_{n,t-1}$ × HC Distance$_{g,n,t-1}$		-0.015**		
		(0.007)		
3rd t. Survival prob.$_{n,t-1}$ × HC Distance$_{g,n,t-1}$		-0.014*		
		(0.008)		
2nd tercile of Sector growth$_{n,t-1}$			0.004	
			(0.006)	
3rd tercile of Sector growth$_{n,t-1}$			-0.004	
			(0.004)	
2nd t. Sector growth$_{n,t-1}$ × HC Distance$_{g,n,t-1}$			-0.011	
			(0.006)	
3rd t. Sector growth$_{n,t-1}$ × HC Distance$_{g,n,t-1}$			-0.008	
			(0.005)	
2nd tercile of Herfindahl index$_{n,t-1}$				-0.019**
				(0.009)
3rd tercile of Herfindahl index$_{n,t-1}$				-0.011
				(0.008)
2nd t. Herfindahl$_{n,t-1}$ × HC Distance$_{g,n,t-1}$				0.023**
				(0.010)
3rd t. Herfindahl$_{n,t-1}$ × HC Distance$_{g,n,t-1}$				0.009
				(0.009)
Controls	Yes	Yes	Yes	Yes
Sector of Origin FE	Yes	Yes	Yes	Yes
Sector of Entry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.090	0.090	0.064	0.090
Observations	74,738	74,738	74,738	74,738

Appendix B Organizational skills and IT services

We define SG&A workers as workers in the human resources, administrative resources, IT services, or marketing department.

Human resources. "Executives specializing in human resources and recruitment" (372c) "Training specialists" (372d) "Masters and Administrative Technicians of Legal Services or Personnel" (461e) "Qualified personnel from personnel and legal departments" (543e).

Administrative resources. "Heads of large companies with 500 or more employees" (231a) "Managers of medium-sized companies, from 50 to 499 employees" (232a) "Building and public works company managers, with 10 to 49 employees" (233a) "Business leaders in industry or transport, with 10 to 49 employees" (233b) "Heads of commercial enterprises, from 10 to 49 employees" (233c) "Heads of service companies, from 10 to 49 employees" (233d) "Administrative, financial and commercial staff of large companies" (371a) "Executives for the organization or control of administrative and financial services" (372b) "Lawyers" (372e) "Documentation and archiving executives (excluding public service)" (372f) "Executives of financial or accounting services of large companies" (373a) "Managers of other administrative departments of large companies" (373b) "Executives of financial or accounting services of small and medium-sized enterprises" (373c) "Executive secretaries, executive assistants (non-executives)" (461b) "Higher-level secretaries (non-executives, excluding executive secretaries)" (461c) "Masters and technicians of financial or accounting services" (461d) "Masters and administrative technicians of other administrative services" (461f) "Interpreters, translators" (464b) "Commercial and administrative managers of passenger transport and tourism (non-executives)" (466a) "Commercial and administrative managers of goods transport (non-executives)" (466b) "Technician-level experts, miscellaneous technicians" (479b).

IT services. "Technical directors of large companies" (380a) "Engineers and study executives, research and development in computer science" (388a) "Engineers and executives in computer administration, maintenance, support and user services" (388b) "IT project managers, IT managers" (388c) "Engineers and technical sales executives in IT and telecommunications" (388d) "Sales and technical sales technicians, IT representatives" (463a) "IT study and development

technicians" (478a) "Production technicians, IT operations" (478b) "Computer installation, maintenance, support and user services technicians" (478c) "Telecommunications and computer network technicians" (478d) "Technician-level experts, miscellaneous technicians" (479b) "Computer operating employees and operators" (544a) "Administrative employees of the technical services of the bank" (545a).

Marketing. "Advertising executives" (375a) "Public Relations and Communications Executives" (375b) "Publicity, Public Relations Assistants" (464a) "Graphic arts, fashion and decoration designers and technical assistants" (465a) "Technical assistants for the production of live and audiovisual shows" (465b) "Photographers" (465c).