Build or Buy?
Human Capital and Corporate Diversification

Paul Beaumont Camille Hebert Victor Lyonnet*

30th May, 2021

Abstract

This paper shows how human capital shapes firms' response to growth opportunities. Using French matched employer-employee data, we show that firms diversify in sectors in which incumbents employ similar occupations as theirs (low "human capital distance"). Conditionally on entering a new sector, firms enter closer sectors by using their internal resources ("build") and more distant sectors by acquiring an incumbent ("buy"). We provide evidence that firms buy when hiring workers externally is too costly. Firms absorb occupation-specific human capital when buying, which allows them to grow more quickly in the sector of entry, and to subsequently diversify in more distant sectors.

Keywords: Diversification, M&As, (In)organic growth, Human capital, Labor market frictions.

JEL codes: L25, J24, J30, G34.

*Paul Beaumont is with the 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, Philippe Aghion, Alice Bonaime, Edouard Challe, Gilles Chemla, Edith Ginglinger, Andrey Golubov, John Graham, Denis Gromb, Johan Hombert, Julien Martin, Adrien Matray, Isabelle Méjean, Paige Ouimet, Gordon Phillips, Manju Puri, Amiyatosh Purnanandam, David Robinson, Christoph Schneider, Matthew Serfling, Merih Sevilir, René Stulz, Geoffrey Tate, Vish Viswanathan, Ramona Westermann, Liu Yang and Sheng-Jun Xu for valuable comments and helpful discussions. We thank also seminar participants at CREST, INSEE, Université Paris-Dauphine, Tilburg University, HEC Paris, Duke University, the Ohio State University, the University of Toronto, McGill University, UQAM, the 5th ECGC workshop on Governance and Control (Best Student Paper Award), the UT Austin Corporate Finance Forum, the 28th Mitsui Labor and Finance Symposium, NFA, SDU workshop on the Firm's Life Cycle, RCEA Growth, Innovation and Entrepreneurship workshop, UNC Private Equity spring symposium 2020, the CEPR-Bank of Italy workshop on Ownership, Management & Firm Performance, AFA 2020, SFS Cavalcade 2020 and MFA 2021. 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 and omissions.

1. Introduction

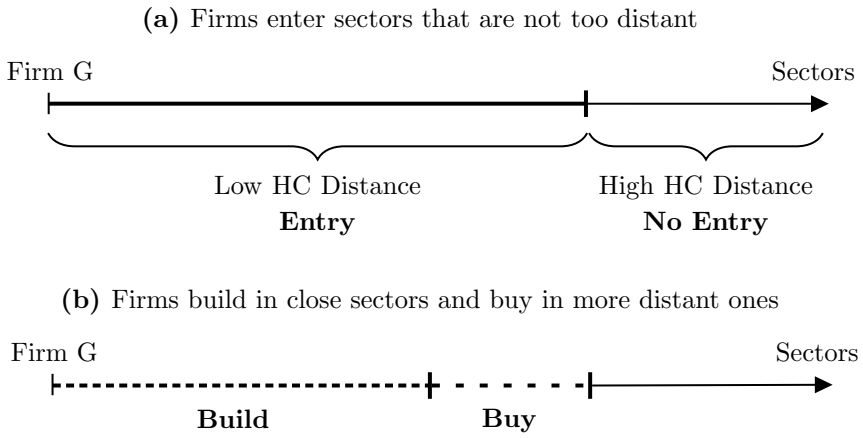
The value of the firm comes from its assets in place and its growth opportunities (Myers, 1977). A key challenge to a complete theory of the firm is the inextricable link between the two. We need a better knowledge of how assets in place define the set of growth opportunities to understand why a firm is worth more than the sum of its parts. This paper proposes to study the link between assets in place and growth opportunities by focusing on what academics and practitioners alike have argued is the most valuable asset of modern corporations: human capital.[1]

In this paper, we ask whether human capital affects firms' value by defining their set of growth options and how they take advantage of these opportunities. We propose a new measure of human capital (HC) distance between a firm and a potential sector of entry, where the firm can invest. We show that the firm's initial human capital endowment explains in which sectors and how firms seize new growth opportunities. Figure 1 illustrates our main results. First, the evidence shows that firms are more likely to enter sectors that require similar skills to benefit from economies of scope (sectors characterized by a low human capital distance Panzar and Willig, 1981; Teece, 1982). Second, conditional on diversifying, firms enter closer sectors in terms of human capital by building on their existing human capital and enter more distant sectors by buying an incumbent company that already employs the complementary skills to operate in the new sector that the acquiring firm cannot replicate itself. In addition, the decision to build or buy in the new sector leads firms to adjust their human capital differently upon entry, which ultimately has implications for firms' performance and the path of subsequent corporate diversifications.

We propose a new measure of human capital distance between a firm and a potential sector of entry and obtain two main findings. Figure 1 illustrates these findings. First, we show that firms are more likely to enter sectors that employ similar workers (low human capital distance). Second, conditional on diversifying, firms enter closer sectors in terms of human capital by building on their existing human capital and enter more distant sectors by buying an incumbent company that already employs the necessary skills to operate in the new sector.

[1]The importance of human capital is reflected in the amount of resources firms devote to hiring, training and retaining employees. See (Abowd and Kramarz, 2003) and (Blatter, Muehlemann, and Schenker, 2012) for hiring costs, (Prendergast, 1993; Jacobson, LaLonde, and Sullivan, 1993; Rajan and Zingales, 2001) for training costs, and (Oyer, 2004; Lustig, Syverson, and Van Nieuwerburgh, 2011) for retention costs.

Figure 1. HC Distance and Entry in New Sectors



In addition, we show that the decision to build or buy leads firms to adjust their human capital differently following the entry, which ultimately has implications for firms' performance and the path of subsequent corporate diversifications.

Our measure of HC Distance is based on the overlap in worker occupations between a company and the sector of entry. This definition considers firms' human capital as a combination of occupation-specific skills (Lazear, 2009), which has the advantage of encompassing several components of the human capital of the entire workforce–including education, experience, and industry skills–associated with the realization of specific tasks (Becker, 1962; Gibbons and Waldman, 2004; Autor and Dorn, 2009). We build this measure for all French firms that enter a new sector during the 2003–2014 period, using the matched employer-employee dataset, which includes a detailed occupation code at the individual-employee level and the exact employment composition of all firms. HC Distance is close to one when the firm does not employ workers in occupations that are heavily represented in the sector of entry, and is close to zero when the firm employs such key workers.[2]

We use exhaustive French administrative data, including the detailed breakdown of firms' sales by sector, linked to M&A deals retrieved from SDC Platinum and Bureau van Dijk databases, to build a novel dataset that contains the universe of corporate diversifications, identified as entries into new sectors.[3]. Instead of taking the firms' portfolio of activities as

[2]Our HC Distance measure is related to Tate and Yang (2016b)'s measure of Human Capital Transferability and Lee, Mauer, and Xu (2018)'s sector-level measure of Human Capital Relatedness. However, one notable difference is that our measure of HC Distance is a firm-level measure of distance to the entry sector and is based on workers' detailed occupations .

[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 sector classification

given, we consider the firm's scope of activity as the result of successive entries into new sectors. Importantly , we define entries at the business-group level. A firm enters a new sector in a given year when its sales in this sector are positive for the first time that year across all of its subsidiaries. This can be due to the acquisition of an existing firm (buy) or the organic entry of one or several of its existing subsidiaries (build).[4] An entry is a "buy" if the firm has acquired the subsidiary that begins selling in the new sector. By contrast, the entry is a "build" if the entry is made through an existing subsidiary. The resulting dataset consists of 35,000 build or buy entries in France from 2003 to 2014. We find that 98% of entries (85% when weighting by entry sales) are made by building, suggesting that the existing M&As fulfills a specific role in corporate diversification.

Our first main result is to show that firms are more likely to enter sectors that share similarities in terms of human capital. We compute each firm's HC Distance to every sector of entry and we compare firms in the same sector of origin, entering the same new sector in the same year. This approach accounts for unobservable positive or negative demand shocks that may affect the sector of entry or the sector of origin and other unobservable complementarities between sectors.[5] We find that a one-standard-deviation increase in the firm's HC Distance (0.21) is associated with a 0.53 percentage point decrease in the probability to diversify in a given sector. This negative correlation suggests that firms are more likely to enter sectors that require similar skills to benefit from economies of scope, and possibly to overcome asymmetries of information associated with performing in a new sector.

Our second main result is that conditional on diversifying, firms build in closer sectors in terms of human capital and buy in more distant ones. In our baseline specification, we compare firms that operate in the same sector of origin and enter the same sector of entry in the same year.[6] A one-standard-deviation increase in HC Distance is associated with a 0.71 p.p. increase in the likelihood of buying. This relationship is sizable, equal to 25% of the unconditional probability of buying. The relationship is unchanged when we control for firm's size, age, profitability, capital intensity, cash holdings, access to external financing (i.e.,

[4]To match SDC Platinum and Zephyr to the French administrative data, we develop a web crawler that takes as input the acquiring and target firms' names and addresses to link SDC Platinum and Zephyr deals to the French administrative data. We also discard LBO deals from the sample.

[5]We also control for firm' size, profitability, capital intensity, and cash holdings. We show that our main findings are robust to various combinations of fixed effects, including size buckets and the second main sector interacted with the sectors of origin-entry fixed effects, and firm-year fixed effects.

[6]This specification also includes firm' size, profitability, capital intensity, and cash holdings. In addition, we show that our main findings are robust to various combinations of fixed effects, including size buckets and the second main sector interacted with the sectors of origin-entry fixed effects.

public versus private firms), or the sector of entry's expected profitability (e.g., number of new firms, sectoral growth, or competition). Moreover, our inferences are robust to a number of alternative measures of HC Distance and imposing different sales thresholds to identify entries in new sectors. The evidence suggests that build in closer sectors where they can redeploy their human capital, and they buy to get access the necessary complementary skills to operate in the sector of entry. However, consistent with the lack of such complementarities, we find that firms are unlikely to enter unrelated sectors that are too distant from the sector of origin. [7]

We exploit the panel dimension of our dataset and its quasi-exhaustive nature to move closer to a causal interpretation. First, the correlation between HC Distance and the decision to build or buy remains positive and significantly different from zero when including firm fixed effects. Even within firms, build in closer sectors and buy in more distant ones. This finding invalidates alternative explanations based on unobservable differences between firms that build or buy. Second, we show that HC Distance predicts the choice of the sector of entry and the decision to build or buy even when calculated at the first year the firm appears in the sample period, that is, 4.9 years before the entry on average. This specification exploits the fact that firm's human capital is sticky over time, so that it is unlikely that the relationship between HC Distance and the decision to buy stems from firms that adjust their workforce in anticipation of entering the new sector (i.e., reverse causality). Third, we show that this relationship holds at the sector level. These findings mitigate concerns that firms' human capital reflects omitted variables related to the decision to build or buy.

We then provide evidence that the acquisition of skilled workers for the sector of entry drives the decision to build or buy. Firms should be more likely to buy when they do not employ the specific skills to operate in the sector of entry, and when these skills are in short supply on the external job market. We use occupation-level data on the local labor market (LLM) tightness that captures the extent to which specific occupations for the sector of entry are in short supply in the geographic area where the firm enters.[8] We find that firms are more likely buy when LLMs for the key specific occupations are tight. The relationship between the decision to build or buy and HC Distance is driven by the highest tercile of LLM Tightness .[9]

[7] Another reason why firms do not enter too distant sectors is that they may face a larger adverse selection toward the acquisition of a human capital they do not know.

[8] We obtained the data about occupations in short supply from the French unemployment agency to test this hypothesis. 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 in a position in this occupation. The data report occupations "in short supply" across 350 LLMs.

[9] By focusing on the geographic area where the firm enters, we take the choice of the area of entry as given. In

Next, we study the human capital adjustments associated with the decision to build or buy upon entry. First, we focus on firms that build and compare the employment dynamics of firms in the same sector of origin that build in the same sector of entry in the same year. We show that firms that build in more distant sectors experience a significantly higher employment growth after entering the new sector, relative to firms that build in closer sectors to their sector of origin for which they already employ the necessary skills. Employment growth is driven by workers hired in the top 5 occupations for the sector of entry, which ultimately reduces the firm's HC Distance after entry in the new sector. Second, we study how firms readjust the acquired target's human capital, which is the entity operating in the new sector. We do not find significant differences in employment growth between targets of low and high HC Distance acquiring firms, suggesting the acquiring firm's pre-entry human capital is independent from human capital adjustments happing at the target level. We then decompose the effect by occupations. We find that a positive employment growth in the top 5 occupations for the sector of entry, of targets acquired by high-HC Distance firms, and a negative employment growth in other occupations. Overall, the evidence is consistent with the idea that buy entries are motivated by the acquisition of targets' occupation-specific human capital, whereas build entries rely on the firm's endowed human capital and adjust the workforce to the sector of entry at the margin.

We show that human capital adjustments upon entry associated with the decision to build or buy are associated with better performance of firms in the new sector. Regarding, firms that build in the sector of entry, our results show that a one standard deviation increase in HC Distance is associated with 14% lower sales in the year following entry in the new sector. The performance gap between high and low HC Distance firms remains significantly different from zero up to three years after entry, but decreases in magnitude over time, consistent with the evidence that human capital adjustments are more likely to take place at high HC Distance firms. In addition, we find that high-HC Distance firms are significantly more likely to exit the sector of entry over the course of the next three years.

Consistent with our previous results on human capital adjustments, we show that the target's post-acquisition performance, measured by sales and the probability of survival, is unrelated to the acquiring firm's pre-entry HC Distance. This findings suggests that entering

an additional test, we show that within firms, entries are more often made through subsidiaries with a low HC Distance to the sector of entry, suggesting that the choice of the subsidiaries through which the entry is made, in a given area is itself driven by labor adjustment costs considering within firms.

by buying allows firms to acquire a human capital "off the shelf" and be quickly profitable in the sector of entry.[10] By contrast, our results on firms that build suggests that developing new skills organically takes time and that firms that build perform well in the sector of entry when their pre-entry human capital is adapted to the sector of entry.

Finally, we show that the human capital adjustments associated with the decision to build or buy ultimately modify firms' ability to seize new opportunities in the future. In line with this intuition, we find that firms that buy's initial human capital becomes relatively less predictive of the subsequent sectors of entry relative to firms that enter the new sector by building. Thus, the acquisition of the target allows the acquiring firms to enter sectors that were initially too distant from the sector of origin, but also dramatically reshape the firm's human capital by integrating workers with different skills, which redefines the firm's set of growth opportunities in the future. By contrast, human capital adjustments that have occurred at firms that build marginally modifies its set of opportunities, which leads however, without modifying its growth path.

The existence of post-entry human capital adjustments and the fact that the relationship between human capital and the decision to build is driven by occupations in short supply support the hypothesis that human capital plays a specific role in firms' entry decisions. Still, the correlation between HC Distance and the decision to buy could be overestimated in the presence of latent factors driving firms to buy in distant sectors or to build in close sectors. In the last part of the paper, we explore the role of other potential determinants of the decision to build or buy. First, we show that firms do not choose to build or buy only to minimize physical capital adjustments costs. Our results hold when we control for capital expenditures or when we focus on sectors of entry in which physical capital is less likely to play a role. Second, we find that the relationship between HC Distance and the decision to build or buy is robust to controlling for vertical integration (input-output links, downstream and upstream relationships), product market distance, or geographical proximity (i.e, whether the firms enter a new area while entering the new sector).[11] Third, we exclude firms that could have specific motives for

[10] In line with this idea, Cisco's Chief Strategy Officer writes that acquisitions "provide a capability acceleration potential or earlier sectoral entry compared to partnering or developing in-house." (Hilton Romanski, Cisco's blog, 2017). Another illustration is Disney's acquisition of BAMTech in 2017 and 21st Centrury Fox in 2019, in preparation for the launch of the Disney+ streaming platform.

[11] Our product market distance is based on Bloom, Schankerman, and Van Reenen (2013) and uses the correlation between the product market portfolio of firms mainly operating in the sector of entry and the product market portfolio of firms mainly operating in the sector of origin. The distance is equal to one when the portfolio of sales of firms operating in the sector of origin is orthogonal to the firms' portfolio of activity operating mainly in the sector of entry. Note that we cannot replicate Hoberg and Phillips (2010)'s product market distance based

building in close sectors or buying in distant ones. Specifically, we remove firms that shift a substantial part of their activities in the sector of entry (as they may want to build to reallocate their workforce from an under-performing business segment to a better performing one, see e.g., Baghai et al. (2020a)) and serial acquirers (as they may buy in distant sectors for risk-diversification or empire building motives, see e.g., Golubov, Yawson, and Zhang (2015)). The HC Distance coefficient remains statistically significant, suggesting that the explored alternative determinants of the decision to build or buy do not confound the role of human capital.

Related literature. Our paper contributes to the rapidly growing literature on labor and finance, and specifically, to a strand of papers that considers human capital as a determinant of M&As.[12] Ouimet and Zarutskie (2016) and Chen, Gao, and Ma (2020) show that human capital acquisition is an important motive of M&As. Closely related to our paper, Tate and Yang (2016b) and Lee, Mauer, and Xu (2018) find that diversified M&As are more likely to take place between industries that share similar human capital.[13] Our contribution to this literature is threefold. First, our unique data allow us to measure firms' human capital at the firm-level. Our measure predicts the choice of the sector of entry and the decision to build or buy. Second, our analysis goes beyond cross-industry M&As and considers organic entries as an alternative mode of entry. We show that firms build in sectors that share a similar human capital and buy in more distant sectors that require a human capital they do not employ. Third, we show that firms buy to acquire occupation-specific human capital when hiring workers on the external job market is too costly.

We also contribute to the small literature on the determinants of the decision to build or buy. Nocke and Yeaple (2007) analyze the choice to export rather than to engage in cross-border M&As (see also Frésard, Hege, and Phillips (2017)). Phillips and Zhdanov (2013) and Akcigit, Celik, and Greenwood (2016) explore how firms' patent acquisitions affect their own

on textual analysis of the 10k filings of US public information that is not available in France for both public and private firms.

[12]Other papers focus on the consequences of M&As on employment and wages (Tate and Yang, 2016a; Dessaint, Golubov, and Volpin, 2017; Ma, Ouimet, and Simintzi, 2018; Lagaras, 2019; He and le Maire, 2020). Other strands of the labor and finance literature has recently explored the implications of labor for corporate investment, cash holdings and capital structure decisions (Merz and Yashiv, 2007; Matsa, 2010; Agrawal and Matsa, 2013; Simintzi, Vig, and Volpin, 2014; Baghai et al., 2020b; Serfling, 2016; Israelsen and Yonker, 2017; Ghaly, Anh Dang, and Stathopoulos, 2017; Hombert and Matray, 2017) and for asset prices (e.g., Eisfeldt and Papanikolaou, 2013; Eiling, 2013; Donangelo, 2014; **?**).

[13]Similarly, firms are more likely to merge when they are alike (Rhodes-Kropf and Robinson, 2008), sell similar products (Hoberg and Phillips, 2010), rely on similar technology (Bena and Li, 2014) or share similar corporate cultures (Li, Qiu, and Shen, 2018).

innovation activities. We contribute to this literature in two ways. First, we provide an economy-wide analysis of build and buy entries into new sectors and show that M&As represent only approximately 2% (15% when weighted by sales) of entries in new sectors in France.[14] Second, our results highlight that after absorbing targets' human capital, firms that buy diversify in sectors that are more distant from their pre-entry human capital. Our work relates in that sense to research investigating the role of human capital acquisition in firms' growth path (e.g., Custódio, Ferreira, and Matos, 2019; Akcigit, Alp, and Peters, 2021).

Finally, our paper relates to a large body of work exploring the determinants of the firm's scope of activity (Lucas, 1978). Research in both corporate finance (Maksimovic and Phillips, 2002) and international trade (Bernard, Redding, and Schott, 2010) finds evidence that the number of sectors (or products) of the firm is determined by its ability to be productive for a wide range of activities.[15] Hoberg and Phillips (2018, 2020) and Boehm, Dhingra, and Morrow (2019) explore how firms determine their portofolios of activities and show that firms co-produce in sectors that require similar inputs. Similarly, our paper shows that firms' human capital predicts in which sectors firms diversify and that the acquisition of occupation-specific human capital determines whether they build or buy upon entry. Furthermore, our results imply that firms do not face a fixed set of resources, but rather, that this set evolves with the decisions they make along their life-cycle. We show that M&As have the potential to radically modify the firm's human capital, thereby reshaping the firm's subsequent diversification decisions.

2. Hypothesis Development and Empirical Strategy

This paper studies how human capital determines *which* sectors firms enter, and *how* they enter these sectors. This section develops testable hypotheses and proposes an empirical strategy to test them. Appendix B includes a model that theoretically backs the predictions tested in the analysis.

[14] Bernard, Redding, and Schott (2010) document that US multi-product firms often vary their product mix but infrequently do so through an acquisition (only 7% of cases). Only a few other papers in economics and finance have looked at the decision to build or buy in the context of diversification McCardle and Viswanathan (1994); Elsas, Flannery, and Garfinkel (2014). The strategy literature has been more prolific on this question (e.g., Yip, 1982; Lee and Lieberman, 2010; Capron and Mitchell, 2012).

[15] A related strand of the literature studies business groups or conglomerate firms' value creation through reallocation of capital (Shin and Stulz, 1998; Schoar, 2002; Giroud and Mueller, 2015) or labor (Tate and Yang, 2016a; Cestone et al., 2018; Faccio and O'Brien, 2018) when they face sectoral demand shocks.

2.1. Hypothesis development

Choice of the sector. We start by studying which sector firms enter. We conjecture that firms enter a sector that requires similar skills for three reasons.

First, if growth opportunities are not perfectly observable (Kirzner et al., 1978; Gompers et al., 2010), firms may be better at identifying opportunities in sectors requiring skills that overlap with those of their existing employees. Indeed, firms are better able to detect synergies and accurately forecast their chance of success in sectors that are close in terms of HC Distance. Second, a firm that enters a sector can train existing workers to learn new skills or hire new workers who already have the right skill set. However, this process is costly and frictional: for instance, hiring costs account for about one-third of wages and they increase in the specificity of skills (Abowd and Kramarz, 2003; Blatter, Muehlemann, and Schenker, 2012). Therefore, in order to leverage their existing human capital, firms are likely to enter sectors that require a similar skill set, that is, close sectors in terms of HC Distance. Third, expanding in related sectors allows firms to benefit from economies of scope by exploiting skill or knowledge synergies across activities (Panzar and Willig, 1981; Henderson and Cockburn, 1996).

Build or buy? Next, we ask how firms enter the new sector. We argue that firms' human capital explains the decision to build or buy in the new sector.

On the one hand, building becomes more attractive as the HC Distance to the sector of entry decreases. A low HC Distance implies that the firm already employs the occupation-specific human capital to operate in the new sector. Therefore, we expect economies of scope to be large in this case, so that the adjustment costs associated with the training of existing workers or the hiring of complementary workers are limited. Moreover, existing workers are more likely to invest in human capital that is complementary to their existing skills (Cunha and Heckman, 2007), which facilitates the integration of the new skills required for the new sector into the existing production process.

On the other hand, buying becomes more attractive as the HC Distance to the sector of entry increases. In more distant sectors, acquiring a target's human capital allows acquirers to avoid the search and hiring costs of the key workers for the new sector. By definition these costs are higher for distant sectors because firms would need to hire many workers to enter a sector that is distant in terms of HC distance. In closer sectors, buying is likely to be more

costly. Indeed, the costs of firing duplicate employees involved with the restructuring process following M&As (Maksimovic, Phillips, and Yang, 2013; Dessaint, Golubov, and Volpin, 2017) are larger in close sectors than in more distant sectors. Relatedly, when firms buy in close sectors, the integration process may create a more competitive environment for the target's and the acquirer's workers, which may impede synergies between teams or may lead the most productive workers to leave (Lagaras, 2019; Agrawal and Tambe, 2019; Fulghieri and Sevilir, 2020). As a result, firms should build in close sectors and buy in more distant sectors.[16]

Human capital adjustments and post-entry performance. So far, we have developed hypotheses about the ex-ante effects of human capital on the choice of the sector and the decision to build or buy. To confirm that human capital is a key channel behind the decision to build or buy in the new sector, we test three separate sets of hypotheses. First, the human capital channel implies that firms build less and buy more when hiring workers is more costly (section 5.1). Second, our hypothesis implies that building firms adjust their workforce after entry by hiring specialized workers for the sector of entry, whereas buying firms acquire targets to take over their specialized human capital and thus maintain their target's workforce after the acquisition (section 5.2). Finally, because hiring and integrating new workers to the production process takes time, and because buying firms directly access their target's human capital, we expect the performance of building firms to be worse than the targets of buying firms in the short-term (section 5.3).

2.2. Empirical strategy

Choice of the new sector. We start by studying the firm's choice to enter a new sector. We focus on the choice of sector of entry conditional on diversifying at least once in the sample period.[17] We use our measure of firms' HC Distance to the sector of entry to predict the firm's

[16] In our model reported in appendix B, the firm chooses its optimal scope of activity based on its core resources and the profitability of entering a new sector by building or buying. When a firm builds in a new sector, its marginal cost of production in the new sector depends on its core competence and its distance to the sector of entry (Eckel and Neary, 2010). When a firm buys an incumbent company in the new sector, it accesses the target's resources (including its human capital) to operate in the new sector. Moreover, we assume that acquiring firms can better assimilate workers with skills similar to those used in their sector of origin. We formally show that in the presence of acquisition costs, existence of human capital distance threshold such that it is optimal for a firm to build in sectors close to its core competence and buy in more distant sectors.

[17] We focus on the choice of the new sector, instead of the decision to diversify per se. We do so for two reasons. First, we consider all sectors in which firms could enter at any given point in the sample period. This procedure yields a sample of more than 86 million observations, such that it is computationally challenging to also consider firms that do not diversify in the sample period. Second, we are mainly interested in the firms' choice of the sector of entry rather than changes in the total number of sectors firms operate in.

choice of a sector of entry. We expect that firms enter sectors that are similar in terms of human capital. We estimate a linear model for the probability of firm g entering industry n between time $t-1$ and t:

$$\mathbb{1}(\text{Diversify})_{g,n,t} = \lambda_{g,t} + \lambda_{o,n,t} + \delta_d \cdot \text{HC Distance}_{g,n,t\text{-}1} + \beta_d' X_{g,n,o,t-1} + \eta_{g,n,o,t} \qquad (1)$$

where $\mathbb{1}(\text{Diversify})_{g,n,t}$ is a dummy variable that takes value one if firm g that mainly operates in the sector of origin o enters sector n in year t, and zero otherwise. $\lambda_{o,n,t}$ is the sector of origin $\times$ entry $\times$ year fixed effect which captures any economic changes that determine entry into a particular sector at a particular point in time (such as demand or cost shocks that affect all incumbent firms in n), but also all stocks that may make firms in sector o more or less likely to enter sector n (such as merger waves, or shocks affecting the input-output structure of the economy). In another specification, we add a firm $\times$ year fixed effect, $\lambda_{g,t}$, to capture unobservable firm-level determinants of the decision to diversify in a given year (such as the average rate of entering sectors for each firm-year). In that case, the regression identifies the direction of change in the industry space and not changes in the number of sectors the firm operates in. We also add a vector of control variables $X_{g,n,o,t-1}$, which includes other firm characteristics that may influence the choice of the sector of entry.[18] Standard errors are clustered at the firm level, the sector of origin level, and the sector of entry level to account for correlations of errors within firms, across firms within the sector of origin, and across firms within the sector of entry, respectively.

The main independent variable, HC Distance$_{g,n,t-1}$, is the distance between firm g's human capital and the sector of entry n in year $t-1$ (see Section 3.1.3). In this specification, HC Distance is defined to every potential sector of entry n where the firm can enter. We expect $\delta_d < 0$, which implies that firms are on average more likely to enter sectors that are close in terms of human capital.

Decision to build or buy. Next, we analyze the firms' decision to build or buy conditional on the sector of entry. We consider the decision to build *or* buy within a given sector of entry

[18] 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 the firm size) by the number of employees because this variable is additive when we aggregate subsidiaries to the firm level. Appendix B contains the definitions of the variables used in the empirical analysis.

and we estimate the following regression:

$$\mathbb{1}(\text{Buy})_{g,n,t} = \lambda_{n,o,t} + \delta_b \text{HC Distance}_{g,n,t-1} + \beta_b^{'} X_{g,n,t-1} + \varepsilon_{g,n,t}, \quad (2)$$

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 3.1.2). The main independent variable is still HC Distance$_{g,n,t-1}$. The vector $X_{g,n,t-1}$ includes the same control variables as in (1). $\lambda_{n,o,t}$ is the sector of origin $\times$ entry $\times$ year fixed effects that allows comparing firms that build and firms that buy in the same sector of entry n, in the same year t, and that operate in the same sector of origin o. The fixed effects capture unobservable synergies between sectors, as well as unobservable factors related to the sector of entry (e.g., fixed costs and barriers of entry) or the sector of origin (e.g., demand shocks). We double-cluster standard errors at the sector of origin and sector of entry levels to control for correlations across firms within the sector of entry and within the sector of origin.

$\delta_b < 0$ would imply that acquisitions dominate in sectors that are close to the firm in terms of human capital, in line with "like-buys-like" theories of M&As. In contrast, $\delta_b > 0$ would imply that firms are more likely to buy when their human capital is relatively distant from the sector of entry. Thus, on average, firms would be more likely to build in the new sector when it is close in terms of human capital.

Specific empirical concerns. To mitigate the role of unobservable determinants of the decision to build or buy, we estimate equation (2) on a matched control group of firms that build. We construct the matched sample of building and buying 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. More specifically, we retain build entries within the same (origin, entry, size quartile) triplet for every buy entry. In an alternative specification, we add firm fixed effects in (2), thus analyzing the effect of a firm's human capital on the decision to build or build within firms over time (see Table 4).

The interpretation of our findings is based on the assumption that our main independent variable HC Distance is not correlated with omitted variables that also drive the choice of the sector of entry or the decision to build or buy. Such a correlation would occur if, for instance, the firm hires employees in key occupations for the sector of entry in anticipation of the mode of

entry. We deal with this endogeneity concern in Section 4.5. Finally, our analysis is also based on the assumption that our results are not affected by common unobservable latent variables that may drive the choice of the sector and the decision to build or buy. We relax this assumption in Section 4.4.

3. Data and Descriptive Statistics

3.1. Data sources and main variables

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

3.1.1 M&A deals

We collect M&A deals from SDC Platinum and Bureau van Dijk Zephyr between January 2003 and December 2014. We exclude leveraged buyouts and private equity deals from the sample. We focus on majority deals in which the acquirer owns less than 50% of the target shares before the acquisition and more than 50.1% 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. We build a Python webcrawler to retrieve them.[19] These steps allow 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.

[19]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.

3.1.2 Firm-level data

Ownership link dataset. Our analysis is at the business group level (hereafter, the "firm"). A firm includes the parent company and majority-owned subsidiaries connected through ownership links.[20] By placing the analysis at the firm level, we ensure that we observe the same information as managers when deciding on entries in new sectors (e.g., sales and workforce composition of the subsidiaries). We use the ownership link dataset (*Enquête sur les Liaisons financières entre sociétés*, LIFI) to retrieve the structure of firms and identify all entities linked to the diversifying firm. We also use the ownership link dataset 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.[21]

Tax files. The tax files (*Bénéfices Industriels et Commerciaux*) provide detailed yearly accounting information (balance sheet and income statements) at the subsidiary level.[22] We retrieve cash holdings, total sales, 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 include 272 different sectors. We use the firm registry (*SIRENE*) to obtain the geographic locations of firms' establishments. Consistent with the approach on firms involved in M&A deals, we consolidate all variables at the business group level.

Build or buy? To identify entries into 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.[23]

[20] 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.

[21] In addition, using the ownership dataset, we verify that the target of a 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.

[22] We exclude from our sample firms in the financial, agricultural and public sectors because they use different accounting systems. We focus on firms filing under the standard tax regime (*Régime normal*). We also discard firms that use the simplified regime (*Régime simplifié*) because some important information is not reported. Firms in the standardized regime represent on average 94% of total value-added.

[23] The data are available at the subsidiary-level and are exhaustive for entities with at least 20 employees. The survey randomly includes smaller firms and covers at least 85% of sales realized in every sector. We check that subsidiaries included in the survey cover 96% of sales in the manufacturing sector. We assume that smaller

A firm enters a new sector if at least one of its subsidiaries begins selling in that sector and if none of its other subsidiaries had been operating in the sector the year before the entry. A firm "buys" into a new sector, i.e., enters by acquiring an incumbent company, if it acquires a subsidiary that enters the new sector simultaneously to enter the new sector. By contrast, a firm "builds", i.e., enters organically, if the entering subsidiary was already controlled by the firm before the entry or created the year of the entry in the new sector. We impose an entry threshold to ensure that the entry is economically meaningful from the firm's perspective. We impose that sales in the new sector represent at least 5% of the firm's total sales in other sectors the year before the entry. We also check that buy entries are not the byproduct of horizontal acquisitions. We remove buy entries for which the acquiring firm and the target company realized more than 50% of their sales in the same sectors before the acquisition. We check our results' robustness to these definitions in Table A3 reported in the appendix.

3.1.3 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.[24] 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).

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

$$\text{HC Distance}_{g,n} = 1 - \frac{\sum_i s_{g,i} \cdot s_{n,i}}{\sqrt{\sum_i s_{g,i}^2}\sqrt{\sum_i s_{n,i}^2}}, \tag{3}$$

where $s_{g,i}$ is the share of employees in firm g employed in occupation i, and $s_{n,i}$ the share of employees in sector n employed in occupation i.[25] The vector s_n represents the occupation

subsidiaries for which the sales breakdown is not available in the survey sell only in their main sector of activity, which is directly retrieved from the tax files.

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

[25] For simplicity, we omit the t subscripts from the explanations even though our distance varies across years. Note that because we do not observe the allocation of employees across the firm's sectors of activity, we only retain firms operating only in sector n for the computation of $s_{n,i}$.

composition of the "representative firm" resulting from the consolidation of all single-sector firms in sector n.[26] Thus, HC Distance is equal to one minus the cosine similarity between vectors s_g and s_n.

The variable HC Distance$_{g,n}$ ranges from 0 to 1, with 1 being the maximum possible human capital distance from firm g to sector n. Intuitively, HC Distance measures the extent to which firm g's production function (as reflected by its occupational mix) resembles the representative firm's production function in sector n. If the firm employs the same occupations as the representative firm in the new sector n, the firm's HC Distance to the sector of entry is small. We expect the firm to perform well in the new sector without having to adjust its workforce composition. Instead, if the firm employs different workers than those employed by the "representative firm" in the new sector, the HC Distance is large, and we do not expect the firm to perform well in the new sector. Another way to think about HC Distance is that it yields a unique Hotelling-like human capital space between firms and sectors (see Hoberg and Phillips, 2016, 2020, for a different application).

As an illustration, suppose that when aggregating all the firms operating in the civil engineering sector (1), we find that the aggregate workforce in the sector is evenly divided among 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 divided between 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 both firms' human capital would be equidistant to the civil engineering sector.

3.1.4 Local labor market tightness

To test the role of local labor market (LLM) tightness, we retrieve information collected by the French national unemployment agency (*Pôle emploi*). The data available from 2010 to 2014

[26] 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 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 employees in a firm's workforce employed in occupation i, we use the share of employees $s_{o,i}$ in a given sector of origin o's aggregate workforce employed in occupation i as follows: HC Distance$_{o,n} = 1 - \frac{\sum_i s_{o,i} \cdot s_{n,i}}{\sqrt{\sum_i s_{o,i}^2}\sqrt{\sum_i s_{n,i}^2}}$ We use these alternative definitions of HC Distance in Table 6 and Appendix Table A2.

identifies 350 different LLMs, which corresponds to commuting zones. An occupation is flagged as "in short supply" in an LLM if (i) the number of job ads posted on the 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. We use the list of occupations "in short supply" by LLM to construct a sector- and LLM-specific measure of LLM Tightness as follows:

$$\text{LLM Tightness}_{z,n,t} = \sum_i \mathbb{1}(\text{Occupation in short supply})_{i,z,t} \cdot s_{n,i,t}$$

where $\mathbb{1}(\text{Occupation in short supply})_{i,z,t}$ is a dummy variable that is equal to one if occupation i is in short supply in z. $s_{n,i}$ is the share of employees employed in occupation i in sector n. A high value of $\text{LLM Tightness}_{z,n,t}$ means that firms in the area z are likely to face difficulties finding workers in occupation i that are key to operate in sector n. For each entry in a new sector n, we use the value of $\text{LLM Tightness}_{z,n}$ in the local labor market z in which the subsidiary that enters the new market is located.[27]

3.2. 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 we weight entries by sales in the sector, the year of the entries, buy entries represenon average of 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% across the French SIC's different industry classification levels (Panel B).

[Insert Table 1 here]

Table 2 reports summary statistics on the variable used in our analysis. In the regression sample, buy entries represent about 2% of the entries, such that 98% are build entries. Besides,

[27] In the case of a multi-establishment subsidiariy which operates in several LLMs, we consider entries in each LLMs in which the subsidiary is located.

within firms that enter a new sector at least once during our sample period, the entry is made thought 29% of their subsidiaries.

HC Distances. Regarding the main independent variables used in our analysis, we find that the average HC Distance between the diversifying firm and the sector of entry that is actually chosen is 0.76, suggesting that the average firm lacks employees in occupations that are key to operate 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 weights overlapping occupations by the number of employees in the sector of entry. If instead, we use the number of hours or wages as weights, the HC Distances remains stable with a mean and a standard deviation very close to the baseline measure. These figures suggest that the distribution of wages and hours worked are similar to the distributions of employees by occupation. We also calculate the average HC Distance between the diversifying firm and incumbent firms in the sector of entry. The (simple or weighted) average HC Distances are larger than the baseline measure, suggesting that firms within the sector of entry display significant variations in their human capital composition.

In addition, we report HC Distance$_{g,n,t_0}$, calculated at t_0, which is the first year the firm appears in the sample period, corresponding to 4.9 years before the entry. HC Distance$_{g,n,t_0}$ is slightly larger than the baseline measure, HC Distance$_{\text{g,n,t-1}}$, suggesting that firms' human capital composition is sticky over time but gets closer to the sector of entry's representative firm. The sector-level measure of HC Distance, HC Distance$^{\text{Sector}}_{\text{o,n,t-1}}$ defined by equation (5), is also very close to the baseline measure, HC Distance$_{\text{g,n,t-1}}$.

We create a heatmap based on the HC Distance$^{\text{Sector}}_{\text{o,n,t-1}}$ between sectors in 2014. We report in Figure 2 HC Distance between sectors at a one-digit sector classification (A21). Each cell of the heatmap corresponds to the HC Distance between the two sectors located on the cell's 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 mean more distant sectors, whereas lighter cells mean 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).[28] Others are more or less distant

[28] 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.

from any given sector (e.g., the *Consultancy and Engineering* sector is close to the *Scientific and Technical Activities* sector but distant from the *Accommodation and Food Services* sector).

[Insert Figure 2 here]

Other variables. Table 2 also reports the distributions of the control variables we include in our main tests. The average firm in our sample employs 36 workers (= $e^{3.60}$), hired in 14 different occupations (= 36 × 0.4). It produces approximately €60,000 in value-added per worker, owns €60,000 in tangible assets per employee, and holds €30,000 in cash per worker. The average firm is 22 years old and invests €65,000 at the time of the entry. 47% of diversifying firms were already diversified before the entry into the new sector. Only 2% of them include a listed subsidiary in the business group.[29]

Firms report on average €1 million in sales in the sector of entry in the first year, with a large dispersion around the mean. Besides, 18% of the diversification event identified at the 3-digit of the French SIC happens to be in the same 2-digit industry, and 35% happens in the same 1-digit industry. In addition, we find that 24% of the diversification occurs in an upstream sector and 19% in a downstream sector. In addition, only 4% of the diversification occurs in a different region, and 7% occurs in a different department. 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]

4. Main Results

4.1. Human capital and corporate diversification

We begin the analysis by investigating what drives the firm's choice to enter a new sector. Specifically, we test that the firm's HC Distance to the sector of entry explains in which sector the firm diversifies. Our approach consists of computing the firm's HC Distance to every potential sector where the firm could diversify, and then of explaining the actual choice of the sector,

[29] Our sample counts approximately 700 (=2% ×35,000 observations) entries into new markets involving the subsidiary of a publicly listed firm. This figure can be compared to the number of current publicly listed firms in France, which is equal to approximately 750 firms.

the firm makes by the HC Distance.[30] 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. The main independent variable, $\text{HC Distance}_{\text{g,n,t-1}}$, is high when the diversifying firm employs only a few employees in key occupations for the sector of entry n. Table 3 reports the results with different specifications and sets of fixed effects. In column 1, the results show that firms are less likely to enter sectors that are distant in terms of human capital. We find a negative correlation between the choice to enter sector n and HC Distance. In column 2, we add firm-level control variables that could also drive the choice of the sector of entry. We find that the coefficient of our main independent variable does not change in the presence of these control variables . In column 3, we add sector of origin $\times$ entry $\times$ year fixed effects to account for unobservable complementarities between sectors that could drive the entry in a given sector. The HC Distance coefficient drops by a factor of two but remains significantly different from zero. A one-standard-deviation increase in the firm's HC Distance to the sector of entry (0.21) is associated with a 0.53 percentage point (p.p.) decrease in the probability to diversify in a given sector.

The relationship holds when we add firm-level controls in column 4. Finally, we add firm $\times$ year fixed effects in column 5 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. Moreover, since the coefficient is only estimated on firms that diversify at least once in a given year, this specification allows focusing on the role of HC Distance on the choice of the sector of entry rather than the choice of diversifying or not. We find that the relationship still holds, with a coefficient not statistically significant from the previous specification.

[Insert Table 3 here]

4.2. Human capital and the decision to build or buy

Next, we analyze the firm's decision to build or buy conditional on the sector of entry. We test that firms are more likely to buy when their human capital is distant from the sector of entry, and that they build in closer sectors. The dependent variable, $\mathbb{1}(\text{Buy})_{g,n,t}$, is a dummy

[30] Note that this specification multiplies the number of observations by the number of potential sectors and by the number years the firm can potentially enter into.

variable equal to one if firm g enters a new sector n in year t through the acquisition of an incumbent firm, and zero if it enters by building on its existing human capital. The main independent variable is HC $\text{Distance}_{\text{g,n,t-1}}$ which is high when the diversifying firm employs few employees in the key occupations for the sector of entry. Table 4 reports the results with different specifications and sets of fixed effects.

The baseline specifications in columns 1 and 2 include 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 the same year t, while controlling for unobservable synergies between the sector of origin, and that of entry.

We show that the firm's HC Distance to the sector of entry is positively correlated to the probability of acquiring an incumbent company in the sector of entry. A one-standard-deviation increase in HC Distance is associated with a 0.57 to 0.71 p.p. increase in the probability to buy (columns 1 and 2). This relationship is sizable and is equal to approximately 25% of the unconditional probability of buying and is significant at the 5% level. The point estimate is not significantly different when we add control variables in column 2, suggesting that the control variables are mostly uncorrelated with the HC Distance.

In columns 3 to 4, we test the robustness of our main finding to different combinations of fixed effects. In column 3, we include interacted sector of origin $\times$ entry $\times$ quartile of firm size fixed effects to compare firms of the similar size in the sector of origin and before the entry in the same sector. In column 4, we compare firms that initially operate in the same *main* sector of origin, same *second* sector of origin, and enter the same sector. This specification accounts for unobservable synergies between the firm's main and secondary sector of activity and the sector of entry. The HC Distance coefficient is robust to these alternative combinations of fixed effects and remains very similar in these specifications.

In column 5, we add firm and year fixed effects in addition to the interacted sectors fixed effects. The HC Distance coefficient drops by half but 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 unobservable omitted firm-level variables.

Columns 2 to 4 include firm-level control variables. The evidence shows that firms that

employ a relatively more diverse set of occupations are more likely to buy. In line with the existing literature on the determinants of M&A deals, we find that larger firms, cash-rich firms, and high-wage-paying firms are significantly more likely to buy . However, several of these relationships become negative when comparing the same firm over time (column 5), suggesting that firms that enter multiple sectors over time tend to buy at a stage of their life-cycle in which they are relatively smaller and less profitable than when they build (e.g., declining firms).

In conclusion, the empirical analysis presented in Table 4 supports our main hypothesis. Firms build in close sectors and buy in distant sectors, that is, in sectors for which they do not have the right human capital. Our results highlight the acquisition of complementary skills as a motive of M&As and economies of scope as a determinant of building. We show that our results hold in the cross-section of firms, as well as within firms over time. It is also robust to various combinations of sector fixed effects and to the introduction of other determinants of M&As.

[Insert Table 4 here]

4.3. Robustness checks

Alternative measures of HC Distance. In appendix table A2, we test the robustness of our results to alternative measures of HC Distance. The baseline measure of HC Distance uses the exact firms' workforce composition and weights occupations by the number of workers employed in this occupation in the sector of entry. This measure may underestimate the weight of a few key individuals, e.g., CEO, or overestimate the weights of numerically important occupations, e.g., part-time workers.[31] In addition, using the aggregate composition of firms' workforce in the sector of entry may over-weight larger firms' idiosyncrasies and dim firms' heterogeneity in their workforce composition. In panel A, columns 2 and 3, we consider the firms' workforce composition weighted by the share an occupation represents in the total wage bill and in the number of hours worked. In panel B, we replace HC Distance with the average distance between the diversifying firm and incumbent firms in the sector of entry. We put different weights on incumbent firms (equal weights, total sales, number of employees) to account for the relative importance of firm size in the definition of HC Distance (see Section 3.1.3 for more details about the alternative measures). Consistent with our baseline results, we find that firms build in close

[31] This definition of HC Distance also addresses the specific concern that CEOs' and other high-level executives' wages may not reflect only their contribution to the firm's performance and maybe the result of agency conflicts.

sectors and buy in distant sectors. The results are statistically significant and economically close to the baseline specification.

Alternative thresholds in the sector of entry. In appendix table A3, we test the robustness of our specification to the entry size threshold used to define our estimation sample. We define the entry size as the ratio of sales realized in the new sector relative to total sales the firm realized in other sectors the year before entry. In column 1, we do not impose any threshold. Entries equal to one euro are included in the sample. In columns 2 to 4, we impose an entry size threshold of respectively equal to 1%, 5% (baseline), and 10%. The magnitude of the effect of HC Distance on the likelihood to buy is economically and statistically stronger when we impose a higher threshold of entry sales, suggesting that the decision to buy or build is more sensitive to the firm's human capital when the entry in the new sector is meaningful at the scale of the diversifying firm. In other words, when the entry is small, possibly driven by experimental motives, buying an incumbent firm is less of an option, and the workforce composition being not as crucial as in larger entries. We confirm this intuition in column 5 by running a weighted least square (WLS) regression, where observations are weighted by entry sales.

HC Distance within firms that build. 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 A4, we change the unit of observation and calculate HC Distance at the subsidiary-level. We test that the entry in the new sector is made by subsidiaries that are close to the sector of entry in terms of human capital. Specifically, within firms that build, 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, i.e., low HC Distance subsidiary.[32] Our results confirm that within the firm, the entry is more likely to be made through the subsidiaries with a low HC Distance to the sector of entry. This finding not only validates our measure of HC Distance to explain firms' entry decisions but also suggests that

[32]We test this prediction for build entries only. It is unclear whether the human capital of firms that buy' subsidiaries necessarily predicts which subsidiary should 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, and zero otherwise. We include firm $\times$ 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 stand-alone firms as it relies on the variation in HC Distance between subsidiaries.

firms try to minimize labor adjustment costs when entering a new sector.

4.4. Self-selection and corporate diversification

In the previous sections, we study the selection of the sector of entry and decide to build or buy independently. However, we observe the decision to build or buy conditional on the firms have selected a given sector of entry. Thus, if latent unobservable factors drive both the choice of the sector and the decision to build or buy, our estimates may be biased, and our interpretations misled. Specifically, our estimation of equation (2) is subject to a self-selection bias if the residuals ϵ of this equation are correlated with the residuals η in the selection equation (1). Assuming that η and ϵ, the residuals of the selection and entry mode equations, are jointly normal with a correlation term γ, we get that the estimated coefficient $\hat{\delta}_{Buy}$ obtained from the estimation of equation (2) verifies under the standard Heckman (1979) setting

$$\hat{\delta}_{Buy} = \delta_{Buy} - \gamma \cdot \Gamma(X) \cdot \delta_{Diversify} \tag{4}$$

where γ is the correlation between these error terms and $\Gamma(X)$ is the function of the observable characteristics that co-determine the choice to build or buy. $\Gamma(X)$ can be shown to be positive.

In Section 4.1, we have estimated equation (1), and we have shown that $\delta_{Diversify} < 0$, implying that firms are unconditionally less likely to diversify in distant sectors distant in terms of human capital. Therefore, the self-selection bias may lead to overestimating δ_{Buy} if the correlation γ between the error terms is positive.[33]

In theory, self-selection bias can be addressed using a Heckman selection model (Campa and Kedia, 2002; Li and Prabhala, 2007). However, this method requires finding an instrument that affects the choice to diversify in a given sector but not the type of entry (therefore, the instrument must be sector-specific). Finding such an instrument appears to be a difficult, if not impossible, task in our setting. We adopt a more targeted approach 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 attempt to isolate the bias and replicate our results.

[33]Note that we are not worried about cases where γ is negative because the self-selection bias would instead lead to underestimating the coefficient δ_{Buy}, which would play against us.

Shifting firms. First, we identify firms that operate in a declining sector and shift a substantial part of their activities in the sector of entry. Diversification is usually associated with good pre-entry performance (see section 5.3). Firms operating in a declining sector would be less likely to diversify in the first place ($\eta < 0$). However, conditional on diversifying, these firms would certainly try to shift their activity toward better-performing sectors that employ similar types of workers, i.e., low 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 estimated $\hat{\delta}_{Buy}$ coefficient in Section 4.2 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 5 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_{Buy} estimated without shifting firms are not statistically and economically different from those reported in table 4. 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 the specific case of serial acquirers. We define serial acquirers as firms that have bought at least two firms in our sample period. These firms represent 50% of buy entries in our sample and may buy companies for reasons that are disconnected from human capital considerations (Golubov, Yawson, and Zhang, 2015). For instance, serial acquirers may buy to diversify their portfolio's sector-specific risk or because of managers' optimism or for empire-building (Malmendier and Tate, 2008). As a result, serial acquirers may choose to enter more distant sectors ($\eta > 0$) and do so by buying incumbent comapanies in those sectors ($\epsilon > 0$). We exclude serial acquirers from our estimation sample and rerun our main build or buy model (equation (2). In column 4 of Table 5, we show that excluding serial acquirers does not significantly affect our estimate of coefficient δ_{Buy}.

Life-cycle. Third, firms may favor build or buy entries at a different stage of their life-cycle, e.g., experimental build entries in close sectors early on and conglomerate expansion by acquisition in more distant sectors at a later stage (Matsusaka, 2001; Hoberg and Phillips, 2020). To address the self-selection bias in relation to the firm's life cycle, we adopt a fixed-effects approach. We compare firms that build or buy coming from the same sector of origin, entering the same sector, and with the same number of sectors of origin before the entry (column 5),

and in the same age decile (column 6), respectively. Our baseline estimates once again remain unchanged.

In none of the above situations, self-selection can fully explain the decision to build or buy and overrule the role of HC Distance. While our approach can only invalidate some specific situations, we believe that this approach significantly raises the bar for other self-selection mechanisms that would lead to fully rule out the role of human capital in the decision to build or buy.

4.5. Endogeneity of human capital

Sector-level human capital. The interpretation of our results presented in Section 4.2 is based on the assumption that the firm's human capital is not correlated with unobservable omitted firm's characteristics that would drive diversification strategies. This condition may be violated if, for instance, firms that are more likely to diversify, systematically rely on different occupations than other firms. To deal with this endogeneity, we replicate our main results at a more aggregate level. We compute the total number of build and buy entries by sector of origin and sector of entry. We also calculate a sector-level HC Distance between the sector of entry and the sector of origin (see Section 3.1.3 for more details). Sector-level HC Distance is based on aggregate workforce composition by sectors. Thus, this approach ensures that unobservable firm-level characteristics do not confound our results. This specification also includes sector of origin, sector of entry, and year fixed effects, as well as sector-level control variables.[34]

Table 6, panel A reports the results. In column 1, we find that the sector-level HC Distance is negatively and significantly correlated with the (logarithm) total number of entries in the new sector. Consistent with the results reported in table 3, the evidence shows that firms are less likely to enter sectors distant from their sector of origin's human capital. In columns 2 and 3, we decompose the total number of entries between build or buy entries. We find that the sector-level HC Distance is also negatively and strongly related to the number of build and buy entries, respectively. However, the magnitude of the effect is stronger for build entries, suggesting that more firms tend to buy rather than building in distant sectors from the sector of origin. Column 4 confirms our results reported in Table 4. In this model, we keep only

[34]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. See definitions of the sector-level variables in Appendix B.

triplets of the origin-entry-year sector for which we observe both build and buy entries. We regress the fraction of buy entries over build entries on the sector-level HC Distance. We show that HC Distance between the sector of origin and that of entry predicts a higher frequency of buy entries relative to build entries in the sector.

Lagged human capital. The interpretation of our main results may also suffer from a reverse causality endogeneity bias if firms anticipate entry in the new sector and modify their workforce composition accordingly. For instance, firms may hire employees in key occupations the year before to build in the sector of entry. By contrast, firms may fire employees in overlapping occupations in anticipation of an acquisition in the sector of entry. To address this concern, we compute HC Distance$_{g,n,t_0}$ to the sector of entry n in the first year t_0 the firm appears in our sample period. The upcoming entry, on average, takes place 4.9 years after the firm first appears in our sample. We show that firms' human capital is sticky over time. We argue that firms' workforce composition at that time t_0 is unlikely to reflect changes made to prepare the upcoming entry in the new sector. Table 6, panel B reports the results.

In columns 1 and 2, we replicate our main results using HC Distance$_{g,n,t_0}$. We find that HC Distance at t_0 still predicts the sector of entry and the decision to build or buy in this sector at time t. Therefore, our results show that firms' human capital drives corporate diversification, not the opposite. that the evidence is based on the fact that firms' human capital is sticky over time, which arguably allows for a higher degree of exogeneity. However, these tests using HC Distance at t_0 do not impose any restrictions on firms' past activities reflected by the sector of origin at t_0. Thus, the estimated coefficients may pick up variations in human capital from differences in past growth trajectories. In columns 3 and 4, we compare firms' choice of a new sector and decision to build or buy across firms that enter the same new sector n the same year t and that come from the same sector of origin as observed at t_0 (the first year where the firm appears in our sample) instead of $t-1$. We find that the HC Distance$_{g,n,t_0}$ coefficient in the sector of entry choice regression drops by one-third. In addition, the HC Distance$_{g,n,t_0}$ coefficient is no longer significantly related to the decision to build or buy in the sector of entry. The evidence suggests that the firm's initial activity, defined by its sector of origin at t_0 determines to a large extent firms' workforce composition over their life-cycle. Our results' key identification comes from the path-dependence of firms' activities over their life-cycle, reflected in their sticky human capital. In section 6, we further investigate this intuition. We explore

how firms' sticky human capital shapes firms' subsequent diversification choices and how M&As actually enables firms to deviate from their initial path.

[Insert Table 6 here]

5. The Labor Channel in the Decision to Build or Buy

5.1. The role of local labor market tightness

We showed that firms build in sectors that are close in terms of human capital and buy in more distant ones. Given the cost associated with an acquisition, how do firms find it profitable to buy a target in the sector of entry instead of hiring new workers in key occupations? One answer is that firms may not be able to find the key employees for the sector of entry in the external job market. This is the case if key occupations are in short supply and thus, local labor markets (LLM) are tight for these occupations.

To test this hypothesis, we construct a measure of LLM Tightness based on occupations in short supply or in short supply in the location of the firm's entry (see section 3.1.4).[35] We estimate our main equation 2 on subsamples corresponding to terciles of the LLM Tightness$_{z,n,t-1}$. We add to our preferred specification LLM fixed effects. The first tercile corresponds to LLMs without key occupations in short supply for a given sector. The third tercile corresponds to tight LLMs where key occupations for the sector of entry are the most scarce. If human capital is a first-order concern to explain the decision to build or buy, the effect should be driven by entries in tight LLMs, when it is more difficult to find key workers. Table 7 reports the results, and figure 3 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 short supply in the LLM.

In columns 1 and 2, we regress the choice to build or buy in the sector of entry on firms' HC Distance in locations where workers employed in key occupations can easily be found on the external job market. We find a positive but not significant relationship between the likelihood to buy and HC Distance, suggesting that human capital does not play a significant role when it is not scare enough. By contrast, in column 3, we find that HC Distance is positively and

[35] 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 decreases the number of observations in the estimation sample.

significantly related to the likelihood to buy in the new sector when the LLM for the sector of entry's key occupations are tight. Hence, firms' decision to build or buy is driven by entries in the tightest LLMs. The evidence is consistent with firms that, despite the cost of buying, buy in distant sectors to overcome the difficulty of hiring new key workers for the sector of entry.

[Insert Table 7 here]

5.2. Human capital distance and post-entry labor adjustments

This section shows that entries into new sectors are associated with workforce adjustments that are specific to the selected mode of entry. The graphs above Table 8 plot the dynamics of post-entry labor adjustments four years before to four years after the entry. We plot labor dynamics of firms that build or buy separately as we expect labor adjustments to be different. We identify "high HC Distance firms" by computing a dummy equal to one of the firm's HC Distance to the sector of entry is above the median in each sector of origin $\times$ entry $\times$ year triplet.[36] We focus on the employment dynamics of the subsidiary responsible for the entry in the new sector. The main independent variable is HC Distance calculated at the firm level.[37]

The graphs show that high and low HC Distance firms do not display significant differences in employment size before the entry, neither for build nor buy entries. However, after the entry in the new sector, high HC Distance firms employ 2 to 4% more workers than low HC Distance firms when they build, suggesting that these find adjust their workforce more following the entry. In contrast, we do not observe significant post-entry differences between low and high HC Distance firms that buy.

We further analyze these post-entry labor dynamics and report the results in Table 8.[38] First, in column 1 of Table 8 Panel A, we regress the firms' that build employment growth from year $t-1$ to $t+3$ on the firm's HC Distance calculated the year before the entry. We compare firms that operate in the same sector of origin, build in the same sector of entry in the same

[36]The graphs plot the coefficients of the regression of the (log) number of employees and include subsidiary fixed effects, sector of origin $\times$ entry $\times$ year of entry fixed effects, and years to entry fixed effects. The high HC Distance dummy variable is interacted with years to entry fixed effects.

[37]Two main reasons motivate this specification. First, the decision to build or buy is likely to be made at the firm level (business group). Second, the entity that actually operates in the new sector after the entry is arguably the one that needs to adjust its human capital. As a result, we expect most of the changes in the employment composition to happen at the subsidiary level. In addition, note that in the case of buy entries, the subsidiary through which the entry is performed is the target company. In the case of build entries, this subsidiary is not directly associated with changes in business groups' workforce composition .

[38]Note that the number of observations in Table 8 is lower than in other estimation tables as we focus on firms surviving at least three years after the entry to compute the three-year-employment growth.

year. We show that HC Distance is significantly and positively associated with employment growth. A one standard deviation increase in pre-entry HC Distance is associated with a 1.5 p.p. increase in employment growth around the entry in the new sector. In columns 2 and 3, we decompose $\Delta\text{Employment}_f$ between the 5 most important occupations for the sector of entry, and the rest of occupations.[39] Our results show that high HC Distance firms grow significantly more in top 5 occupations than low-distance firms. Finally, in column 4, we show that a one standard deviation increases in HC Distance the year before the entry is associated with an 8.3 p.p. decrease in HC Distance to the sector of entry three years after the entry. Overall, Panel B's results show that firms that build restructure their workforce after entry in the new sector to bring their human capital closer to the sector of entry.

Next, we report the post-entry labor dynamics of firms that buy in Panel B of Table 8. We find a negative relationship between HC Distance and post-entry employment growth. We decompose the employment growth between top five occupations and the rest, and we show that target companies acquired by high HC Distance firms grow more in the top 5 occupations after the acquisition (column 2). However, targets' employment growth in other occupations is negatively related to the acquiring firms' HC Distance, suggesting that the target companies focus on occupations that are key for the sector of entry and layoff redundant workers in other occupations (column 3). In column 4, we find that the target firm's employment composition gets closer to the sector of entry's representative firm's employment composition. Overall, the evidence shows a null to a negative relation between the acquirer's HC Distance and the target's total employment growth, highlighting the idea that acquisiting key, complementarity human capital is the primary motive for the acquisition.

[Insert Table 8 here]

5.3. Human capital distance and post-entry performance

Next, we investigate the idea that not having the right human capital to operate in the sector of entry affects post-entry performance. We measure performance using realized sales in the sector of entry and the probability to survive in this sector after three years. The graphs above Table 9 plot the dynamics of post-entry sales for firms with a high HC Distance relative

[39] $\Delta\text{Employment}_f = \Delta\text{Top5 occupations}_f + \Delta\text{Other occupations}_f$, such that the coefficients in columns 2 and 3 add up to the coefficient of column 1.

to firms with a low HC Distance to the sector of entry. The plotted coefficients represent the difference in the (log) realized sales in firms with a high versus low HC Distance to the sector of entry. We also separately plot build and buy entries to compare post-entry firm performance, which we expect to vary depending on whether firms build or buy.[40]

The graphs above Table 9 show that high HC Distance firms realize about 25% lower sales in the sector of entry, relative to firms that have a low HC Distance. The gap is persistent but decreases over time, consistent with the idea that firms that build, hire significantly more employees in top 5 occupations after the entry to reduce their HC Distance and thus adapt their human capital to the sector of entry. Consistently, in Panel A of Table 9, we show that a one standard deviation increase in the firm's HC Distance the year before the entry is associated with 15% lower sales at the end of the year of entry and a 4.1 p.p. lower probability of survival in the sector of entry after 3 years (column 2).

Panel B of table 9 reports the entry sales and the survival probability of firms that buy in the sector of entry. Consistent with the graphical evidence, we do not find that the target's post-entry performance is significantly correlated with the acquiring firm's human capital before entry. This finding suggests that target firms are acquired for their human capital by acquirers missing the key human capital. We conclude that the target's human capital matters to explain firms' success that buy in the sector of entry, as opposed to the acquirer's human capital.

To summarize, when firms build despite a human capital that is not adapted to the sector of entry, they experience significantly lower sales in the sector of entry. They are significantly less likely to survive in this sector after three years. To mitigate this problem, firms that build adjust their human capital following entry and hire employees in the key occupations for the sector of entry, reducing their HC Distance to the sector of entry over time. Instead, firms' human capital does not explain the acquired target company's performance in the new sector. After the acquisition, the target's employment composition appears to refocus around key occupations for the sector of entry. These findings are consistent with our interpretation that the target's human capital motivates the acquisition to overcome shortcomings in the acquirer's human capital. Moreover, the evidence on post-entry performance and labor dynamics of building and buying firms validate the hypothesis that the firm's employment composition is a key

[40]Specifically, the graphs display the results of the regression of the (log) sales on sector of origin × entry × year of entry fixed effects and years to entry fixed effects. In addition, the high HC Distance dummy variable is interacted with years to entry dummies.

determinant of corporate diversification strategies.

[Insert Table 9 here]

6. Build, buy, and firms' growth path

Firms buy in sectors that are distant in terms of human capital (section 4.2), and acquire their target's human capital to successfully operate in the sector of entry (sections 5.2 and 5.3). Therefore, buy entries often lead to dramatic changes to the firm's human capital, build entries happen in close sectors and lead to incremental changes to the firm's human capital. In this section, we show that the choice to build or buy upon entry affects the firm's human capital differently, thereby also affecting which sectors the firm enters next. As a result, we expect future entries to depend less on firms' initial human capital after a buy entry relative to a build entry. Given that buy entries lead to more dramatic changes to the firm's human capital, we expect the initial human capital of firms that buy matter less for the next sector of entry than building firms. We propose two tests of this path dependence hypothesis in table 10.

First, in columns (1) and (2), we compare firms' choice of the subsequent sector of entry after having built or bought in their first entry. The subscript t_1 stands for the first entry into a new sector observed in our sample period. The dummy variable $\mathbb{1}(t > t_1)$ denotes events taking place after the first entry at t_1. The dummy variable $\mathbb{1}(\text{First entry by buying})_{g,t_1}$ takes the value one if the first entry at t_1 was a buy entry. The independent variable HC Distance$_{g,n,t_0}$ is the HC 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 (see Section 4.5). We test whether that firms enters sectors that are more distant from their initial human capital after a buy entry at t_1. We expect the coefficient of the triple interaction of these main independent variables to be positive.

Second, in columns (3) and (4), we focus on the sample of firms that buy at least once during the sample period. t_{Buy} stands the first year a firm buys. The dummy variable $\mathbb{1}(t > t_{Buy})$ denotes events taking place after the buy entry at t_{Buy}. The interaction of these two independent variables captures the extent to which subsequent entries depend on the initial firm's human capital, HC Distance at t_0. We expect the coefficient of the double interaction to be positive. We introduce firm fixed effects to test this effect within firms. All models reported in table 10

include sector of origin-entry-year fixed effects as well as time to entry fixed effects.[41]

The two tests give a consistent picture. After a first buy entry, the firm's initial human capital becomes relatively less predictive of the subsequent sector of entry compared to entries after a first build entry (columns 1 and 2). Overall, the evidence shows that acquisitions allow firms to access their target's human capital which is necessary to enter sectors that were previously too distant to enter. In other words, acquisitions reshape the firms' set of future growth opportunities and disrupt their path of growth.

[Insert Table 10 here]

7. Other Determinants of the Decision to Build or Buy

The decision to build or buy may depend on other considerations than human capital alone. This section considers other determinants of firms' decisions to build or buy.

Physical capital. HC Distance may capture differences in investment in physical capital needed to enter a new sector. Although the evidence on labor adjustments suggests that our results cannot fully be explained by the role of physical capital, in Appendix table A5, we explicitly control for investment in physical capital. In column 1, we rank firms into ten deciles by the amount of capital invested the year of the entry. We run our baseline regressions with interacted sector of origin $\times$ entry $\times$ investment decile fixed effects and year fixed effects.[42] This specification compares firms that operate in the same sector of origin, enter the same sector in the same year, and invest in similar amounts when entering the new sector. In column 2, we control for investment, and we add the (log) amount of capital expenditures at the time of entry. The point estimates are not statistically different from the baseline specification. In columns 3 to 5, we rerun our main specification in three sub-samples splits by sector-level capital intensity. We find that even when entering low capital intensive sectors (for which physical capital is less likely to matter), firms are still more likely to build in close sectors and buy in more distant sectors.

[41] Time to entry fixed effects are defined as $t - t_1$ in columns (1) to (2) and as $t - t_{Buy}$ in columns (3) to (4).

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

Size and access to financing. In Appendix Table A6, we study the role of financing constraint. We interact the firm's 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 column 1, the coefficient on the interaction between employment size and HC Distance is negative, suggesting that larger firms build in more distant sectors. In column 2, we find a negative, although not significant, the relationship between firms' cash holdings and HC Distance. The evidence also shows that diversified firms build in more distant sectors. Finally, in column 4, we find that listed firms are 14% more likely to buy. The interaction coefficient shows that listed firms are not significantly more likely to build in more distant sectors though. Overall, the evidence shows that the decision to build or buy is more sensitive to the firm's human capital in smaller firms and firms that are not diversified in the first place. The evidence suggests that larger and already diversified firms that have arguably a more diverse set of employees may be less finally constraint and find alternatives to buying and substitute for missing key human capital (e.g., internal labor markets).

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 choose to buy instead. In Appendix Table A7, we investigate the role of geographical distance between the initial firm's location and the effective location of the new activities. We construct two dummy variables, *Department* and *Region*, respectively, equal to 1 if the firm diversifies in a new geographic area defined as a department (column 1) and a region (column 2). The results reveal that firms are respectively 5.3% and 3.8% more likely to buy when the diversification is associated with a new department or a new region. The coefficients of the interaction between HC Distance and the geographical distances are not significantly different from zero. In addition, we show that 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. The evidence suggests that the role of geography in the firm's decision to build or buy is distinct from human capital.

Industry and product market distance. In appendix table A7, we also 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 $\mathbb{1}(SameIndustry)_{o,n}$ is equal to one of the firm's sector of origin

and sector of entry, defined at the three-digit industry code belong to the same one-(two-)digit industry code. To estimate the effect of industry distance on build or buy, we modify the set of fixed effects and use a combination of sector of origin × year and sector of entry × year fixed effects. In columns 3 and 4 of Appendix table A7, we find that firms are unconditionally and significantly less likely to buy in a different one-(two-)digit industry than within the same industry. The interaction between HC Distance and the dummy variable $\mathbb{1}(SameIndustry)_{o,n}$ is negative and significantly different from zero, suggesting that entering a distant sector amplifies the importance of human capital in the decision to build or buy.

Next, we construct a product market distance adapted from Bloom, Schankerman, and Van Reenen (2013).[43] The product market distance is equal to one when the sales portfolios of firms operating in the sector of origin are orthogonal to the sales portfolio of firms operating in the sector of entry and equal to zero when the two vectors are colinear. In columns 5 and 6, we show that the firms are more likely to buy in distant product markets. The HC Distance coefficient remains positive and significantly different from zero when controlling for the product market.

Firms may enter upstream sectors by acquiring suppliers or downstream sectors by acquiring customers to transfer goods along the production chain, irrespective of human capital considerations. In columns 7 to 10, we test the effects of upstream and downstream vertical connections 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.[44] We define vertical links between a firm and the sector of entry when more than 5% (10%) of the firm's sector of origin's inputs come from the sector of entry. We do not find that vertical integration considerations affect the decision to build or buy in a new sector. Moreover, the role of HC Distance is not significantly affected when controlling for industry vertical links.

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

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

with $v_{o,q}$ the share of sales firms operating mainly in the sector of origin o realize in sector q, and $v_{n,q}$ the share of sales firms operating mainly in the sector of entry n realize in sector q.

[44]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.

Sector of entry profitability and competition. Finally, we test the effects of sector of entry's characteristics on the decision to build or buy. First, we test that firms may build in more distant sectors from their core competence when expected profits in these sectors are high and offset the need for adapted human capital to build. In Appendix table A8, we use four proxies of expected profitability of the sector of entry: the number of firms created in the sector, the 3-year survival probability of new firms, the growth in aggregate sales, and the level of competition as measured by the Herfindahl index. We include a combination of sector of origin, sector of entry, and year of entry fixed effects to identify the effects of the sector of entry's characteristics interacted with HC Distance. We find supporting evidence for our hypothesis. The coefficient of the interactions between HC Distance and the third tercile of new firms creation and aggregate sales growth, respectively, is significantly different from zero and negative. The evidence suggests that firms build in more distant sectors when these sectors deliver high expected profits, measured by high levels of firm creation in the sector and aggregate sales growth. However, the evidence becomes less clear regarding the relationship between the decision to build or buy and the Herfindahl index and newly created firms' probability to survive.

8. Conclusion

Our paper shows that human capital shapes firms' diversification strategies. We develop a measure of "human capital distance" (HC Distance) between a firm and potential sectors of entry based on the overlap in worker occupations between the firm and sectors of entry. Our measure of HC Distance predicts *which* sectors firms enter and *how* they enter them. Using French administrative data, including the exhaustive matched employer-employee dataset and firms' breakdown of sales by sector, as well as M&A deals retrieved from SDC and Bureau van Dijk Zephyr, we construct a new and comprehensive dataset that to track firms' next entries into new sectors. Consistent with the presence of complementarities, we find that firms enter close sectors in terms of human capital, that are sectors that employ similar worker types. Conditional on the selected sector of, we then study of how firms perform the entry in a new sector, that is whether they build on their human capital or whether they buy and incumbent company in the sector of entry. The raw data shows that 98% of entries into a new sector (85% when weighted by entry sales) come from firms that build rather than the acquisition of an existing company. We find that firms build in closer sectors in terms of human capital and buy

in more distant ones.

This finding is explained by firms that seek to minimize labor adjustment costs when considering expanding the scope of the firm and deciding to build or buy in the new sector. We provide additional pieces of evidence consistent with this motive. First, within firms, entries are more often made through subsidiaries with a low HC Distance to the sector of entry, suggesting that firms that build intend to minimize labor adjustments within the internal labor market, in the first place. Second, firms choose to buy in a new sector when the cost of hiring workers in key occupations is too high. Using unique data about occupations in short supply, we create a measure of LLM Tightness. We show that firms are more likely to buy when their HC Distance to the new sector is high and when LLMs are tight, such that firms are likely to buy in closer sectors and incure the cost of an acquisition when hiring key workers on the exernal job market is costly. We then study the labor adjustments and performance implications associated with the decision to build or buy. We show that firms that build without having the key human capital for the sector of entry experience significant employment growth after entering a distant sector. We find that these firms adjust their human capital to the new sector; employment growth is driven by growth in key occupations for the sector of entry and decline in other occupations. Moreover, our results show that when entering by buying an incumbent firm in a new sector, acquiring firms adjust their target firm's human capital by increasing employment in the key occupations for the sector of entry and decreasing employment for other occupations.

Human capital distance in connection with the chose mode of entry also determines firms' post-entry performance. Firms that build in a new sector despite a high HC Distance realize about 25% fewer sales in this new sector than firms with more adapted human capital. These firms are also significantly more likely to exit the sector of entry within three years. Regarding buy entries, we do not find that the newly acquired target's post-entry performance depends on the acquiring firm's human capital. These findings are consistent with the idea that firms that build rely on their existing human capital, whereas firms that buy rely on their target's human capital to operate in the sector of entry successfully.

The decision to build or buy itself affects the firms' set of future growth opportunities through the acquired human capital or labor adjustments made around that time. We show that firms' initial human capital becomes less predictive of their subsequent sectors of entry when firms first choose to buy rather than build. The evidence suggests that firms that build

follow path-dependent trajectories determined by the initial firm's human capital. In contrast, acquisitions allow firms to disrupt the initial path of their growth opportunities by acquiring new skills.

References

Abowd, John M and Francis Kramarz. 2003. "The costs of hiring and separations." *Labour Economics* 10 (5):499–530.

Agrawal, Ashwini and David Matsa. 2013. "Labor unemployment risk and corporate financing decisions." *Journal of Financial Economics* 108 (2):449–470.

Agrawal, Ashwini and Prasanna Tambe. 2019. "Takeovers and endogenous labor reallocation." *Working Paper* .

Akcigit, Ufuk, Harun Alp, and Michael Peters. 2021. "Lack of selection and limits to delegation: firm dynamics in developing countries." *American Economic Review* 111 (1):231–75.

Akcigit, Ufuk, Murat Alp Celik, and Jeremy Greenwood. 2016. "Buy, keep, or sell: Economic growth and the market for ideas." *Econometrica* 84 (3):943–984.

Autor, David and David Dorn. 2009. "This job is "getting old": measuring changes in job opportunities using occupational age structure." *American Economic Review* 99 (2):45–51.

Baghai, Ramin, Rui Silva, Viktor Thell, and Vikrant Vig. 2020a. "Talent in distressed firms: Investigating the labor costs of financial distress." *Journal of Finance* .

———. 2020b. "Talent in distressed firms: Investigating the labor costs of financial distress." *Journal of Finance* .

Becker, Gary S. 1962. "Investment in human capital: A theoretical analysis." *Journal of political economy* 70 (5, Part 2):9–49.

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

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

Blatter, Marc, Samuel Muehlemann, and Samuel Schenker. 2012. "The costs of hiring skilled workers." *European Economic Review* 56 (1):20–35.

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." .

Campa, Jose Manuel and Simi Kedia. 2002. "Explaining the diversification discount." *Journal of Finance* 57 (4):1731–1762.

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. 2020. "Human capital-driven acquisition: Evidence from the inevitable disclosure doctrine." *Management Science* .

Cunha, Flavio and James Heckman. 2007. "The technology of skill formation." *American Economic Review* 97 (2):31–47.

Custódio, Cláudia, Miguel A Ferreira, and Pedro Matos. 2019. "Do general managerial skills spur innovation?" *Management Science* 65 (2):459–476.

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

Donangelo, Andres. 2014. "Labor mobility: Implications for asset pricing." *Journal of Finance* 69 (3):1321–1346.

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.

Eiling, Esther. 2013. "Industry-specific human capital, idiosyncratic risk, and the cross-section of expected stock returns." *Journal of Finance* 68 (1):43–84.

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

Elsas, Ralf, Mark J Flannery, and Jon A Garfinkel. 2014. "Financing major investments: Information about capital structure decisions." *Review of Finance* 18 (4):1341–1386.

Faccio, Mara and William O'Brien. 2018. "Business groups and employment." *Working Paper* .

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

Frésard, Laurent, Ulrich Hege, and Gordon Phillips. 2017. "Extending industry specialization through cross-border acquisitions." *The Review of Financial Studies* 30 (5):1539–1582.

Fulghieri, Paolo and Merih Sevilir. 2020. "Human capital integration in mergers and acquisitions." *Working Paper* .

Ghaly, Mohamed, Viet Anh Dang, and Konstantinos Stathopoulos. 2017. "Cash holdings and labor heterogeneity: the role of skilled labor." *Review of Financial Studies* 30 (10):3636–3668.

Gibbons, Robert and Michael Waldman. 2004. "Task-specific human capital." *American Economic Review* 94 (2):203–207.

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

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

Gompers, Paul, Anna Kovner, Josh Lerner, and David Scharfstein. 2010. "Performance persistence in entrepreneurship." *Journal of financial economics* 96 (1):18–32.

He, Alex Xi and Daniel le Maire. 2020. "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: Journal of the econometric society* :153–161.

Henderson, Rebecca and Iain Cockburn. 1996. "Scale, scope, and spillovers: the determinants of research productivity in drug discovery." *The Rand Journal of Economics* :32–59.

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.

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

———. 2018. "Conglomerate industry choice and product language." *Management Science* 64 (8):3735–3755.

Hoberg, Gerard and Gordon M Phillips. 2020. "Scope, Scale and Competition: The 21st Century Firm." *Available at SSRN* .

Hombert, Johan and Adrien Matray. 2017. "The real effects of lending relationships on innovative firms and inventor mobility." *The Review of Financial Studies* 30 (7):2413–2445.

Israelsen, Ryan D and Scott E Yonker. 2017. "Key human capital." *Journal of Financial and Quantitative Analysis* 52 (1):175–214.

Jacobson, Louis S, Robert J LaLonde, and Daniel G Sullivan. 1993. "Earnings losses of displaced workers." *The American economic review* :685–709.

Kirzner, Israel M et al. 1978. "Competition and Entrepreneurship." *University of Chicago Press Economics Books* .

Lagaras, Spyridon. 2019. "Corporate takeovers and labor restructuring." *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.

Lustig, Hanno, Chad Syverson, and Stijn Van Nieuwerburgh. 2011. "Technological change and the growing inequality in managerial compensation." *Journal of Financial Economics* 99 (3):601–627.

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

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

Maksimovic, Vojislav, Gordon Phillips, and Liu Yang. 2013. "Private and public merger waves." *Journal of Finance* 68 (5):2177–2217.

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

Matsa, David A. 2010. "Capital structure as a strategic variable: evidence from collective bargaining." *Journal of Finance* 65 (3):1197–1232.

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

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

Merz, Monika and Eran Yashiv. 2007. "Labor and the market value of the firm." *American Economic Review* 97 (4):1419–1431.

Myers, Stewart C. 1977. "Determinants of corporate borrowing." *Journal of financial economics* 5 (2):147–175.

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. 2016. "Acquiring labor." *Working Paper* .

Oyer, Paul. 2004. "Why do firms use incentives that have no incentive effects?" *The Journal of Finance* 59 (4):1619–1650.

Panzar, John C and Robert D Willig. 1981. "Economies of scope." *The American Economic Review* 71 (2):268–272.

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.

Prendergast, Canice. 1993. "The role of promotion in inducing specific human capital acquisition." *The Quarterly Journal of Economics* 108 (2):523–534.

Rajan, Raghuram G and Luigi Zingales. 2001. "The firm as a dedicated hierarchy: A theory of the origins and growth of firms." *The Quarterly Journal of Economics* 116 (3):805–851.

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." *The Journal of Finance* 57 (6):2379–2403.

Serfling, Matthew. 2016. "Firing costs and capital structure decisions." *Journal of Finance* 71 (5):2239–2286.

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

Simintzi, Elena, Vikrant Vig, and Paolo Volpin. 2014. "Labor protection and leverage." *Review of Financial Studies* 28 (2):561–591.

Tate, Geoffrey and Liu Yang. 2016a. "The bright side of corporate diversification: evidence from internal labor markets." *Review of Financial Studies* 28 (8):2203–2249.

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

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

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. *Sources:* SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown. *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 classification 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. Descriptive Statistics

This table reports descriptive statistics (number of observations, average, standard deviation, 5th, 25th, 50th, 75th, and 95th percentiles) for the variables used in the analysis. The variables are defined in Appendix Table A. *Sources:* SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, 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.

	#	Mean	SD	P5	P25	P50	P75	P95
Dependent variables:								
$\mathbb{1}(\text{Buy})_{g,n,t}$	34,045	0.02	0.14	0.00	0.00	0.00	0.00	0.00
$\mathbb{1}(\text{Diversification})_{g,n,t\text{-}1}$	86,661,511	0.00	0.04	0.00	0.00	0.00	0.00	0.00
$\mathbb{1}(\text{Diversifying Subsidiary})_{f,n,t}$	32,967	0.29	0.45	0.00	0.00	0.00	1.00	1.00
Human Capital Distances:								
HC $\text{Distance}_{g,n,t\text{-}1}$	34,045	0.76	0.22	0.27	0.66	0.84	0.93	0.99
HC $\text{Distance}_{g,n,t\text{-}1}$ (wages)	34,045	0.76	0.22	0.29	0.65	0.83	0.93	0.99
HC $\text{Distance}_{g,n,t\text{-}1}$ (hours)	34,045	0.76	0.22	0.27	0.66	0.84	0.93	0.99
HC $\text{Distance}_{g,n,t0}$	31,541	0.80	0.21	0.33	0.72	0.87	0.96	1.00
HC $\text{Distance}_{g,n,t\text{-}1}$ (simple average)	33,972	0.86	0.14	0.55	0.82	0.90	0.95	0.98
HC $\text{Distance}_{g,n,t\text{-}1}$ (weighted: employment)	33,972	0.87	0.14	0.53	0.83	0.91	0.96	0.99
HC $\text{Distance}_{g,n,t\text{-}1}$ (weighted: sales)	33,972	0.87	0.14	0.55	0.83	0.91	0.96	0.99
HC $\text{Distance}_{o,n,t\text{-}1}$ (sector-level)	23,452	0.77	0.21	0.28	0.68	0.83	0.92	0.98
Other Firm-level Independent Variables:								
$\text{Log}(\#\text{ Employees})_{g,t\text{-}1}$	34,045	3.60	1.23	1.39	2.94	3.58	4.26	5.66
$\#\text{ Occupations}/\#\text{ Employees}_{g,t\text{-}1}$	34,045	0.40	0.27	0.08	0.21	0.33	0.53	1.00
$\text{Value Added}/\#\text{ Employees}_{g,t\text{-}1}$	34,045	0.06	0.05	0.02	0.04	0.05	0.07	0.14
$\text{Tangible Assets}/\#\text{ Employees}_{g,t\text{-}1}$	34,045	0.06	0.16	0.00	0.01	0.02	0.05	0.18
$\text{Total Wages}/\#\text{ Employees}_{g,t\text{-}1}$	34,045	0.03	0.01	0.02	0.02	0.03	0.03	0.05
$\text{Cash}/\#\text{ Employees}_{g,t\text{-}1}$	34,045	0.03	0.07	0.00	0.00	0.01	0.03	0.11
$\text{Sales Shift (100\%)}_{g,t\text{-}1,t}$	34,045	0.01	0.12	0.00	0.00	0.00	0.00	0.00
$\mathbb{1}(\text{Public})_{g,t\text{-}1}$	34,045	0.02	0.14	0.00	0.00	0.00	0.00	0.00
$\text{Serial Acquirer}_{g}$	34,045	0.01	0.10	0.00	0.00	0.00	0.00	0.00
$\mathbb{1}(\text{Diversified})_{g,t\text{-}1}$	34,045	0.47	0.50	0.00	0.00	0.00	1.00	1.00
$\text{Age}_{g,t\text{-}1}$	33,392	22.61	16.86	3.00	11.00	20.00	31.00	51.00
$\text{Log(1+Investment)}_{g,t}$	28,146	4.22	2.12	0.00	2.88	4.26	5.58	7.67
Post-entry Performance:								
$\text{Log(Employment)}_{f,t}$	238,926	3.45	1.05	1.61	3.00	3.50	3.99	5.10
$\Delta\ \text{Employment}_{f,t\text{-}1,t+3}$	16,338	-0.34	0.68	-1.00	-1.00	-0.18	0.04	0.43
$\Delta\ \text{Top 5 Occupations}_{f,t\text{-}1,t+3}$	16,338	-0.07	0.26	-0.57	-0.11	0.00	0.00	0.21
$\Delta\ \text{Other Occupations}_{f,t\text{-}1,t+3}$	16,338	-0.28	0.60	-1.00	-0.80	-0.17	0.04	0.42
$\Delta\ \text{HC Distance}_{f,t,t+3}$	16,338	-0.35	0.58	-1.00	-1.00	-0.05	0.00	0.16
$\text{Log(Sales)}_{g,n,t}$	73,028	6.92	1.40	4.50	6.10	7.07	7.96	8.88
$\text{Survival}_{g,n,t+3}$	16,338	0.21	0.41	0.00	0.00	0.00	0.00	1.00
Sector-level Dependent Variables:								
$\text{Log(1+Entries)}_{o,n,t}$	23,452	0.68	0.73	0.00	0.00	0.69	1.10	2.08
$\text{Log(1+Build)}_{o,n,t}$	23,452	0.66	0.73	0.00	0.00	0.69	1.10	2.08
$\text{Log(1+Buy)}_{o,n,t}$	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:								
$\text{LLM Tightness}_{z,n,t\text{-}1}$	8,396	0.20	0.18	0.01	0.06	0.14	0.29	0.58
$\text{Capital Intensity}_{n,t}$	34,045	109.72	324.53	15.83	28.25	39.49	74.81	321.27
$\mathbb{1}(\text{New Department})_{g,t}$	33,934	0.07	0.25	0.00	0.00	0.00	0.00	1.00
$\mathbb{1}(\text{New Region})_{g,t}$	33,934	0.04	0.19	0.00	0.00	0.00	0.00	0.00
$\text{Product Market Distance}_{o,n,t\text{-}1}$	34,045	0.04	0.20	0.00	0.00	0.00	0.00	0.01
$\mathbb{1}(\text{Same 1-digit Industry})_{o,n}$	34,045	0.35	0.48	0.00	0.00	0.00	1.00	1.00
$\mathbb{1}(\text{Same 2-digit Industry})_{o,n}$	34,045	0.18	0.39	0.00	0.00	0.00	0.00	1.00
$\mathbb{1}(\text{Upstream Link } >5\%)_{o,n}$	34,045	0.24	0.42	0.00	0.00	0.00	0.00	1.00
$\mathbb{1}(\text{Upstream Link } >10\%)_{o,n}$	34,045	0.16	0.36	0.00	0.00	0.00	0.00	1.00
$\mathbb{1}(\text{Downstream Link } >5\%)_{o,n}$	34,045	0.19	0.39	0.00	0.00	0.00	0.00	1.00
$\mathbb{1}(\text{Downstream Link } >10\%)_{o,n}$	34,045	0.13	0.34	0.00	0.00	0.00	0.00	1.00
$\#\text{ New Firms}_{n,t\text{-}1}$ (thousands)	34,045	2.97	4.02	0.08	0.59	1.30	3.07	13.05
$\text{Sector New Firm Survival}_{n,t\text{-}1}$	34,045	0.74	0.12	0.46	0.68	0.77	0.83	0.89

Continued next page

Summary statistics (continued)

	N	Mean	SD	P5	P25	P50	P75	P95
Sector Sales Growth$_{n,t-1}$	34,045	0.06	0.14	-0.11	-0.01	0.05	0.11	0.29
Herfindahl Index$_{n,t-1}$	34,045	0.02	0.04	0.00	0.00	0.01	0.02	0.06

Figure 2. Heatmap of Human Capital Distance between Sectors

The heatmap shows the pairwise HC Distance between sectors in 2013. The sector-level HC Distance measures the distance between the vectors containing the share of workers by occupation in the aggregate sectoral workforce (see equation (5)). The measure ranges from zero to one. It is equal to zero when the allocation of workers across occupations is similar in two sectors and close to one when there is no overlap in worker occupations 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.

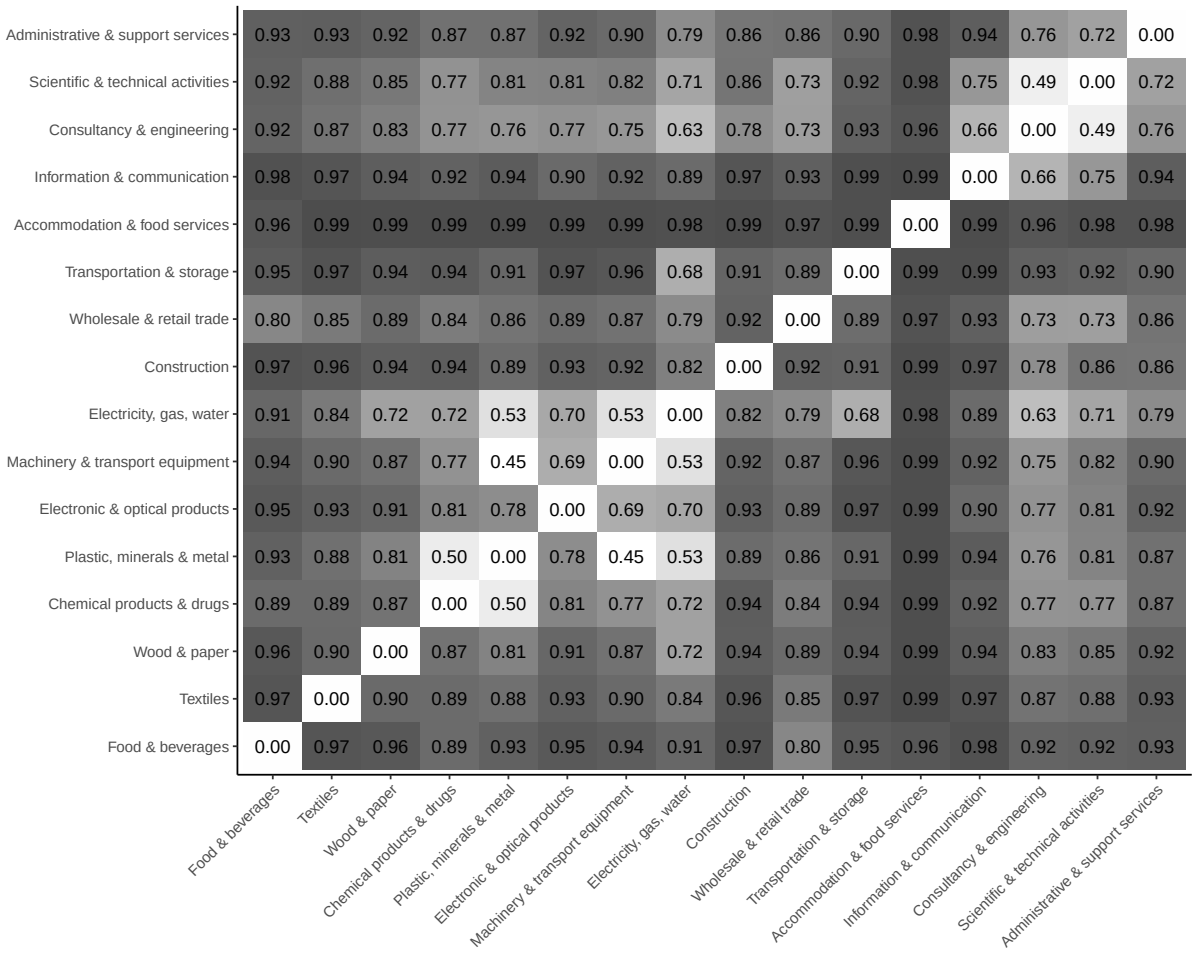


Table 3. Human Capital Distance and the Sector of Diversification

The table reports OLS estimates of the effect of HC Distance on the choice of the sector of diversification. The dependent variable $\mathbb{1}(\text{Diversification})_{g,n,t}$ 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 Distance$_{g,n,t-1}$, measures the distance of firm g's vector of occupations to sector of entry n's vector of occupations 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 $t-1$. The 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. *Sources:* SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, and matched employer-employee dataset. *Sample:* Firms that enter a new sector in the 2005-2014 period.

Dependent variable:	$\mathbb{1}(\text{Diversification})_{g,n,t}$				
	(1)	(2)	(3)	(4)	(5)
HC Distance$_{g,n,t-1}$	-0.014***	-0.014***	-0.008***	-0.008***	-0.009***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Controls	No	Yes	No	Yes	No
Sector of Origin-Entry-Year FE	No	No	Yes	Yes	Yes
Firm-Year FE	No	No	No	No	Yes
Adjusted R^2	0.002	0.002	0.033	0.033	0.035
Observations	86,661,511	86,661,511	86,661,511	86,661,511	86,661,511

Table 4. Human Capital and the Decision to Build or Buy

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, and matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 5% 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, $\mathbb{1}(\text{Buy})_{g,n,t}$, is a dummy variable that takes value one if firm g buys in sector of entry n in year t, 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$_{g,n,t-1}$, measures the distance of firm g's vector of occupations to sector of entry n's vector of occupations 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 various combinations of fixed effects. 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 $t-1$. Sector of origin (second) denotes the second sector (in terms of sales) in which the firm realizes most of its sales in $t-1$. The sector of entry n denotes a new sector in which the firm g starts reporting sales. t is the year of the 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:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
	(1)	(2)	(3)	(4)	(5)
HC Distance$_{g,n,t-1}$	0.027**	0.034***	0.029***	0.024**	0.010**
	(0.012)	(0.011)	(0.009)	(0.010)	(0.005)
N. Occupations/N. Employees$_{g,t-1}$		0.018**	0.018**	0.012***	-0.030**
		(0.009)	(0.008)	(0.005)	(0.014)
log(N. Employees)$_{g,t-1}$		0.008***	0.010***	0.005**	-0.049***
		(0.003)	(0.003)	(0.002)	(0.011)
Tangible Assets/N. Employees$_{g,t-1}$		-0.031**	-0.036***	-0.002	-0.037
		(0.014)	(0.013)	(0.006)	(0.027)
Cash/N. Employees$_{g,t-1}$		0.087**	0.079**	0.052	0.019
		(0.034)	(0.037)	(0.032)	(0.054)
Value added/N. Employees$_{g,t-1}$		0.003	0.014	-0.024	-0.200***
		(0.044)	(0.038)	(0.031)	(0.065)
Total wages/N. Employees$_{g,t-1}$		1.123**	1.066**	0.731**	0.298
		(0.495)	(0.413)	(0.294)	(0.462)
Controls	No	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Year FE	Yes	Yes	No	No	No
Sector of Origin × Size × Entry FE	No	No	Yes	No	No
Sector of Origin (main) × Origin (second) × Entry FE	No	No	No	Yes	No
Sector of Origin × Entry FE	No	No	No	No	Yes
Firm FE	No	No	No	No	Yes
Year FE	No	No	Yes	Yes	Yes
Adjusted R^2	0.208	0.217	0.170	0.264	0.725
Observations	34,045	34,045	32,435	27,152	19,472

Table 5. Investigating the Role of Self- Selection

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, and matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 5% 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, $\mathbb{1}(\text{Buy})_{g,n,t}$, takes value one if firm g buys in sector of entry n in year t, zero if it builds. The main independent variable, HC Distance$_{g,n,t-1}$, measures the distance of firm g's vector of occupations to sector of entry n's vector of occupations in year $t-1$. In columns (1) to (3), we estimate our main model on subsamples excluding firms that shift a substantial part of their activity. Shifting firms decrease their activity in one of their existing sectors of activity when they enter sector of entry n. For each firm in the sample, we compute the growth rate of sales between $t-1$ and t in each sector where the firm operates in $t-1$. We take the firm-level 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 the time period. In column (5), we interact the number of sectors of origin o where the firm is present in $t-1$ with sector of origin-entry fixed effects, we add year fixed effects, and we estimate our main model. In column (6), we interact the firm's age deciles with sector of origin-entry fixed effects, we add 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:	$\mathbb{1}(\text{Buy})_{g,n,t}$					
	% Sales shift$_{g,n,t,t-1}$			Excluding	Life Cycle	
	100%	50%	25%	Serial Acquirers	# Sectors$_{g,t-1}$	Age$_{g,t-1}$ decile
	(1)	(2)	(3)	(4)	(5)	(6)
HC Distance$_{g,n,t-1}$	0.032***	0.040***	0.050***	0.016**	0.036***	0.029**
	(0.012)	(0.013)	(0.017)	(0.007)	(0.008)	(0.009)
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

Table 6. Exploring Endogeneity: Sector-level Analysis, Sticky Human Capital and Entry

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, and matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 5% of their sales in the new sector in the 2005-2014 period. The table reports OLS estimates and first analyzes the effect of two plausibly more exogenous human capital variables on the decision to diversify in a given sector and second examines the decision to build or buy. In panel A, we test the relationship between HC Distance and corporate diversification strategies 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 coming from firms in sector of origin o 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 in year t. Since we study the intensive margin, the regression in column (4) only retains observation. The main independent variable, HC Distance$_{o,n,t-1}$, measures the distance between firms in sector of origin o's vector of occupations and sector of entry n's vector of occupations in year $t-1$. Control variables include the product market distance, vertical links upstream and downstream, and a 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 in all models in panel A. In panel B, we examine the role of past workforce composition and sales composition in firms' diversification strategies. The dependent variable in columns (1) and (3) is $\mathbb{1}(\text{Diversification})_{g,n,t}$ is a dummy variable that takes value one if firm g enters sector n in year t, zero otherwise. In columns (2) and (4), the dependent variable, $\mathbb{1}(\text{Buy})_{g,n,t}$, takes value one if firm g buys in sector of entry n in year t, zero if it builds. The main independent variable, HC Distance$_{g,n,t_0}$, measures the distance of firm g's vector of occupations to sector of entry n's vector of occupations in year t_0, which is the first year that firm g appears in our sample. 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) and (2) include sector of origin o in year $t-1$-sector of entry n-year $t-1$ fixed effects. The models in columns (3) and (4) include sector of origin o in year t_0-sector of entry n-year $t-1$ 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.

Panel A. Sector-level Human Capital and Entry

Dependent variable:	Log(1+Entries)$_{o,n,t}$ (1)	Log(1+Build)$_{o,n,t}$ (2)	Log(1+Buy)$_{o,n,t}$ (3)	Buy frequency$_{o,n,t}$ (4)
HC Distance$_{o,n,t-1}$	-1.061***	-1.050***	-0.071***	0.438***
	(0.105)	(0.106)	(0.018)	(0.140)
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	0.298	0.292	0.116	0.330
Observations	23,452	23,452	23,452	438

Panel B. Sticky Human Capital and Entry

Dependent variable:	$\mathbb{1}(\text{Diversify})_{g,n,t}$ (1)	$\mathbb{1}(\text{Buy})_{g,n,t}$ (2)	$\mathbb{1}(\text{Diversify})_{g,n,t}$ (3)	$\mathbb{1}(\text{Buy})_{g,n,t}$ (4)
HC Distance$_{g,n,t0}$	-0.013***	0.015***	-0.009***	0.005
	(0.002)	(0.005)	(0.001)	(0.004)
Controls	No	Yes	No	Yes
Sector of Origin $(t-1)\times$ Entry × Year FE	Yes	Yes	No	No
Sector of Origin $(t_0)\times$ Entry × Year FE	No	No	Yes	Yes
Firm × Year FE	Yes	No	Yes	No
Adjusted R^2	0.034	0.222	0.033	0.133
Observations	78,587,798	31,129	73,816,135	27,544

Figure 3. Local Labor Market Tightness in France

Source: French unemployment agency (Pole Emploi). This figure plots the distribution of local labor market (LLM) tightness in 2013. LLM Tightness$_{z,n,t}$ measures labor scarcity by LLM and corresponds to the number of occupations in short supply in a given LLM. Occupations in short supply 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 LLMs. Darker shades of blue indicate a higher degree of LLM Tightness$_{z,n,t}$.

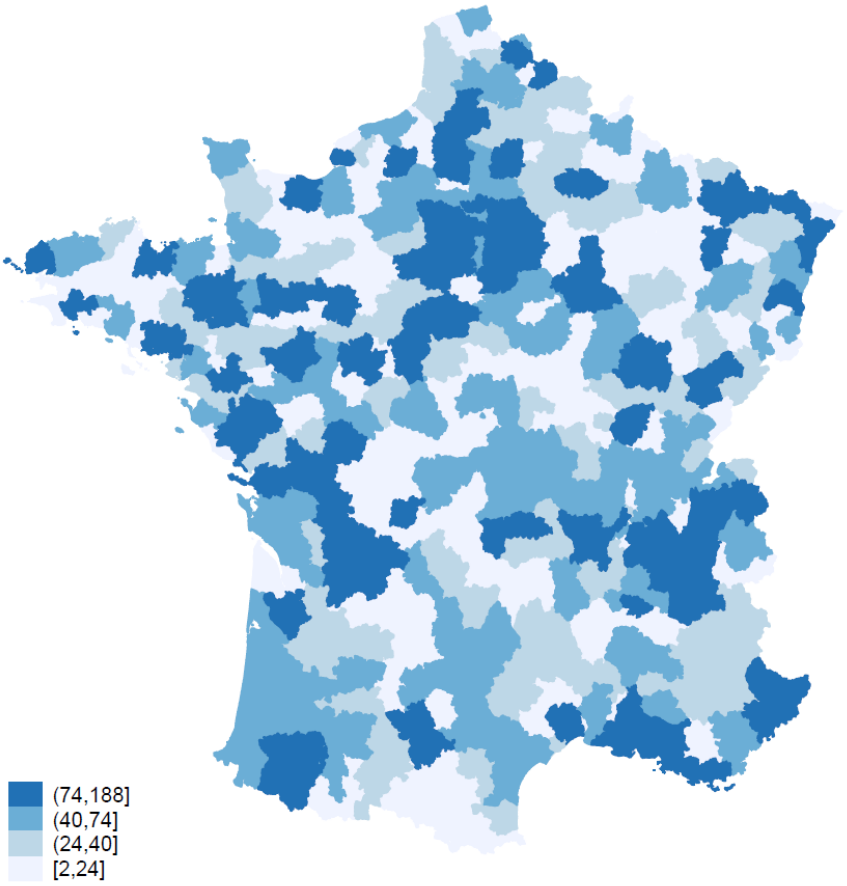
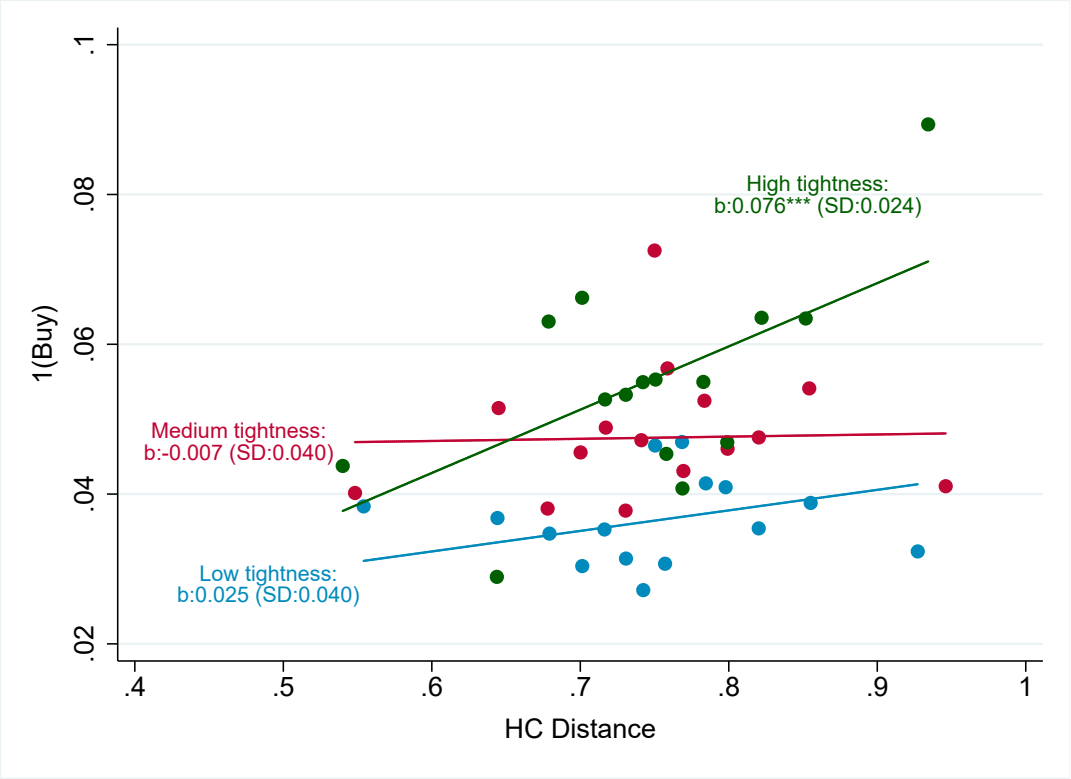


Table 7. Build or Buy and Local Labor Market Tightness

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, matched employer-employee dataset, unemployment agency dataset. *Sample:* Firms that enter a new sector and realize at least 5% of their sales in the new sector in the 2010-2014 period. The table reports OLS estimates and tests the effect of human capital interacted with local labor market (LLM) tightness 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, 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 vector of occupations to sector of entry n's vector of occupations 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 LLM and corresponds to the number of occupations in short supply in a given LLM. Occupations in short supply 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 LLMs. 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 LLM fixed effects. Standard errors are 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.

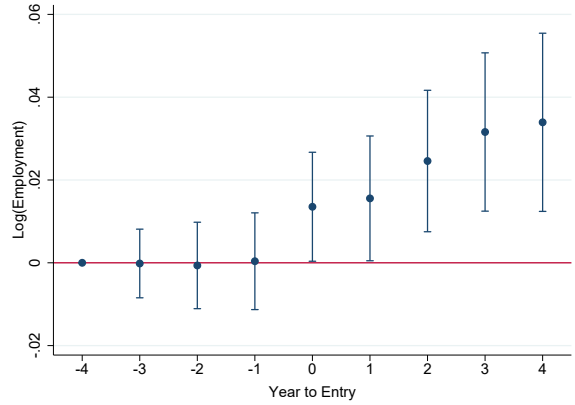


Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$		
Subsamples by LLM Tightness:	Tercile 1	Tercile 2	Tercile 3
	(1)	(2)	(3)
HC Distance$_{g,n,t-1}$	0.025	-0.007	0.076***
	(0.040)	(0.041)	(0.025)
Controls	Yes	Yes	Yes
Origin × Entry × Year FE	Yes	Yes	Yes
Local labor market FE	Yes	Yes	Yes
Adjusted R^2	0.481	0.523	0.367
Observations	1,667	1,516	1,604

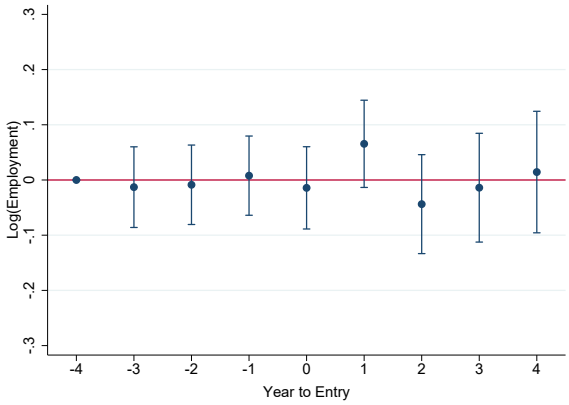
Table 8. Human Capital and Post-entry Labor Adjustments

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 5% 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 a subsidiary fixed effect, a sector of origin × entry × year of entry fixed effect, a time to entry fixed effect and the high distance dummy coefficient interacted with time to entry dummies (plotted on the graph; the first year is set as the baseline). We proceed with separate estimations for build and buy entries. 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). Only firms that survive three years after entry are retained. The dependent variable in column (1) is the growth rate of the number of workers between $t-1$ and $t+3$ employed by 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 $\Delta\text{Employment}_f = \Delta\text{Top5 occupations}_f + \Delta\text{Other occupations}_f$, meaning that the coefficients in columns (2) and (3) sum to the coefficient of column (1). The dependent variable in column (4) is the variation in the subsidiary's HC Distance between t and $t+3$. The main independent variable, $\text{HC Distance}_{g,n,t-1}$, measures the distance of firm g's vector of occupations to sector of entry n's vector of occupations 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 and local labor market 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)
$\text{HC Distance}_{g,n,t\text{-}1}$	0.071**	0.567***	-0.496***	-0.396***
	(0.032)	(0.034)	(0.033)	(0.051)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.525	0.418	0.502	0.553
Observations	17,752	17,752	17,752	16,882

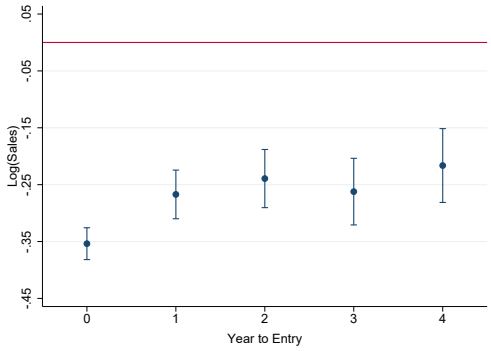
Panel B: Buy entries

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)
$\text{HC Distance}_{g,n,t\text{-}1}$	-0.518*	0.256*	-0.774**	-0.540**
	(0.303)	(0.131)	(0.288)	(0.206)
Controls	Yes	Yes	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.445	0.306	0.350	0.660
Observations	233	233	233	240

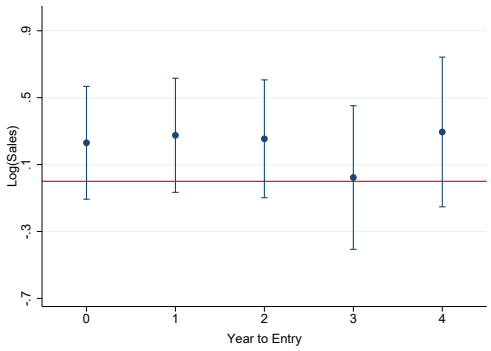
Table 9. Human Capital and Post-entry Performance

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, matched employer-employee dataset. *Sample:* Firms that enter a new sector and realize at least 5% 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 shows 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 in the 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 vector of occupations to sector of entry n's vector of occupations 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 and local labor market 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$_{g,n,t-1}$	-0.721***	-0.196***
	(0.113)	(0.022)
Controls	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes
Adjusted R^2	0.566	0.185
Observations	18,928	18,928

Panel B: Buy entries

Dependent variable:	$\text{Log(Sales)}_{g,n}$	$\text{Survival}_{g,n}$
Time interval:	t	$t+3$
	(1)	(2)
HC Distance$_{g,n,t-1}$	0.369	-0.260
	(0.737)	(0.191)
Controls	Yes	Yes
Sector of Origin × Entry × Entry Year FE	Yes	Yes
Adjusted R^2	0.410	0.230
Observations	129	294

Table 10. Build, Buy, and Firms' Path of Entry

Sources: SDC Platinum, BvD Zephyr, tax files, ownership links dataset, survey of sales breakdown, 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 $\mathbb{1}(\text{Diversification})_{g,n,t}$ 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 Distance$_{g,n,t0}$, measures the distance of firm g's vector of occupations to sector of entry n's vector of occupations in year t_0, which is the first year that firm g appears in our sample. In the first two columns, we include all firms and compare firms' diversification depending on whether they built or bought the first time they entered a new sector in our sample (denoted by t_1). $\mathbb{1}(\text{First entry by buying})_{g,t,1}$ is a dummy equal to one if firm g buys at time t_1. $\mathbb{1}(t > t_1)$ is a dummy equal to one after time t_1. In the last two columns, we only include firms that buy and compare within-firm diversifications before and after the first buy entry. The time of the first buy entry is denoted by t_{Buy}, and $\mathbb{1}(t > t_{Buy})$ 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:	$\mathbb{1}(\text{Diversify})_{g,n,t}$			
Estimation:	Across		Within	
Sample:	All firms		Firms that buy	
	(1)	(2)	(3)	(4)
HC Distance$_{g,n,t0}$	-0.011***	-0.011***	-0.019***	-0.049***
	(0.002)	(0.002)	(0.004)	(0.007)
$1(t > t_1)\times$ 1(First entry by buying)$_g$		-0.006**		
		(0.003)		
1(First entry by buying)$_g$ × HC Distance$_{g,n,t0}$		0.001***		
		(0.001)		
$1(t > t_1)\times$ 1(First entry by buying)$_g$ × HC Distance$_{g,n,t0}$		0.006*		
		(0.003)		
$1(t > t_{Buy})\times$ HC Distance$_{g,n,t0}$				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

Appendix A Description of Variables

Variables	Description
Main Dependent Variables:	
$\mathbb{1}(\text{Diversify})_{g,n,t}$	Dummy variable that takes value one if firm g enters sector n in year t, zero otherwise. *Sources:* SDC Platinum, Zephyr, Ownership links dataset, Sales breakdown survey.
$\mathbb{1}(\text{Buy})_{g,n,t}$	dummy variable that takes value one if firm g buys in sector of entry n in year t, zero if it builds. *Sources:* SDC Platinum, Zephyr, Ownership links dataset, Sales breakdown survey.
$\mathbb{1}(\text{Diversifying Subsidiary})_{f,n,t}$	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 survey.
Human Capital Distances:	
HC Distance$_{g,n,t}$	Human capital distance is defined by equation (3): $$\text{HC Distance}_{g,n,t} = 1 - \frac{\sum_i s_{g,i,t} \cdot s_{n,i,t}}{\sqrt{\sum_i s_{g,i,t}^2}\sqrt{\sum_i s_{n,i,t}^2}},$$ where $s_{g,i,t}$ is the share of employees in firm g employed in occupation i and $s_{n,i,t}$ is the share of employees in sector n employed in occupation i. *Source:* Matched employer-employee dataset, Ownership links dataset.
HC Distance$_{g,n,t}$ (wages)	Human capital distance is defined by (3), 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 sector n's total wage bill that goes to employees in occupation i. *Source:* Matched employer-employee dataset.
HC Distance$_{g,n,t}$ (hours)	Human capital distance is defined by (3), 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 in sector n by employees in occupation i. *Source:* Matched employer-employee dataset.
HC Distance$_{g,n,t_0}$	HC Distance of firm g to sector n (in terms of number of workers) the first year firm g appears in the sample. *Source:* Matched employer-employee dataset.
HC Distance$_{g,n,t}$ (simple average)	Simple average of the HC Distance of firm g to firms operating in sector n (in terms of the number of workers). *Source:* Matched employer-employee dataset.
HC Distance$_{g,n,t}$ (weighted average: employment)	Average of the HC Distance of firm g to firms operating in sector n (in terms of the number of workers) weighted by the number of workers of incumbent firms. *Source:* Matched employer-employee dataset.
HC Distance$_{g,n,t}$ (weighted average: sales)	Average of the HC Distance of firm g to firms operating in sector n (in terms of the number of workers) weighted by sales of incumbent firms. *Source:* Matched employer-employee dataset.
HC Distance$_{o,n,t}$	Sector-level human capital distance between sector of origin o and sector of entry n given by equation (5): $$\text{HC Distance}_{o,n,t} = 1 - \frac{\sum_i s_{o,i,t} \cdot s_{n,i,t}}{\sqrt{\sum_i s_{o,i,t}^2}\sqrt{\sum_i s_{n,i,t}^2}}. \quad (5)$$ where $s_{o,i,t}$ is the share of workers employed in occupation i in sector of origin o at time t and $s_{n,i,t}$ is the share of workers employed in occupation i in sector of entry n at time t. *Source:* Matched employer-employee dataset, Ownership links dataset.
HC Distance$_{f,n,t}$	Human capital distance at the subsidiary f level adapted from (3), where $s_{f,i,t}$ is the share of employees in subsidiary f employed in occupation i, and $s_{n,i,t}$ is the share of employees in sector n employed in occupation i. *Source:* Matched employer-employee dataset, Ownership link dataset.

Continued next page

Description of Variables (continued)

Variables	Description
Other Firm-level Independent Variables:	
$\log(\#\text{ Employees})_{g,t}$	Logarithm of the number of employees in firm g in year t. *Source:* Matched employer-employee dataset, Ownership links dataset.
$\#\text{ Occupations}/\#\text{ Employees}_{g,t}$	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, Ownership links dataset.
$\text{Value added}/\#\text{ Employees}_{g,t}$	Total value added generated by firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Ownership links dataset, Matched employer-employee dataset.
$\text{Tangible Assets}/\#\text{ Employees}_{g,t}$	Total value of tangible assets held by firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Ownership links dataset, Matched employer-employee dataset.
$\text{Total wages}/\#\text{ Employees}_{g,t}$	Total wages of firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Ownership links dataset, Matched employer-employee dataset.
$\text{Cash}/\#\text{ Employees}_{g,t}$	Total cash holdings of firm g in year t, scaled by the number of employees in the firm. *Sources:* Tax files, Ownership links dataset, Matched employer-employee dataset.
$\text{Entry threshold}_{g,n,t}$	Ratio of firm's realized sales in sector of entry n in year t divided by the firm's total sales realized in sector of origin at $t-1$. We consider four entry thresholds: none, 1%, 5% (baseline specification), and 10%. *Sources:* Sales breakdown survey, Ownership links dataset.
$\text{Sales shift}_{g,t\text{-}1,t}$	A shifting firm is a firm that decreases its activity in sector of origin o when it enters sector of entry n. Shifting firms are identified by computing the growth rate of sales between $t-1$ and t in 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 survey, Ownership links dataset.
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 Amadeus.
Serial acquirer_g	A serial acquirer is a firm that enters at least a new sector by acquisition at least twice during the time period. *Sources:* SDC Platinum, BvD Zephyr, Ownership links dataset, Sales breakdown survey.
$\mathbb{1}(\text{Diversified})_{g,t}$	Dummy variable equal to one if firm g is diversified in year t, i.e., the firm operates in more than one sector. A sector is defined at the 3-digit level. *Sources:* Ownership links dataset, Sales breakdown survey.
$\text{Age}_{g,t}$	Age is defined as the maximum of age of subsidiaries composing the firm. *Sources:* Ownership links dataset, Tax files.
$\text{Log(1+Investment)}_{g,t}$	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:* Ownership links dataset, Tax files.
Post-entry Performance:	
$\text{Log(Employment}_{f,t})$	Logarithm of the number of employees employed by the subsidiary f associated with the entry of firm g in sector of entry n. *Sources:* Tax files, Sales breakdown survey, Ownership links, Matched employer-employee dataset.
$\Delta\text{ Employment}_{f,t-1,t+3}$	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:* Tax files, Sales breakdown survey, Ownership links, Matched employer-employee dataset.

Continued next page

Description of Variables (continued)

Variables	Description
Δ Top 5 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 subsidiary f associated with the entry of firm g in sector of entry n. It is defined as: $\Delta\text{Top 5 Occupations}_{f,t-1,t+3} = \frac{\text{Top 5 Occupations}_{f,t+3} - \text{Top 5 Occupations}_{f,t-1}}{\text{Employment}_{f,t-1}}$ where Top 5 Occupations$_{f,t}$ is the number of workers employed by 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:* Tax files, Sales breakdown survey, Ownership links, Matched employer-employee dataset.
Δ Other Occupations$_{f,t-1,t+3}$	Defined as $\Delta\text{Other Occupations}_{f,t-1,t+3} = \frac{\text{Other Occupations}_{f,t+3} - \text{Other Occupations}_{f,t-1}}{\text{Employment}_{f,t-1}}$ where Other occupation$_{f,t+3}$ is the number of workers employed by 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:* Tax files, Sales breakdown survey, Ownership links, Matched employer-employee dataset.
Δ HC Distance$_{f,t,t+3}$	Growth in the HC Distance of subsidiary f associated with the entry of firm g in sector of entry n between years t and $t+3$. *Sources:* Tax files, Sales breakdown survey, Ownership links, Matched employer-employee dataset.
Log(Sales$_{g,n,t}$)	Logarithm of sales realized by firm g at time t in sector n. *Sources:* Tax files, Sales breakdown survey, Ownership links dataset.
Survival$_{g,n,t+3}$	Dummy variable that takes value one if the firm g survives at least 3 years in sector n that it enters at time t, zero otherwise. *Sources:* Tax files, Sales breakdown survey, Ownership links dataset.
Path-dependence Analysis:	
$\mathbb{1}(t > t_1)$	Dummy variable that takes value one if the next entry takes place after the first entry at t_1, zero if before. *Sources:* Tax files, Sales breakdown survey, Ownership links dataset.
$\mathbb{1}(t > t_{Buy})$	Dummy variable that takes value one if the next entry takes place after a buy entry at t_{Buy}, zero if before. *Sources:* Tax files, Sales breakdown survey, Ownership links dataset.
$\mathbb{1}$(First entry by buying)$_{g,t_1}$	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 survey, Ownership links dataset.
Sector-level Dependent Variables:	
log(1+Entries)$_{o,n,t}$	Logarithm of the number of entries in sector n at time t coming from firms operating in sector o.
log(1+Build)$_{o,n,t}$	Logarithm of the number of build entries in sector n at time t coming from firms operating in sector o.
log(1+Buy)$_{o,n,t}$	Logarithm of the number of buy entries in sector n at time t coming from firms operating in sector o.
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.
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, Ownership links dataset.

Continued next page

Description of Variables (continued)

Variables	Description
$\mathbb{1}(\text{New department})_{g,z,t}$	Dummy variable that indicates whether firm g enters a new geographical area z in year t. We define the new geographical area z at the department level. France is divided into 101 departments. A firm enters a new geographic zone if (i) it opens a new subsidiary f located in the new department in year t, (ii) if it opens a new plant in the department in year t, or (iii) if it acquires an existing firm in year t that operates in a department where the firm was not present in $t-1$. *Sources:* Ownership links dataset, Firm registry.
$\mathbb{1}(\text{New region})_{g,z,t}$	Dummy variable that indicates whether firm g enters a new geographical area z in year t. We define the new geographical area z at the region level. France is divided into 25 regions over the sample period. A firm enters a new geographic zone if (i) it opens a new subsidiary f located in the new region in year t, (ii) if it opens a new plant in the region in year t, or (iii) if it acquires an existing firm in year t that operates in a region where the firm was not present in $t-1$. *Sources:* Ownership links dataset, Firm registry.
$\text{LLM Tightness}_{z,n,t}$	Local labor market tightness is defined as: $$\text{LLM Tightness}_{z,n} = \sum_i \mathbb{1}(\text{Occupation in short supply})_{i,z} \cdot s_{n,i}, where \mathbb{1}(\text{Occupation in short supply})_{i,z}$$ is a dummy variable that is equal to one if occupation i is in short supply in z and $s_{n,i}$ the share of employees in sector n employed in occupation i. A high value of $\text{LLM Tightness}_{z,n}$ 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 short supply" 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:* Tax files, Sales breakdown survey, Ownership links dataset, French Unemployment Agency dataset.
$\text{Product market distance}_{o,n,t}$	The product market distance between sectors o and n measures whether firms' portfolios of sales in the sector of origin and in the sector of entry are similar. The product market distance is adapted from Bloom, Schankerman, and Van Reenen (2013). It is defined as: $$\text{Product Market Distance}_{o,n,t} = 1 - \frac{\sum_p v_{o,p,t} \cdot v_{n,p,t}}{\sqrt{\sum_p v_{o,p,t}^2}\sqrt{\sum_p v_{n,p,t}^2}}$$ where $v_{o,q,t}$ is the share of sales firms operating mainly in the sector of origin o realize in sector q at time t and $v_{n,q}$ 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:* Ownership links dataset, Tax files, Sales breakdown survey.
Industry classification (1-digit)	Dummy variable equal to one of sector of entry n and sector of origin o, defined at the French 3-digit SIC level, belong to the same French 1-digit SIC sector. *Sources:* Ownership links dataset, Tax files, Sales breakdown survey.
Sector classification (2-digits)	Dummy variable equal to one of sector of entry n and sector of origin o, defined at the French 3-digit SIC level, belong to the same French 2-digit SIC sector. *Sources:* Ownership links dataset, Tax files, Sales breakdown survey.

Continued next page

Description of Variables (continued)

Variables	Description
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:* Ownership links dataset, tax files, 2017 French Input-Output matrix.
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. 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:* Ownership links dataset, tax files, 2017 French Input-Output matrix.
Downstream 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) realize more than 5% of their sales in 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:* Ownership links dataset, tax files, 2017 French Input-Output matrix.
Downstream 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) realize more than 10% of their sales in 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:* Ownership links dataset, tax files, 2017 French Input-Output matrix.
# New firms$_{n,t}$	Number of firms newly created in the sector of entry in year t. *Sources:* SIRENE firm registry
New firm survival$_{n,t}$	Probability that firms in sector of entry n survive at least 3 years after creation. *Sources:* SIRENE 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