Build or Buy?
Human Capital and Corporate Diversification

Paul Beaumont Camille Hebert Victor Lyonnet*

19th November, 2019

Abstract

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? Using a novel dataset constructed by merging French employer payrolls with commercial M&A datasets, we show that firms are more likely to buy when their existing workforce does not include skills needed in the sector of entry. This relationship is more pronounced when labor market frictions make it difficult to hire key workers. Firms that enter by building realize lower entry sales when their existing workforce is not adapted to the sector of entry, especially in the presence of labor market frictions. Our results suggest that firms buy to acquire their target's human capital when adapting their existing workforce is too costly.

Keywords: M&As, diversification, human capital, labor market frictions.

JEL codes: L25, J24, J30, G34.

*Paul Beaumont is at Université Paris Dauphine and CREST (paul.beaumont@dauphine.eu). Camille Hebert is at the University of Toronto (camille.hebert@rotman.utoronto.ca). Victor Lyonnet is at the Ohio State University (lyonnet.1@osu.edu). We thank Manuel Adelino, Philippe Aghion, Fabio Braggion, Edouard Challe, Gilles Chemla, Espen Eckbo, Simon Gervais, Edith Ginglinger, John Graham, Denis Gromb, Johan Hombert, Julien Martin, Adrien Matray, Isabelle Méjean, Mario Milone, Tamara Nefedova, Paige Ouimet, Gordon Phillips, Manju Puri, Amiyatosh Purnanandam, Adriano Rampini, Luc Renneboog, David Robinson, René Stulz, Geoffrey Tate, Vish Viswanathan, Ramona Westermann and Sheng-Jun Xu for valuable comments and helpful discussions. We thank also seminar participants at CREST, Duke University, the 5th ECGC workshop on Governance and Control, HEC Paris, INSEE, the 28th Mitsui Labor and Finance Symposium, NFA, the Ohio State University, Université Paris-Dauphine, the UT Austin Corporate Finance Forum, and Tilburg University. This paper received the best student paper award at the 5th ECGC workshop on Governance and Control. 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

More than twelve thousand cross-industry merger and acquisition (M&A) deals occurred worldwide in 2017 and amounted to a total value of more than 900 billion dollars.[1] Cross-industry M&As allow firms to diversify, obtain specific assets and access new markets.[2] However, the magnitude of these figures can be misleading, concealing the fact that most diversifications result from firms entering new sectors directly by building on their existing capabilities. Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? This question has implications for understanding the role played by M&As in the reallocation of resources in the economy.

In this paper, we compare firms that enter a new sector by building on existing resources to firms that buy an existing company operating in the sector of entry. We focus on the role of human capital as a key resource to successfully operate in the new sector. Although our analysis applies to other assets, the availability of employee-level information and the existence of specific labor market frictions motivate our focus on human capital. We study the costs and benefits associated with each alternative. When a firm buys to enter a new sector, it has to incur both the costs of acquiring and restructuring the target, but the acquiring company also secures access to the target's productive resources. When a firm builds on its existing resources to enter a new sector, it faces the costs of hiring new workers to complement its existing workforce.[3] Therefore, the more key workers there are in a firm's existing workforce, the lower the adjustment costs to the new sector, and the more profitable it is for a firm to build.

We find that firms that have a workforce more adapted to the sector of entry are more likely to "build" than to "buy". The relation between internal human capital and the mode of entry in a new sector is stronger when specific skills are difficult to obtain on the external job market, i.e., when key worker occupations are in short supply.[4] Firms that build have lower

[1] Cross-sector M&As represent approximately 30% of the universe of M&A deals in both number and volume. See KPMG M&A Predictor 2018 Annual Report.

[2] The motives for horizontal M&As usually differ from those of vertical and conglomeral diversifications; they are often motivated by market power consolidation within a sector (e.g., Eckbo, 1983; Farrell and Shapiro, 1990).

[3] Maksimovic, Phillips and Prabhala (2011) shows that acquiring firms enter a costly restructuring process and sell or close 46% of the plants they buy. Hiring costs have been estimated to account for between one-quarter and one-half of wage payments and to increase in the specificity of skills (Abowd and Kramarz, 2003; Blatter, Muehlemann and Schenker, 2012).

[4] In our data, a worker occupation is "in short supply" when it meets two conditions: (i) job offers for this occupation exceed job applications and (ii) surveyed employers anticipate that they will not fill in a job in this occupation.

entry sales when their workforce is not adapted to the sector of entry, especially when key workers are in short supply. These findings are consistent with the idea that firms must search for key assets to expand into a new sector. When search costs increase, buying an existing firm becomes relatively more attractive than building from scratch and hiring new workers.[5]

One key challenge in testing this hypothesis is defining human capital and measuring it at the firm level. Human capital is neither directly observable nor easily defined. To overcome this challenge, we propose a measure based on a model of diversification with endogenous choice of teams. In the model, a firm hires workers with different occupations that relate to different sector-specific skills. If the firm chooses to build, it can draw workers from its existing pool of workers (internal labor market) or from the external job market. Instead, if the firm chooses to buy, it can select workers in the workforce of the target firm but has to pay fixed costs for the acquisition of another firm. The model yields two predictions. First, in equilibrium, the firm chooses to build when its existing *internal human capital* is adapted to operate in the sector of entry. The firm instead buys when its workforce does not include the key worker occupations needed to operate in the sector of entry and does so despite restructuring costs. Second, the relationship between human capital and the decision to build or buy is stronger when key workers are in short supply in the external job market.

The model yields a firm-level measure of human capital that stems from the relation between the fraction of the wage bill that goes to a given type of worker and the contribution of such workers to the firm's output. Empirically, we construct our measure of human capital in two steps. First, we use the French matched employer-employee dataset. In the spirit of Abowd, Kramarz and Margolis (1999), we regress the (log) fraction of the wage bills of all firms in a sector that goes to a given occupation on occupation $\times$ sector fixed effects. We interpret the occupation $\times$ sector fixed effects estimates as a score reflecting the sector-specific human capital of worker occupations. The higher the fixed effect of a given occupation in a sector is, the larger the *average* share of a firm's wage bill that goes to the corresponding worker type. Second, we aggregate worker human capital at the firm level. A firm's internal human capital is the average of the (exponentiated) occupation $\times$ sector fixed effects present in the firm's workforce. Our measure of human capital captures the extent to which the existing workforce is adapted to the

[5]Cisco's acquisition strategy illustrates this trade-off. In 1997, a Cisco analyst described the strategy of the firm as follows: "in today's economy, building work teams from scratch can be yesterday's luxury. So, when you can't build fast enough, you buy". By 2017, Cisco had undertaken more than 200 M&As to "provide a capability, acceleration potential or earlier sectoral entry compared to partnering or developing in-house" (Wysocki, 1997; Romanski, 2017).

sector of entry.[6]

Our measure of human capital has several advantages. First, because the French matched employer-employee dataset contains a detailed occupation code at the individual level, our measure of human capital encompasses various determinants of individuals' human capital, including skills, education and experience (Becker, 1962; Gibbons and Waldman, 2004; Autor and Dorn, 2009). Second, the fixed effects strategy allows us to rank occupations within sectors following their average contribution to firms' output. Third, our measure of human capital does not reflect firm-specific unobservables, such as personnel policies or the influence of unions, that may alter the returns to observable and unobservable dimensions of human capital.

Another empirical challenge for our paper is to identify how firms diversify. We use a detailed breakdown of firms' sales across sectors to identify entries in new sectors.[7] A firm enters a new sector if (i) at least one of its subsidiaries begins selling in the new sector and (ii) none of the other subsidiaries already operates in the sector. We identify buy entries by linking M&A deals retrieved from SDC Platinum and Bureau van Dijk Zephyr to the French administrative data.[8] The entry is a "buy" if the subsidiary that begins selling in the new sector has been acquired by the firm. By contrast, a firm "builds" if entry is made though an existing subsidiary. Our final dataset consists of 75,000 build or buy decisions in France from 2003 to 2014.

We provide the first cross-industry statistics on firms' decisions to build or buy. We find that 98% of entries in a new sector consisted of firms that build on their internal resources. The figure is 90% when weighting by entry sales, meaning that diversification by acquisition represents 10% of the universe of corporate diversification. At the time of entry, firms invest on average one million euros, and approximately half of the firms in our sample continue to operate in the sector of entry after one year. The median (mean) entry sales are equal to 270,000 euros (2.8 million euros).

Our main test analyzes whether firms' existing workforce composition explains the decision to build or buy. We compare firms that operate within the same sector of origin and that

[6]Hereafter, we refer to this variable as "internal human capital" without mentioning that it is specific to the sector of entry.

[7]The detailed breakdown of firm sales across sectors is available in the ESA survey, which is also maintained by the French Bureau of Statistics. In our main analysis, we define sectors according to the 5-digit code of the French Standard Industry Classification (SIC). We test the robustness of our main results by defining sectors at the 1-, 3-, or 5-digit levels of the French SIC.

[8]We develop a web-crawler that captures acquiring and target firms' names and addresses to link SDC Platinum and Zephyr deals to the French administrative data.

then diversify within the same sector of entry in the same year. This specification neutralizes potential unobservable time-varying synergies between the sector of entry and that of origin. We find that firms are more likely to buy when their human capital is not adapted to the sector of entry. A one-standard-deviation decrease in human capital is associated with a 1.1 percentage-point increase in the likelihood of buying. This relationship is sizable, equal to 50% of the unconditional probability of buying. This finding holds when adding firm fixed effects or when controlling for a wide variety of firm characteristics such as size, profitability, capital intensity, cash holdings and other assets.

We then exploit cross-sectional variations in the diversifying and entering firms' characteristics to understand the interactions between human capital and other potential determinants of the decision to build or buy. First, we show that human capital plays a role in the decision to build or buy regardless of the size of the firm, the severity of financial constraints and whether the firm is publicly or privately owned. Second, we construct two measures of distance between the firm's sector of origin and that of entry. The first measure is a product market distance measure that captures the complementarities between the firm's sector of origin and the sector of entry (e.g., production synergies or a common customer base). The second measure captures the physical distance between a firm and the geographical location of its entry market. We find that both product market distance and physical distance to the sector of entry increase a firm's likelihood to enter that market by acquiring an existing company. Moreover, the interaction of human capital with our distance measures shows that the magnitude of the human capital coefficient increases when a firm enters a product market that differs greatly from the firm's sector of origin. This finding suggests that internal resources matter more when the cost of building is likely to be high.

We check the robustness of our results by considering alternative measures of human capital and excluding alternative mechanisms that could drive our results. First, we weight occupations by the number of employees to account for the exact composition of the workforce; we focus on key occupations only, and we exclude CEOs, whose wages may reflect not only productivity but also agency conflicts within the firm, and we modify the unit of observation from firms to plants. Second, we check that the scale of the new activity and possible complementarities between human and physical capital (e.g., equipment, or machinery) do not fully confound the role of human capital in the decision to build or buy. On the one hand, we compare firms

that operate in the same sector of origin, enter the same sector in the same year, and make similar sales. On the other hand, we compare firms that make similar capital expenditures when entering the new sector. Our point estimates remain nearly unchanged. Third, we check that firms that build and have a lower internal human capital adjust their workforce more following entry. Indeed, firms with low human capital hire relatively more new workers, especially in key occupations. Fourth, we exploit the heterogeneity in the workforce composition of subsidiaries within building firms, and we show that entry into the new sector is more often made through subsidiaries with higher internal human capital. This finding supports the view that firms minimize labor adjustment costs when contemplating entry in a new sector.

Furthermore, we attempt to mitigate concerns that firms likely take the decision to diversify *jointly* with the decision to build or buy in the sector of entry. As our sample consists of diversifying firms, self-selection could create a spurious relationship between human capital and a firm's decision to build or buy. We focus on two plausible scenarios of how a firm is likely to jointly select the sector of entry and the mode of entry. In the first scenario, we focus on firms that operate in a declining sector and may thus be willing to shift their operations and reallocate workers to a better-performing sector (Tate and Yang, 2016*b*; Baghai et al., 2018). We would then expect the firm to select a sector in which it can easily redeploy its workforce. In a second scenario, we focus on serial acquirers that are likely to always diversify by acquisition (Golubov, Yawson and Zhang, 2015). To test the first scenario, we exclude firms that shift a substantial part of their operations to another sector. For the second scenario, we exclude serial acquirers from the analysis. In both tests, we find that our main results hold, suggesting a limited role for selection issues.

Our second key finding is that firms are more likely to buy when it is costly to hire key workers on the external job market, i.e., when key worker occupations are in short supply. We test this second prediction from our model using occupation-level data on local labor market (LLM) tightness obtained from the French unemployment agency. The data report worker occupations that are in short supply across 350 LLMs. We document considerable geographic heterogeneity in worker availability across LLMs (Moretti, 2010). We create a measure of LLM tightness based on the weighted average human capital of occupations in short supply within the zip-code area where the firm enters. Our measure of LLM tightness takes higher values if key occupations for the sector of entry are in short supply in the LLM where the firm enters.

We test whether the LLM tightness measure explains the decision to build or buy and interact it with our measure of human capital. We find that firms are more likely to buy when LLMs are tight and that the relationship between human capital and the choice to build or buy is driven by the highest tercile of LLM tightness.

We then explore the performance implications of the choice to build or buy and their relationship with labor market frictions. When LLMs are tight, hiring is costly. Therefore, after entry, we find that firms with less-adapted human capital grow at a slower pace than firms with highly adapted human capital. In other words, firms with less-adapted human capital sell less in the sector of entry. In addition, we find that this relationship is driven by sectoral entries that coincide with tight LLMs. Taken together, these findings suggest that labor market frictions are critical to understanding the role of human capital in diversification decisions.

Finally, we test whether firing costs affect firms' decision to build or buy. When a firm buys to enter a new sector, it has to incur the cost of restructuring the target (e.g., Dessaint, Golubov and Volpin, 2017). According to our model, the higher the restructuring cost, the less attractive is the option to buy. Furthermore, the more adapted an acquiring firm's internal human capital is, the more workers must be laid off after the acquisition because of the greater overlap of key worker occupations. Thus, higher firing costs make the option to buy less attractive for firms with adapted internal human capital. In a first set of tests, we use the ratio of permanent to temporary workers in a firm's workforce to proxy for firing costs, hence the ease with which firms can restructure their workforce. One potential problem with this test is that firms endogenously decide whether to hire workers under permanent or temporary contracts. Therefore, we perform a second set of tests using variation across jurisdictions in the average length of labor case settlements as a proxy for firing costs that is exogenous to firms (see, e.g., Fraisse, Kramarz and Prost, 2015).[9] A longer length of labor case settlements is associated with higher firing costs. In both tests, we find a positive relationship between firing costs and firms' likelihood of growing by acquisition, though this relationship is not statistically significant. However, we find that firing costs interact significantly with our measure of human capital, suggesting that higher firing costs increase in the importance of existing human capital resources in firms' decision to build or buy.

[9]In France, labor disputes between firms and workers are settled by local labor courts. The local court in which the labor case is settled is determined by the firm's location.

Related literature. Our paper contributes to the rapidly growing literature on labor and corporate finance.[10] Specifically, our paper is related to a strand of papers that link the organization of firms and labor economics.[11] Closely related to our paper, Tate and Yang (2016*a*) predict diversified M&As by cross-industry labor flows under the condition that human capital is transferable across sectors. Ouimet and Zarutskie (2016) and Chen, Gao and Ma (2018) show that the desire to gain human capital is an important motive for corporate acquisitions. Lee, Mauer and Xu (2018) find that firms are more likely to merge and have better post-merger outcomes when the target firm has similar human capital. We make three contributions to this literature. First, we propose a firm-level measure of human capital that captures complementarities between firms and sectors. Second, we show that internal human capital resources predict how firms diversify. Third, our results emphasize the importance of labor shortages to understanding the relationship between human capital and M&As. Overall, we show that employment composition shapes the boundaries of the firm.

Our paper also relates to the literature on the determinants of the decision to build or buy across sectors. Very few papers have jointly studied these alternative approaches to diversification.[12] McCardle and Viswanathan (1994) show that firms enter a new sector through acquisition when barriers to entry are high.[13] Phillips and Zhdanov (2013) show that in equilibrium, small firms invest in R&D, whereas large firms buy those small firms that have successfully innovated. Moreover, 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). We contribute to this small strand of the literature by documenting that, at the scale of the French economy, more than 90% of corporate diversifications are made by firms that build from scratch using their preexisting resources rather than buying an incumbent in the sector of entry.

Finally, our paper adds to the literature on the theory of the firm. In the finance literature,

[10]For instance, a strand of the literature has recently explored the implications of labor adjustment costs for corporate investment (e.g., Merz and Yashiv, 2007; Xu, 2018; Bai, Fairhurst and Serfling, 2018) and capital structure decisions (e.g., Matsa, 2010; Agrawal and Matsa, 2013; Simintzi, Vig and Volpin, 2014; Baghai et al., 2018; Serfling, 2016).

[11]Other papers study the consequences of organizational changes, such as M&As, on employment and wages (e.g., Lagaras, 2017; Ma, Ouimet and Simintzi, 2016).

[12]The literature on corporate diversification is large and focuses primarily on the choice of sector in the decision to diversify. See Maksimovic and Phillips (2013) for a review of the literature.

[13]In the strategy literature, Yip (1982)'s empirical study supports McCardle and Viswanathan (1994)'s theoretical analysis and shows that firms are more likely to build in a sector with low barriers to entry. In the international economics literature, Nocke and Yeaple (2007) study the choice of entering a new country via foreign direct investment or acquiring an existing company.

the dominant view has been the "property rights" theory (Grossman and Hart, 1986; Hart and Moore, 1990; Rhodes-Kropf and Robinson, 2008), according to which value is created by combining complementary assets under the control of a single firm.[14] Another view from the early economics and strategy literatures has regained interest in trade and international economics. According to this "resource-based" view, the decision to grow depends on preexisting resources and transferable capabilities (Penrose, 1955; Chandler, 1992). Matsusaka (2001) theoretically shows that firms dynamically modify their portfolio of activities to match their organizational capabilities. Along these lines, Bernard et al. (2007) and Bernard et al. (2018) document that firms are much more likely to produce in certain pairs of industries. Boehm, Dhingra and Morrow (2019) take a step further and show that firms tend to co-produce in industries that require similar intermediate inputs. Our paper builds on this theory and focuses on another type of input: labor. We show that firms are more likely to diversify by building on their existing resources when complementarities exist between the firm's internal human capital and the key skills needed to produce in the new sector. Our results imply that firms buy when they do not have adapted inputs in house, especially when these inputs are scarce on external markets.

2. Theoretical framework: "build" or "buy"?

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? We propose a model that predicts firms' decision to build or buy based on the adaptability of their workforce to the new sector.

2.1. Basic framework

Costs and profits. To enter a new sector, firms must develop new productive capacities, that is, combine additional inputs to produce in the new sector. In the model, we assume that labor L is the only factor of production, so the production function is $Y = \mathcal{L}$.[15] Firms can select workers from three different pools. First, firms can reallocate workers from their *internal* labor market and have them produce in the new sector. Second, firms can hire new workers on the

[14] Rhodes-Kropf and Robinson (2008) show that the search for complementarities implies the existence of an assortative matching between acquirers and target firms, i.e.,"like buys like". An important strand of the empirical M&A literature finds evidence supporting this view: Mergers are more likely to occur and create more value when merging firms have similar human capital (Tate and Yang, 2016*b*; Lee, Mauer and Xu, 2018), sell similar products (Hoberg and Phillips, 2010), use similar technology (Bena and Li, 2014), or share similar corporate cultures (Li et al., 2018).

[15] Note that in the empirical analysis, we consider the interactions between human and physical capital.

external labor market, that is, poach workers already employed or hire unemployed workers. Third, firms can *acquire* an existing company to have workers from the acquired firm produce in the new sector. The production function combines these three sources of labor with a constant elasticity (CES) of substitution:

$$\mathcal{L} = \left(\sum_i L_i^{\gamma} \right)^{\frac{1}{\gamma}}, \tag{1}$$

where $0 < \gamma < 1$ is the elasticity of substitution across worker pools and $i \in \{I, E, A\}$ denotes the different pools: internal, external, and acquired. Firms choose their mix of workers from the three pools to minimize their marginal cost of production c, defined as

$$c = \left(\sum_i c_i^{-\frac{\gamma}{1-\gamma}} \right)^{-\frac{1-\gamma}{\gamma}}, \tag{2}$$

where c_i denotes workers' marginal cost of production in each pool i. We provide microeconomic foundations for each labor type's marginal cost of production in Section 2.3.

We assume that firms engage in monopolistic competition for each variety of product and that consumer preferences exhibit a CES across products. Hence, firms' profit is proportional to their marginal cost of production c:

$$\Pi = Kc^{-(\sigma-1)}, \tag{3}$$

where σ is the elasticity of substitution between goods within a sector, $\sigma > 1$ and K is a constant measuring firms' profitability.[16]

Build. If a firm chooses to enter the new sector by using its existing resources ("build"), it can choose to combine workers from its internal labor market (L_I) and the external labor market (L_E). We assume that building does not have any impact on firms' existing business lines of production. This assumption is realistic if the introduction of new tasks does not disrupt existing tasks, for instance, if internal workers are shifted away from tasks that are either completed or sufficiently automated.

[16] In standard monopolistic competition models, this constant depends on σ and the share of consumer income devoted to a given product. In turn, this share depends on the price of a given product relative to the aggregate price index. The only relevant detail for our analysis is that this constant does not depend on the marginal cost, as is the case in standard monopolistic competition models.

$$\Pi^{Build} = K \left(c_I^{-\frac{\gamma}{1-\gamma}} + c_E^{-\frac{\gamma}{1-\gamma}} \right)^{\frac{(1-\gamma)(\sigma-1)}{\gamma}} . \quad (4)$$

Equation (1) implies that the three pools of workers are combined according to the CES parameter γ. Therefore, in addition to consumers' elasticity of substitution between products σ, a firm's monopolistic profit in (4) depends on the marginal costs of production of the different worker pools, aggregated according to a function of γ.

Buy. If a firm chooses to enter a new sector by acquiring an existing company ("buy"), it accesses the target's pool of workers (L_A), in addition to its internal labor market (L_I) and the external labor market (L_E). Accessing target workers is desirable because target firms' human capital is likely to be more adapted to the sector of entry, as discussed in Section 2.3. However, we assume that incorporating the target's full pool of workers when acquiring a firm's internal labor market leads to post-merger restructuring costs. We model these costs as a fixed cost F. In Appendix A.2, we propose a microfoundation of the fixed cost F as the costs of restructuring the workforce of the acquired firm. When firms buy, their profit can be written as:

$$\Pi^{Buy} = K \left(c_I^{-\frac{\gamma}{1-\gamma}} + c_E^{-\frac{\gamma}{1-\gamma}} + c_A^{-\frac{\gamma}{1-\gamma}} \right)^{\frac{(1-\gamma)(\sigma-1)}{\gamma}} - F. \quad (5)$$

The acquiring firm's access to the target firm's worker pool increases the variable part of (5). When a firm buys, we can theoretically show that it reallocates fewer workers from its internal labor market and hires fewer workers on the external labor market.

Moreover, firms systematically lay off some of the target's workers. These findings provide an interpretation for the fixed cost F as the post-merger restructuring costs, which is in line with Lee, Mauer and Xu's (2018), who show that laying off duplicate workers is a source of economies of scale in mergers.

2.2. Testable predictions: build or buy?

When entering a new sector, the firm compares the profit from building on its existing resources ("build") with the realized profit from acquiring an existing company ("buy"). We show in Appendix A.1 that if the different pools of workers are sufficiently substitutable ($\gamma > (\sigma-1)/\sigma$) and F is low enough, the model implies that there exists a unique threshold $c_I^* > 0$

such that if $c_I > c_I^*$, we have $\Pi^{Buy} > \Pi^{Build}$. Otherwise, if $c_I < c_I^*$, we have $\Pi^{Buy} > \Pi^{Build}$. In this way, we obtain prediction 1.

Prediction 1. *Firms optimally choose to build when their marginal cost of production is low enough, that is, if their existing workforce is more adapted to operate into the sector of entry.*

As we show in Appendix A.1, the model also implies $\frac{\partial c_I^*}{\partial c_E} < 0$, from which prediction 2 follows.

Prediction 2. *Firms are more likely to buy when key workers for the sector of entry are in short supply in the external labor market.*

Prediction 2 implies that prediction 1 is stronger when key workers for the sector of entry are in short supply in the external labor market.

2.3. Micro-foundations of the labor cost

To take predictions 1 and 2 to the data, we need a micro-foundation for the labor costs c_i. In this section, we propose one, based on Cheng and Morrow (2018), in which labor costs depend on the availability and efficiency of workers in each pool.

Worker occupations. We assume that there are different types of workers. Each worker type is employed in a given occupation. Each worker occupation, denoted $o \in \mathcal{O} = \{1, ..., O\}$, is available in quantity $(a_{i1}, ..., a_{iO})$. Importantly, worker availability differs across worker pools $i \in \{I, E, A\}$: The number of workers in each occupation is different in the acquiring firm's internal labor market, in the external labor market, and in the target's internal labor market. Worker characteristics differ across occupations: Workers' wages $(w_1, ..., w_O)$ and efficiency $(m_1, ..., m_O)$ differ across occupations but not across workers within an occupation.[17]

Recruiting process. To recruit workers, the firm conducts interviews with several workers from each pool $i \in \{I, E, A\}$. The match between a firm and a worker is assumed to be of random quality $h \geq 1$ and to follow a Pareto distribution with cumulative density function $\Psi(h) = 1 - h^{-k}$ with $k > 1$. This distribution is assumed to be equal across worker pools; that is, firms' expected match quality is the same regardless of whether workers come from inside

[17] We, therefore, drop the subscript i relative to wages and efficiency.

the firm (i.e., the internal labor market) or from outside (i.e., the external labor market or from the target firm's workforce). This assumption holds as long as workers' new tasks in the sector of entry are sufficiently different from the tasks to which they were previously assigned. In this case, there is no reason for firms to have different expected match quality from workers from different pools.[18] Firms observe the match quality during the interview.

We assume that conducting interviews is labor-intensive and costs f per interview. This assumption implies that the degree of information asymmetry between firms and workers does not vary across the three worker pools. There are two reasons that we assume the cost f to be the same across the different pools of workers. First, this assumption simplifies the theoretical analysis and does not change the predictions. Second, empirically, we cannot observe the cost of interviews and hiring costs across firms or differentiate these costs across worker pools.

The firm selects workers to interview by fixing a match quality threshold $\underline{h}_{io}$ below which they do not hire workers they interview. After the interviews, the firm hires a total number $N_i a_{io} \underline{h}_{io}^{-k}$ of workers in each occupation $o \in \mathcal{O}$. The total labor cost of conducting interviews is then $f N_i c_i$.

Costs and production function. The firm's total labor cost C_i in each pool of workers i sums to the cost of conducting interviews $f N_i c_i$ and the wages of selected workers:

$$C_i = N_i \left(\sum_o a_{io} w_o \underline{h}_{io}^{-k} + f c_i \right). \tag{6}$$

For a given number of interviews N_i and a quality threshold $\underline{h}_{io}$, a firm can estimate its workforce's expected level of human capital. Each worker occupation $o \in \mathcal{O}$ is associated with an expected level of human capital equal to H_{io}.[19] Within each pool of workers, an occupation's expected human capital represents the input in the firm's production function (1) as follows:

$$L_i = \left(H_{i1}^{\theta} + ... + H_{iO}^{\theta} \right)^{1/\theta}, \tag{7}$$

[18] In the empirical analysis, we control for the product market distance between firms' existing business lines and the new sector of production. We find that the data continue to support our model's predictions. Moreover, we find stronger support for our model's predictions when the new sector is distant from firms' existing sectors of activity, that is, when it is most likely that k is the same across worker pools.

[19] The match quality follows a Pareto distribution, which implies a simple formula for each worker occupation's expected human capital: $H_{io} \equiv N_i a_{io} m_o \int_{\underline{h}_{io}}^{\infty} h d\Psi(h) = \frac{N a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}$.

where $\theta < 1$ is the elasticity of substitution between workers' human capital across different occupations. We assume the elasticity to be constant across worker pools $i \in \{I, E, A\}$. This assumption means that the substitutability across occupations is determined by the nature of tasks involved in the production process.

Our model requires that both the interview cost *and* the match quality distribution be constant across worker pools. As a consequence, workers from the internal, external, and the acquired firm's labor markets are perfectly substitutable. Hence, the model does not allow for firm-specific organizational capital or firm-specific human capital.[20] The variation in labor costs across worker pools comes from the different vectors of occupation availability $(a_{i1}, ..., a_{iO})$.

Cost minimization. The firm minimizes the total labor cost C_i in each worker pool (6) by choosing the number of interviews N_i to conduct and the occupation-specific match quality threshold $\underline{h}_{io}$. Note that the firm chooses $\underline{h}_{io}$ and N_i by trading off the quality of the hired workers and the search costs associated with recruiting process. On the one hand, hiring a large number N_i of workers enables firms to select the best matches by choosing high values for $\underline{h}_{io}$. On the other hand, the firm saves on search costs f by choosing a smaller number of workers N_i and low values of $\underline{h}_{io}$. However, the firm takes as given the expected human capital supplied by its workers H_{io} and the production function (7).

We show in Appendix A.3 that this constrained minimization problem results in the following marginal labor cost:

$$c_i = \left(\sum_{o \in O} \left(\frac{a_{io} m_o^k w_o^{1-k}}{f(k-1)} \right)^{\frac{\theta}{\beta}} \right)^{\frac{\beta}{\theta(1-k)}}, \tag{8}$$

where $\beta \equiv \theta + \theta\,(1-k)$.[21] Equation (8) implies that the marginal labor cost in worker pool $i \in \{I, E, A\}$ increases with the interview cost f. Indeed, higher interview costs reduce the optimal number of interviews. Matches' average quality decreases, and, in turn, the human capital supplied by each worker in an occupation is lower. This expression highlights that the labor cost is determined by the byproduct of workers' efficiency and availability. It is large when

[20] One way to model firm-specific organizational or human capital would be to have workers' efficiency m_0 differ across worker pools. This worker-pool-dependent efficiency would act as a subsidy for either build or buy, depending on the firm-specific human capital in the diversifying firm relative to the acquired firm. Our model's qualitative predictions would be unchanged.

[21] The assumption that $k > 1$ implies that $\beta > 0$, in which case it is optimal for firms to hire workers in every occupation (Cheng and Morrow, 2018).

occupations that are efficient in the new sector (high m_o) are in scarce supply in the worker pool (low a_{io}).

Wage-bill share of each occupation. We denote by A_{io} the total number of workers in occupation o hired to produce one unit of output. We show in Appendix A.4 that the share of the wage bill that goes to a given occupation o can be expressed as:

$$\frac{w_o A_{io}}{\sum_o w_o A_{io}} = \frac{\left(a_{io} w_o^{1-k} m_o^k\right)^{\frac{\theta}{\beta}}}{\sum_o \left(a_{io} w_o^{1-k} m_o^k\right)^{\frac{\theta}{\beta}}}. \tag{9}$$

Equation (9) implies that within a worker pool $i \in \{I, E, A\}$, the share of a firm's wage bill that goes to workers in a given occupation depends on (i) occupation-specific wages w_o, (ii) the availability a_{io} of workers in that occupation, and (iii) workers' occupation-specific efficiency m_o. The interpretation is that within a given sector, some occupations receive relatively higher wages than other occupations when they are relatively more efficient at producing in this sector.

3. Empirical strategy

To take predictions 1 and 2 to the data, we need to estimate the marginal labor cost (8) for workers from the internal labor market, i.e., c_I. In Section 4, we explain that neither the availability a_{Io} of internal labor market workers in each occupation nor the occupation-specific efficiency m_o are observable in the data. However, based on our model, we propose a method to estimate c_I using administrative worker-level occupation data.

3.1. Occupation-specific human capital

Equation (9) allows us to estimate the average availability a_{Io} of workers in occupation o and the occupation-specific efficiency m_o. We use the subscripts f, o, n, and t, for firm, occupation, the sector of the firm, and time, respectively. At the firm level, we denote by $\text{Share}_{f,o,n,t}$ the share of the wage bill that goes to occupation o in firm f operating in sector n, i.e., the ratio given by Equation (9). Rewriting worker availability $a_{f,o,n,t}$ as a deviation from the sectoral average $\bar{a}_{o,n,t}$, we have $a_{f,o,n,t} = \bar{a}_{o,n,t} \cdot \widetilde{a}_{f,o,n,t}$. Then, taking the logarithm of

Equation (9), we obtain

$$\log\left(\text{Share}_{f,o,n,t}\right) = \frac{\theta}{\beta}\log\left(\bar{a}_{o,n,t}m_{o,n,t}^{k}w_{o,n,t}^{1-k}\right) - \log(\sum_{o'\in\mathcal{O}_{gt}}\left(a_{g,o',t}m_{o',n,t}^{k}w_{o',n,t}^{1-k}\right)^{\frac{\theta}{\beta}}) + \frac{\theta}{\beta}\log\left(\widetilde{a}_{f,o,n,t}\right), \tag{10}$$

where $\mathcal{O}_{ft}$ is the set of occupations observed in firm f at time t. Equation (10) provides a decomposition of $\text{Share}_{f,o,n,t}$, which can be estimated as the following fixed effects regression:

$$\log\left(\text{Share}_{f,o,n,t}\right) = \mu_{o,n,t} + \nu_{f,t} + \epsilon_{f,o,n,t}, \tag{11}$$

where $\mu_{o,n,t}$ is an occupation $\times$ sector $\times$ time fixed effect capturing the average wage share that goes to occupation o at the level of sector n. $\nu_{f,t}$ is a firm $\times$ time fixed effect, and $\epsilon_{f,o,t}$ is an error term capturing the deviation of occupation o's share in the firm's wage bill from the sectoral average.

The estimation of $\mu_{o,n,t}$ requires firms to hire more than one type of occupation each year and every occupation to be present in more than one firm in a given sector in each year.[22] In Table B1, we report the explanatory power of our first-stage Equation (11). Column 1 shows our first stage as used in the rest of the paper. The dependent variable is the (log) share of a firm's wage bill that goes to a specific occupation. and we regress it on Occupation-Sector-Year FE *and* Firm-Year FE. In columns 2-6, we consider different combinations of fixed effects.We find that 72.3% of the within-firm variation in the log-share allocated to a specific occupation is explained by Occupation-Sector-Year FE and Firm-Year FE. Occupation-Sector-Year FE alone explain 53% of the variation (column 2), and Firm-Year FE alone explain 51% of the variation, suggesting that the combination of the two sets of fixed effects significantly improves the percentage of the variation in within-firm occupation log-share explained by our first stage displayed in column 1.

We use the estimated values, denoted $\hat{\mu}_{o,n,t}$, to construct our main explanatory variable below (Section 3.2). We interpret them as a score reflecting the (wage- and availability-adjusted) human capital of a given occupation at the sectoral level. The higher the fixed effect of a given occupation in a sector is, the larger the score for this occupation in this sector.[23]

[22]To ensure that these two conditions are met, we exclude firms that hire only one type of occupation and employ fewer than 20 workers from the sample. In addition, in the empirical analysis, we exclude occupation-sector-year triplets with fewer than 10 firms to obtain more precise estimates.

[23]As an illustration, Table B2 displays the five occupations with the largest fixed effects in the sectors of pharmaceutical preparations, IT consultancy activities and the manufacture of motor vehicles in 2013. The

3.2. Firm's internal human capital

We rely on the occupation scores to analyze the role of human capital in the decision to build or buy. We use the subscripts g, o n, and t, for the diversifying firm, worker occupation, sector of entry, and time, respectively. The marginal labor cost (8) can then be rewritten as $c_{I,g,n,t} = (f(k-1))^{\frac{1}{k-1}} (\psi_{I,g,n,t\text{-}1})^{\frac{\beta}{\theta(1-k)}}$, where

$$\psi_{I,g,n,t} = \sum_{o} \left(a_{I,g,o,n,t} w_{g,o,n,t}^{1-k} m_{g,o,n,t}^{k} \right)^{\frac{\theta}{\beta}}. \tag{12}$$

We interpret $\psi_{I,g,n,t}$ as a proxy for the human capital of workers from a firm's internal labor market. Indeed, $c_{I,g,n,t}$ is a decreasing function of $\psi_{I,g,n,t}$, that is, the marginal cost of labor of operating in sector n is lower when the human capital of firms' existing workers is high for that sector.

Rewriting $\psi_{I,g,n,t}$ in terms of the occupation $\times$ sector $\times$ time fixed effect $\mu_{o,n,t}$ and the firm-level deviation from the sectoral average availability of workers in an occupation, $\widetilde{a}_{g,o,t}$, we have

$$\psi_{I,g,n,t} = \sum_{o} (\widetilde{a}_{g,o,n,t})^{\frac{\theta}{\beta}} \exp(\mu_{o,n,t}). \tag{13}$$

Predictions 1 and 2 consist of predicting firms' choice to build or buy, based on their existing workforce composition. Therefore, we need to empirically estimate firm g's human capital $\psi_{I,g,n,t-1}$ one year before entering the new sector.

The key challenge to estimating (13) is that we do not observe firms' internal reallocation of existing workers in the sector of entry; that is, we do not observe $\widetilde{a}_{g,o,n,t}$. Therefore, as soon as occupation o is present in g's existing workforce at $t-1$, we assume that $\widetilde{a}_{g,o,n,t} = 1$, meaning that worker availability in the internal labor market is equal to the average availability in the new sector. Otherwise, if $a_{g,o,n,t} = 0$, we assume that $\widetilde{a}_{g,o,n,t} = 0$.[24] We obtain our main independent variable, which is an empirical estimation of the value of (13) at $t-1$:

$$\text{Internal Human Capital}_{g,n,t-1} = \frac{1}{\#O_{g,t-1}} \cdot \sum_{o \in \mathcal{O}_{g,t-1}} \exp\left(\hat{\mu}_{o,n,t-1}\right), \tag{14}$$

selected occupations indeed seem to play an important role in their sectors.

[24] In the empirical analysis, we test several alternative assumptions as robustness checks (see Section 6.1). In particular, we use the fraction of each occupation as weights in the computation of Internal Human Capital$_{g,n,t-1}$. Our results are robust to alternative assumptions in our estimation of Equation (13).

where $O_{g,t-1}$ is the set of occupations present in firm g's internal labor market prior to entering the new sector n (i.e., occupations for which $a_{g,o,n,t-1} > 0$) and the values of $\hat{\mu}_{o,n,t-1}$ come from the estimation of Equation (11).

Internal Human Capital$_{g,n,t-1}$ is a measure of the human capital of firm g's existing workforce for the sector of entry n. It is defined as the average value of the (exponentiated) occupation $\times$ sector $\times$ year fixed effects of the occupations present in the workforce of the firm prior to entry, given by (11). We use the average value instead of the sum to avoid human capital being mechanically larger for firms employing workers in more occupations.[25]

Internal Human Capital$_{g,n,t-1}$ takes a high value when occupations with a high (wage-adjusted) efficiency $m_{o,n,t-1}$ for the sector of entry are already present in a firm's internal labor market prior to entry. We interpret it as a measure of whether a firm's workforce is adapted for a given sector of entry. In the remainder of the paper, we refer to this measure as "human capital" or "internal human capital" without explicitly mentioning that it is specific to a given sector of entry.

3.3. Empirical model

Prediction 1 states that firms are more likely to "buy", as opposed to "build", when their existing internal human capital is not adapted to the sector of entry. This pattern is more pronounced when key workers for the sector of entry are difficult to hire in the external labor market (prediction 2).

We test these predictions by analyzing the link between the type of entry and the constructed measure of internal human capital. The dependent variable is $\mathbb{1}(\text{Buy})_{g,n,o,t}$, a dummy equal to one if firm f enters a new sector n through an acquisition ("buy") and zero if it enters the new sector by building on its existing resources ("build"). o indicates firm f's main original sector of activity, i.e., the sector in which the firm realizes the largest share of its sales prior to entry. Our baseline empirical model is as follows:

$$\mathbb{1}(\text{Buy})_{g,n,o,t} = \lambda_{n,o,t} + \delta \text{Internal Human Capital}_{g,n,t\text{-}1} + \beta X_{g,n,o,t-1} + \eta_{g,n,o,t} \qquad (15)$$

This presence of origin $\times$ entry $\times$ time fixed effects implies that the role of internal human

[25] We also control for firm size using the number of employed workers. In unreported regressions, we control for the number of occupations to closely follow the definition of $\psi_{I,g,n,t\text{-}1}$. Our results are robust to this change.

capital is identified by comparing the diversification strategy of firms operating in the same sector of origin and entering in the same new sector. All unobservable time-varying synergies and complementarities between sectors are therefore captured by the fixed effects $\lambda_{n,o,t}$. In addition, this specification also controls for unobservable factors related to the sector of entry (e.g., fixed costs of entry, barriers to entry. See McCardle and Viswanathan (1994)) or the sector of origin (e.g., ability to collateralize assets to access external finance. See Rajan and Zingales (1998)).

The vector $X_{g,n,o,t-1}$ includes other firm characteristics that may influence the decision to build or buy. At the firm level, such determinants are firm size, cash holding, tangibility, labor productivity and product market relatedness (i.e., the distance to the sector of entry based on sales information). We also control for the variation in labor market tightness across geographical zones and sectors.

By construction, internal human capital$_{g,n,t\text{-}1}$ introduces a measurement error term that generates a correlation between residuals $\eta_{g,n,o,t}$ at the level of the sector of entry. It might also be more or less precisely estimated depending on the sector of origin. We therefore double-cluster standard errors to control for correlations within the sector of entry and within the sector of origin. Finally, we standardize the internal human capital$_{g,n,t\text{-}1}$ variable in the rest of the analysis to interpret regression coefficients in standard deviation units.

4. Data and summary statistics

4.1. Data sources

4.1.1 Firm data

Our primary source of data is French administrative data provided by the French Bureau of Statistics (INSEE). Our definition of a firm includes the parent company and the majority-owned subsidiaries (more than 50% of the shares), which we identify using a dataset containing ownership links (*Enquête sur les Liaisons financières entre sociétés*, LIFI).We recover firms' main sector of activity, balance-sheet information and income statements from the tax files (*Bénéfices Industriels et Commerciaux* and *Bénéfices Non-commerciaux*). We use the SIRENE registry to obtain the geographical location of the different plants of each firm. We then con-

solidate all variables to obtain observations at the parent firm-level.[26]

To identify diversified entries into new sectors, we rely on the subsidiaries' breakdown of sales by sector (*Enquête Annuelle de Production*). This dataset records the detailed amount of sales realized by subsidiaries in every sector and year. The survey is exhaustive for subsidiaries with at least 20 employees and randomly includes smaller subsidiaries, such that the survey covers at least 85% of sales realized within a given sector. For instance, subsidiaries included in the survey cover 96% of sales in the manufacturing sector. For smaller subsidiaries for which sales breakdowns are not available in the survey, we make the assumption that these subsidiaries sell only in their main sector of activity, and these data are retrieved from the tax files. We exclude entries occurring in 2008 because the methodology of the survey changed that year.

Sectors are defined by the French Standard Industry Classification (SIC) (*Nomenclature des activités Francaises, NAF*), which is equivalent to the US SIC. We define sectors by 5-digit French SIC codes. The dataset includes 732 different sectors.[27]

4.1.2 Merger and acquisition deals

We merge the French administrative datasets with a dataset of M&A deals retrieved from SDC Platinum and Bureau van Dijk Zephyr.[28] We collect all deals between January 2003 and December 2014 that involve a French acquirer and a French target. We exclude leveraged buyouts and private equity deals from the sample. We focus on deals in which the acquirer owns less than 50% of the target shares before the acquisition date and more than 50.1% after to identify changes in majority ownership between operating companies.

SDC Platinum and Bureau van Dijk's Zephyr do not provide French firm standardized identifying numbers (SIREN). We proceed in several steps to retrieve the unique firm identifiers. First, we use tickers (available only for publicly traded firms) and the Bureau van Dijk identifiers (available only for Zephyr deals) to recover a fraction of the firm identifiers. Second, we build a Python webcrawler on two websites, which takes as inputs a firm's name and address:[29]

[26]Note that we exclude firms in the financial, agricultural and public sectors because they use different accounting systems, which limits the comparability and relevance of standard variables across sectors.

[27]Note that we cannot replicate Hoberg and Phillips's (2010) product market distance with our data. This measure is based on textual analysis of the 10k filings of U.S. firms, information that is not available in France, especially not for both public and private firms.

[28]Note that an ownership change in the ownership links dataset cannot be directly used to identify M&As deals. Some ownership changes correspond to new entries in the database that do not necessarily correspond to new ownership links.

[29]The webcrawler builds on the Python packages Selenium and Beautiful Soup.

(i) www.bodacc.fr (Bulletin Officiel des Annonces Civiles et Commerciales), which is a governmental website that reports official notifications involving French companies since 2003, and (ii) 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 web-crawler, 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 to the original name is below a certain threshold.[30] We retain observations with a distance above 0.8. Fourth, we manually check the resulting matches. Our final sample contains 7,303 deals from 2003 to 2014. To the best of our knowledge, this is the most comprehensive M&A dataset available to date for the French economy.

4.1.3 Worker-level occupation worker data

We use the French matched employer-employee administrative dataset (*Déclarations Annuelles des Données Sociales*, DADS) to construct the measure of internal human capital. Firms are required by law to report every year detailed information about their workers when filing payroll taxes.[31] The employer must report the type of contract, gross and net wages, the number of hours worked and an occupation code for each worker. The data set also indicates whether the worker is employed with a permanent contract (*Contrat à durée indéterminé,* or CDI). Note that before 2003, occupations are often missing from the dataset. This is why we begin the empirical analysis at this date. Occupations are reported as 4-digit codes. The French nomenclature of occupations (*Nomenclatures des professions et catégories socio-professionnelles des emplois salariés des employeurs privés et publics*, PCS-ESE) consists of 414 different occupations, including, for instance, 28 different types of engineers (e.g., logistics, IT, electrical, or mechanical).

[30]The Jaró-Winkler distance measures the number of characters in common between strings under the assumption that differences near the start of the string are more significant than differences near the end of the string.

[31]Note that reporting of the occupation code is required for firms that employed at least 20 employees in a given year and optional for firms below the threshold.

4.1.4 Labor market frictions and firing costs

We measure hiring frictions at the LLM level using data from the French national unemployment agency (*Pôle emploi*). The unemployment agency lists job vacancies, helps unemployed people find jobs and produces national and local unemployment statistics, which we use in this paper. In particular, the unemployment agency tracks occupations in short supply in 350 different LLMs starting in 2010. Occupations are flagged as being in short supply when (i) job offers exceed job applications and (ii) surveyed employers anticipate that they will not fill in a job. Figure 3 maps the number of occupations in short supply by LLM in 2013.[32] Darker shades of blue indicate a higher degree of shortage in the LLM. Interestingly, we see that labor market tightness is not systematically related to population density, as tight LLMs can be observed in both urban and rural areas.

To measure firing costs, we use 2010 data from the French Ministry of Justice website on local labor courts that deal with labor disputes between firms and workers (*Prud'hommes*). Elected judges representing employers and employees sit on the local labor courts. The local court in which the labor case is settled is determined by the firm's location. Due to a judicial reform, the number of local labor courts decreased from 271 to 210 in 2008. Therefore, to ensure constant coverage over time, we aggregate LLM-level observations at a more aggregate level of the judicial map, called a "jurisdiction". There are 140 jurisdictions over our sample period.

To proxy for firing costs, we use variation across jurisdictions in the average length of labor case settlements. The implicit assumption here is that the length of local labor case settlements reflects a high firing cost for firms. Long settlements are likely to reflect frequent worker litigation (Fraisse, Kramarz and Prost, 2015). Moreover, they increase the administrative burden associated with worker dismissals.

4.2. Main variables

4.2.1 Type of entry

The main dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable equal to one if the entry of firm f in sector n at time t is made through an acquisition ("buy") and equal to zero if the entry is made internally ("build"). We consider a firm to have entered a new sector if (i) at

[32] Appendix Table B4 lists occupations that are in short supply in 2013.

least one of its subsidiaries begins selling in that sector and (ii) none of the other subsidiaries already operates in the sector.

A firm enters a new sector by acquisition if the entity that reports sales in the new sector ("entering subsidiary") becomes one of the acquirer's subsidiaries after the M&A. By contrast, a firm builds if the entering subsidiary had already been controlled by the firm.

We also consider whether an entry in a new sector is associated with an entry in a new geographical zone. We use "region" to define the geographical zone (France had 25 regions over the sample period). A firm begins operating in a new geographical zone if either (i) the entering subsidiary is created at time t and located in a new region, (ii) the entering subsidiary opens a plant in a new region at time t, or (iii) the entering subsidiary is acquired at time t and operates in a region in which the firm was not present at $t-1$.

4.2.2 Product market distance between sector of origin and sector of entry

To study how the complementarities between the sector of entry and the sector of origin interact with internal human capital, we construct a product market distance measure based on Bloom, Schankerman and Van Reenen (2013). The firm-level variable measures the distance between the firm's sectoral repartition of sales ("sector portfolio") prior to entry to the sector portfolios of firms already operating in the sector of entry. The idea of the metric that if the sector portfolio of the entering firm is unusual compared to firms already present in the sector of entry, then the firm is "distant" to the sector of entry - otherwise it is "close".

For a firm f entering a new sector at time t, we denote by $S_g = (S_g^1, ..., S_g^N)$ the vector of sales at time $t-1$ broken down by sectors ($n = 1, ..., N$). S_g^{-n} is the vector of sales excluding sales in sector of entry n.[33] For a given sector of entry n, we define the distance $d_{g,h}^n$ between the firm and any firm h already operating in the new sector n, as (one minus) the uncentered Pearson correlation between vectors S_g^{-n} and S_h^{-n}:

$$d_{g,h}^n = 1 - \frac{(S_g^{-n} \cdot S_h^{-n})}{\sqrt{(S_g^{-n} \cdot S_g^{-n})}\sqrt{(S_h^{-n} \cdot S_h^{-n})}}$$

Finally, we define the product market distance of firm f to sector n as the weighted average of

[33] We exclude sales in sector n because a firm that enters sector n at time t necessarily reports zero sales in n at time $t-1$.

the distance $d^n_{g,h}$ for all firms h operating in sector n at $t-1$:

$$\textit{Product Market Distance}_{g,n} = \frac{\sum_h \omega^n_h d^n_{g,h}}{\sum_h \omega^n_h}$$

where the weights ω^n_h are given by the share of sales realized by firm h in sector n at $t-1$: $\omega^n_h = \frac{\text{Sales of firm } h \text{ in sector } n}{\text{Sales of firm } h}$. The weights ensure that the distance between the diversifying firm and firm h matters more if firm h realizes a large part of its sales in sector n.

4.2.3 Local labor market tightness and length of local labor case settlements

We use the list of occupations in short supply by LLM to build a time-, sector- and geographic-level measure of labor market tightness. We define the *Local Labor Tightness*$_{n,z,t-1}$ variable as the average of the (exponentiated) occupation × sector × year fixed effects of the occupations in short supply in LLM z:

$$\textit{Local Labor Tightness}_{n,z,t-1} = \frac{1}{N.L_{n,t-1}} \sum_{l \in L} \mathbb{1}(l \text{ in short supply in LLM } z) \times \exp(\hat{\mu}_{o,n,t})$$

where $N.L_{n,t-1}$ is the number of occupations in sector n at time $t-1$.[34] This sector-specific measure of LLM tightness takes high values if there are occupations in short supply in LLM z at time $t-1$ that are key for firms already operating in sector n. Figures 4a and 4b plot the geographical distribution of the variable for manufacturers of pharmaceutical preparations and motor vehicles. The graphs show that firms face, on average, more difficulties finding key workers in the second sector than in the first sector. Moreover, the measure appears to vary significantly across both LLMs and sectors.

We then use the average length of labor case settlements as a time- and geographic-level proxy for firing costs. The average length of labor procedures is 12.5 months (standard deviation: 3.6). We rely on the location of the entering subsidiary to determine the intensity of LLM frictions faced by the firm.

[34] In a robustness check, we directly use the percentage of jobs in short supply as a proxy for LLM tightness:

$$\textit{Local Labor Tightness}_{n,z,t-1} = \frac{1}{N.L_{n,t-1}} \sum_{l \in L} \mathbb{1}(l \text{ in short supply in LLM } z).$$

Our results are robust to this alternative specification.

4.3. Summary statistics

Build or buy? Panel A of Table 1 and Figure 1a present the evolution of the proportion of build and buy entries between 2004 and 2013 (excluding 2008). While at the beginning of the period, approximately 1.8% of entries are made by acquisition, this figure increases over the sample period to reach 2.49% of total entries in 2014. However, buy entries are, on average, larger than build entries. Although sales are higher for build entries than for buy entries in the aggregate (Figure 1b), when weighting by entry sales, buy entries increase to an average of 8.5% of total entries between 2004 and 2013 (Panel B of Table 1).

We then examine how these figures vary with the definition of sector in Panel B of Table 1. The proportion of buy entries remains stable at 1.7-1.9% whether we define a sector using the 1-digit or 5-digit code of the French SIC (our baseline). This finding suggests that firms do not tend to enter more by acquisition in sectors that are classified as similar in the French administrative data.[35] However, we find that build entries tend to be larger in sectors that are close to the sector of origin. Indeed, entries by acquisition represent 15% of entry sales when we use the 1-digit classification level but only 8% when we rely on the most detailed definition of sector (5-digit code).

[Insert Table 1 here]

Internal human capital. Figure 2 plots the probability density functions of firms' human capital by type of entry. The variable Internal Human Capital$_{g,n,t\text{-}1}$ is normalized to have a mean of zero and a standard deviation of one. On average, the human capital of firms that enter a new sector by acquisition is less adapted to the sector of entry than that of firms that choose to build.

[Insert Figure 2 here]

Other variables. Table 2 Panel A reports summary statistics on the different control variables we include in our baseline specification (see Section 3.3). Consistent with Table 1, the average number in the panel of firms diversifying through an acquisition is approximately 2%. Firms

[35] In the empirical analysis, we control for product market distance using our measure from Section 4.2.2 instead of the French SIC codes and find that firms tend to buy more frequently in more distant product markets (3). Hoberg and Phillips (2016) discuss several improvements over the core method underlying static sector classifications.

report on average €2.82 million sales in the sector of entry in the first year, with a very large dispersion around the mean. The same year, we find that they invest on average €960,000. Firms enter industries that are vertically integrated into their original industry in 67% of the cases and usually in sectors that are close to their sector of origin. They employ approximately 630 workers, produce approximately €50,000 of value added per worker, own €40,000 in fixed assets per worker, and hold approximately €20,000 in cash per worker.

Panel B of Table 2 compares the characteristics of firms that enter a new sector by acquisition with those that build on their preexisting internal human capital. The results show that firms that build have significantly smaller sales in the sector of entry than firms that buy, with €9.39 million less in sales on average in the year of entry. Building firms are also 13% less likely to stay in the sector of entry at a one-year horizon. They invest €9.95 million less in the year of entry and employ approximately 1,770 fewer workers on average. In addition, firms that diversify by acquisition are more profitable, more capital intensive and have higher internal funds. These significant differences in observable characteristics between firms that build and firms that buy emphasize the importance of including control variables in the empirical analysis.

[Insert Table 2 here]

5. Human capital and corporate diversification

5.1. Main results

Table 3 presents our main results. We test our model's prediction that firms are more likely to build when their human capital is more adapted to operate in the sector of entry (prediction 1). The dependent variable $1(Buy)_{g,n,t}$ is a dummy variable equal to one if firm f enters a new sector n at time t through the acquisition of an existing firm and zero if it enters by building on its own resources. The main independent variable *Internal Human Capital* measures the extent to which a firm already employs workers from key occupations for the sector of entry prior to diversification. All our regressions include interacted sector of origin $\times$ sector of entry $\times$ year fixed effects. This specification neutralizes potential unobservable time-varying synergies between the sector of entry and that of origin. The idea is to compare firms that operate in the same sector of origin and enter the same new sector in the same year to isolate the effect of human capital on firms' decision to enter through an acquisition or by building on their own

resources. All specifications control for firms' size (log number of workers), and we also control for firms' total cash holdings, tangible assets, and value added, with the latter variables being scaled by the number of workers.

Consistent with our model's prediction and Figure 2, we find that the internal human capital of the firm prior to entry is negatively correlated with the probability to enter by acquisition. A one-standard-deviation increase in internal human capital is associated with a 1 percentage-point decline in the likelihood of entering by acquisition (columns 1 and 2). This relationship is sizable, equal to 50% of the unconditional probability of buying, and significant at the 1% level. The point estimate is unchanged when we add control variables in column 2, suggesting that the control variables are uncorrelated with our key dependent variable *Internal Human Capital*. We conclude that firms possessing human capital adapted to the sector of entry are more likely to enter by building on their own resources. The estimates in Table 3 also show that the likelihood of buying relative to building increases with firm size. By contrast, cash holdings, tangibility and profitability do not appear to be significantly associated with the mode of entering a new sector.

An alternative explanation is that firms anticipate the mode of entry by adjusting the composition of their internal human capital several years before diversifying. If this were the case, human capital would be endogenous to the mode of entry, and our interpretation would be biased. To overcome this problem, we test whether lagged values of internal human capital also predict the mode of entry. We find that firms' decision to build rather than buy is still negatively correlated with internal capital two and three years before entry (columns 3 and 4, respectively). The point estimates on each lagged measure of human capital remain unchanged, suggesting that firms do not significantly modify their workforce composition during the years preceding entry.

Finally, to address the concern that our results could be confounded by firms having developed a certain expertise for a given type of entry irrespective of the composition of their internal human capital, we adopt a within-firm estimation in column 5. Again comparing firms that operate in the same sector of origin and enter the same new sector in the same year (i.e., with sector of origin $\times$ sector of entry $\times$ year fixed effects), we find that within firms, the magnitude of the relationship between internal human capital and the likelihood of entering by acquisition remains significant at the 5% level. However, the economic magnitude of this

relationship is only one-third of that in the across-firm, within-sector specification (column 1). Note that in the within-firms specification, the coefficient is identified on firms that perform several diversified entries during the sample period and switch their mode of entry. In addition, the coefficients of firm size and value added per worker in column 5 have a different sign than in the other columns. Our interpretation is that firms that diversify multiple times tend to buy at a stage of their lifecycle in which they are relatively smaller and less profitable.

To conclude, Table 3 supports our model's prediction 1; that is, firms choose to build rather than buy when their existing workforce is more adapted to operate into the sector of entry. According to our model, one rationale for this result is that hiring workers in key occupations is too costly because of tight LLMs. We discuss the role of labor market tightness in Section 8.1.

[Insert Table 3 here]

5.2. The role of size and financial constraints

Firms that buy tend to be larger (Table 2). One potential concern is that only large firms may be able to pay the fixed costs associated with an acquisition. Hence, small firms would always end up entering by building, and human capital considerations would be irrelevant. We should expect in that case no significant relationship between human capital and the type of entry for smaller firms. In columns 1 and 2 of Table 4, we interact our human capital measure with firm size. We use terciles of firm size (number of workers) to allow for non-linear effects of firm size on the decision to build or buy. We find that the interaction terms are not significantly different from zero, suggesting that human capital plays a role in the decision to build or buy for both small and large firms.

A potential explanation for why larger firms tend to grow more by acquisition is that they are less financially constrained than smaller firms. In columns 3 and 4, we test whether our measure of internal human capital interacts with internal financial resources as proxied by cash holdings. The results show that the relationship between internal human capital and the probability of building or buying does not differ across firms with different levels of cash earnings.

Finally, public firms may have different growth strategies than private firms because they

can raise equity to fund acquisitions. We test whether the relationship between internal human capital and the decision to build or buy differs for firms that are privately and publicly owned. We re-estimate our baseline specification separately on public and private firms in columns 5 and 6 of Table 4.[36] We find that the negative relationship holds both for public and private firms. Note that given the small number of public firms in France, we only perform those regressions with sector of origin × sector of entry interacted separately with fixed effects and year fixed effects. In summary, Table 4 confirms that the main findings hold across different types of firms.

[Insert Table 4 here]

5.3. The role of product market and physical distances

Firms do not diversify in random sectors but consider complementarities and synergies between the sectors of origin and entry.[37] Specifically, one may argue that firms choose to diversify by acquisition in more distant sectors from the sector of origin and may instead enter close sectors by building, irrespective of labor considerations. Geographical distance between the firm and the sector of entry can also play an important role, with physically more distant firms lacking the local resources to enter a sector by building on their existing resources. Instead, we expect firms are more likely to build with their existing workforce if they are physically close because it is less costly for them to encourage their current employees to travel and/or relocate.

First, we investigate the effect of product market distance on a firm's decision to build or buy. We interact internal human capital with a firm-level measure of the firm's distance to the sector of entry. This distance is based on the correlation between the product market portfolio of a firm prior to entry and product market portfolios of firms already operating in a sector (see Section 4.2.2 for details). A distance close to zero, for instance, means that the entering firm has very similar activities to incumbent firms.

In column 1, we use terciles of the distance, and in column 2, we interact them with the measure of internal human capital in column 2 to investigate how human capital interacts with

[36] In the Appendix, Table B7 interacts our covariates with a dummy equal to 1 when the target is publicly listed. Our main results on the role of internal human capital are unchanged, although we find public targets to be acquired relatively more when their ratio of fixed assets to the total number of workers is lower.

[37] Hoberg and Phillips (2010) show that product market distance plays an important role in determining M&A patterns. Firms that make similar products tend to merge more with one another, further enhancing value creation. The authors develop a measure based on textual analysis of 10k filings available for publicly listed firms in the US.

the product market distance between the sector of origin and the sector of entry. We find that firms tend to buy more frequently in more distant product markets (third tercile). The point estimate for human capital is very close to that in Table 3, which suggests that workforce composition is not explained by sectoral similarities. Interestingly, in column 2, we find that the internal human capital coefficient is stronger for firms in the 2^{nd} and 3^{rd} terciles of distance. This finding suggests that diversification in distant product markets amplifies the importance of labor costs in the decision to build or buy.

Second, in columns 3 and 4 of Table 5, we test the effect of vertical integration in the decision to build or buy. Labor cost considerations may be less relevant in the presence of important vertical links. Firms might be willing to enter upstream sectors to acquire suppliers and to facilitate transfers of goods along the production chain. To account for the firm's position in the production chain, we measure vertical integration following Fan and Goyal (2006). Using the 1995 input-output (IO) matrix for France compiled by the Organization for Economic Cooperation and Development (OECD), we identify vertical links between a firm and the sector of entry when more than 5% of the inputs used by the sector of origin come from the sector of entry (we also use 1%, 10% and 20% as thresholds).[38] Because the variable is defined at the level of a sector of origin and the sector of entry, we replace the interacted fixed effects with separate origin, entry and year fixed effects. We do not find any systematic relationship between the presence of vertical links and the decision to build or buy. In particular, firms that start operations in an upstream sector do not seem to enter more often by acquisition.

Third, we investigate the role of the physical distance between the firm and the market of entry. On the one hand, one could expect geographical diversification to induce firms to buy rather than build if tapping a new market requires a physical presence, e.g., to build a local customer capital or if a physical presence is a proxy for a better knowledge of the local market. On the other hand, if potential targets are geographically close, then it is cheaper to build to enter a geographically distant market. To test the effect of geographical distance, we construct a dummy variable, *New geographic zone*, that is equal to 1 if the firm diversifies in a new geographic area.[39] The results reveal that 4.2% of entries by acquisition occur in a new geographic area, and only 1.8% of entries occur by building. In column 5 of Table 5, we find

[38]Note that the IO matrix from France that we use is rather coarse (35 industries by 35 industries). The regression excludes observations for which the industry of origin and the industry of entry are the same because the vertical link variable is not defined in that case.

[39]In untabulated results, we find similar results when defining a geographic zone using departments instead of regions.

that an entry in a new geographic area is positively associated with the likelihood of diversifying by acquisition. Hence, physically distant firms are more likely to buy, whereas firms that stay in the same area are more likely to build on their existing workforce. Moreover, we do not find the economic or statistical significance of our internal human capital measure to be affected when we control for entry in a new geographic area. The results in column 6 indicate that the interaction of physical distance with human capital is not significantly different from zero. Overall, the evidence suggests that the role of physical distance in a firm's decision to build or buy is distinct from that of internal human capital.

[Insert Table 5 here]

6. Robustness checks and alternative mechanisms

6.1. Alternative measures of human capital

In this section, we check the robustness of our results to alternative definitions of the measure of internal human capital. To do so, we replicate the main specification in Table 3 using different versions of *Internal human capital*. Table 6 reports the results. In column 1, the main independent variable is a dummy variable taking value one if the firm does not employ any worker in the top-10 most important occupations for the sector of entry. The ranking of occupations within sectors is based on the estimated values of the occupation $\times$ sector fixed effects (see Section 3). We find that firms that do not hire any workers in the top-10 occupations for the sector of entry are 40% more likely to buy (column 1).

In column 2 of Table 6, we take the weighted average of the fixed effects with weights equal to the share of workers in a given occupation in the workforce of the firm. This alternative measure of internal human capital assumes that the firm allocates workers to the new sector in proportion to the existing occupational structure (see Section 3.2). In column 3, we exclude CEOs from the set of occupations used to build the measure of internal human capital. One concern with CEOs' wages is that they may be determined by factors other than their contribution to the firms' performance, for instance, moral hazard or information asymmetries.

Finally, in column 4, we change the unit of observation used in the estimation of the occupation $\times$ sector $\times$ year fixed effects from firms to plants. Plants are assumed to be less

diversified entities than firms; thus, estimating the fixed effects at the plant level should yield more precise estimates. Indeed, when we estimate fixed effects at the firm level, we make the implicit assumption that every worker is involved in the production process of the firm's main activity.

Our main result is not affected by any of these alternative measures of internal human capital: The negative relationship between human capital and the probability to buy remains economically and statistically significant. Moreover, the point estimates are very similar across specifications.

[Insert Table 6 here]

6.2. Does selection into diversification drive our results?

A potential limitation of our analysis is that we focus on the build versus buy trade-off without eliciting the decision to diversify in the first place. If the entry and the type of entry in the new sector are jointly determined and driven by unobservable factors, the OLS estimates may be biased.

Our approach to address this issue is to identify plausible scenarios in which self-selection could invalidate our results.[40] First, firms with under-performing business segments may be willing to shift their activities to sectors with better prospects. Thus, firms are likely to choose sectors in which they can easily redeploy their existing workforce. In this scenario, the likelihood of an entry by acquisition would also be negatively related to internal human capital.

We test the influence of sectoral reallocation in Table 7. We identify "shifting firms" as those firms that enter a new sector while using internal resources to shift a substantial part of their activities. Under the tested hypothesis, a negative coefficient for human capital on the likelihood to enter by building would be driven by the sub-sample of shifting firms. Excluding them should result in a non-significant coefficient for human capital. In contrast, the results in columns (1) to (3) show that the point estimates for our main variable remain unchanged.

[40] A classic solution to address selection issues in diversifying choices is to find an instrumental variable that affects the choice of sector but is plausibly orthogonal to the main dependent variable. For instance, Tate and Yang (2016*b*) uses Tobin's Q as an instrument in a two-stage Heckman selection model. Applied to our context, we would need an instrument that affects the choice to diversify but not the type of entry. In addition, this instrument would need to be sector-specific because we would have to instrument not only for the choice to build or buy but also for the decision to diversify in a given sector. Finding such an instrument appears to be a difficult task; therefore, we choose to focus on specific scenarios in which such selection issues are likely to arise.

Consider then the other polar case. Firms with specific expertise or organizational capital for acquisition may be willing to always diversify by acquisition irrespective of internal human capital considerations. Since firms are unlikely to employ the right set of workers to enter a new sector if they never expect to build using their existing resources, this could translate into an observed negative relationship between internal human capital and the propensity to buy. To address this particular issue, in column (4) of Table 7, we focus on firms that enter multiple sectors by acquisition ("serial acquirers").[41] We still find a significant, negative coefficient for human capital when excluding those serial acquirers. Overall, these findings about serial acquirers and firms likely to shift their actives to sectors with better prospects suggest a limited role for selection issues.

[Insert Table 7 here]

6.3. The role of scale and physical capital

We investigate the issue of the scale of the new activity in the sector of entry. If the entry is small, existing workforce slack can be used for it conditional on the workforce having the right skills. It is also the case that if the entry is small, the firm does not have to hire many workers. When size is large, it becomes much more difficult to enter the new sector by building on existing resources because it requires workers who can function together, so hiring one worker at a time can be inefficient and time-consuming. Another related issue is the role of physical capital adjustment (e.g., equipment or machinery), which may act as a confounding factor in the relationship between human capital and firms' decision to build or buy.

First, to investigate the role of the scale of the new activity, we compare the build or buy decision of firms that realized similar entry sales in the year of entry. As firms should enter a new sector only if they anticipate high enough entry sales to offset entry costs, we use entry sales to proxy for entry costs. In columns 1 and 2 of Table 8, we rank entry sales into 10 deciles and run our baseline regressions with interacted sector of origin $\times$ sector of entry $\times$ year $\times$ sales decile fixed effects. This specification allows us to compare firms that operate in the same sector of origin, enter the same sector in the same year, and make similar sales when entering the new sector. Although this specification creates many singletons that are dropped from the sample (the number of observations drops from 75k to 45k), our point estimates remain nearly

[41] Among firms that enter a sector using external resources at least once, 50% are classified as serial acquirers.

unchanged.

Second and similarly, we proxy for physical capital adjustments using the volume of capital expenditures made by firms when entering the new sector. To do so, we compare firms investing similar amounts in the year of entry. For firms that build, we measure investment using capital expenditures in the year of entry. For firms that buy, we measure investment as the amount of fixed assets in the target. We run our baseline regression with the interacted sector of origin $\times$ sector of entry $\times$ year $\times$ investment decile fixed effects. The idea is to isolate the effect of human capital on the build or buy decision, irrespective of differences in capital expenditures. In columns 3 and 4 of Table 8, we find that the economic magnitude of the role of human capital is reduced by approximately half. It remains negative and statistically significant at the 5% level, with or without the inclusion of control variables. Based on the results in Table 8, we conclude that human capital matters in firms' decision to build or buy, irrespective of the scale of the new activity.

[Insert Table 8 here]

7. Evidence of labor adjustment costs

The previous sections established that firms that do not employ the right set of workers tend to enter a new sector by acquisition. This finding, we argue, suggests that firms prefer to pay the costs associated with acquiring and restructuring a target when the costs of adjusting the existing workforce are large. In this section, we focus on firms that diversify by building on their existing resources to highlight the existence of such adjustment costs.

7.1. Within-firm human capital and diversification choice

We define a firm as a set of subsidiaries. Therefore, the composition of the workforce may vary within a firm, with some subsidiaries being better prepared to enter the new sector than others. Within firms that build, we should expect the entry to be made through subsidiaries that already employ the right set of occupations to minimize reallocation costs in the internal labor market. We construct our measure of internal human capital for each subsidiary and

estimate the following model:

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

The dependent variable is a dummy variable that takes value one if the entry into the new sector is made through subsidiary f, zero otherwise. We include firm $\times$ sector of entry $\times$ year fixed effects to compare the internal capital of the different subsidiaries of the same firm g. This specification leads mechanically to the exclusion of stand-alone firms because we focus on the heterogeneity of human capital across subsidiaries.

In column 1 of Table 9, we find that within a firm, entry is more likely to be achieved through a subsidiary with the appropriate set of occupations for the new sector. In column 2, we add control variables at the subsidiary level. The estimates show that large, productive and cash-rich subsidiaries are more likely to diversify in a new sector. In addition, in columns 3 and 4, we show that lagged values of internal human capital are also positively correlated with the entry dummy. As for firms, internal human capital at the subsidiary level is sticky over time, suggesting that the composition of the workforce is not adjusted in anticipation of diversification. Overall, these findings are consistent with the presence of reallocation costs in the internal labor market.

[Insert Table 9 here]

7.2. Internal human capital and workforce adjustment

Our measure of human capital captures the extent to which the firm's workforce is adapted to the sector of entry. Therefore, among firms that enter a new sector by developing their own resources, we should find that firms with lower human capital hire relatively more workers to adjust their workforce. We test this hypothesis by examining employment growth within subsidiaries. Precisely, we focus on the subsidiaries that, within diversifying firms, begin selling in the new sector. This choice allows us to observe more precisely the adjustment in labor associated with entry because it allows us to abstract from employment variations in the other subsidiaries.

In Table 10, we demonstrate the existence of a negative relationship between internal human capital at $t-1$ and the growth rate of employment between $t-1$ and $t+1$ (with t being the year

of entry into the new sector). In column 1, we find that the higher the internal human capital is, the fewer additional workers the subsidiary hires to operate in the new sector. In columns 2 and 3, we examine the timing of new hiring. We find that subsidiaries with lower internal human capital do not hire more workers prior to entry (column 2). Instead, subsidiaries tend to adjust their workforce after having entered the new sector (column 3). In column 4, we show that the newly hired workers are relatively more adapted to the sector of entry because they imply an increase in firms' human capital. Our interpretation is that when entering a new sector by developing their own resources, firms adjust their internal resources for the new sector.

[Insert Table 10 here]

8. The role of labor market frictions

8.1. The effects of local labor market tightness for key occupations

Given the costs associated with an acquisition, why do firms not hire new workers instead of buying an existing firm? The model predicts that firms are more likely to buy than to build on their existing workforce when workers in key occupations are in short supply in the external labor market (prediction 2). Specifically, the negative relationship between internal human capital and the likelihood of buying should be stronger when it is costly for firms to hire workers in the external labor market. In this section, we empirically assess the role of LLM tightness in the decision to build or buy using the *Local Labor Tightness* variable described in 4.2.3.

Columns 1 and 2 of Table 11 show that the point estimates on the second and third terciles of LLM tightness are positively related to the decision to buy. Thus, firms are significantly more likely to buy when labor markets for key occupations are tighter. Moreover, the interaction terms show that the link between internal capital and the type of entry is stronger in the presence of greater market frictions.

These findings hold both with and without control variables (columns 1 and 2). We confirm this finding when we divide the sample by tercile of LLM tightness and re-estimate the regression separately on each sub-sample: The coefficient on internal human capital increases in absolute value and is significant at 10% only for the last tercile of LLM tightness. Overall, the finding

is consistent with the prediction that human capital determines firms' decision to build or buy and is especially acute in the presence of labor market frictions.

[Insert Table 11 here]

To check the robustness of this result, we conduct the same analysis using an alternative measure of LLM tightness. We use the definition in Equation (34) in Section 4.2.3, based on the percentage of occupations in short supply. Table B6 presents the results, which are very similar to our earlier findings: Firms buy more when LLMs are tight, and the role of human capital is stronger when LLMs are tight. When using the interacted sector of origin × sector of entry × (geographical) LLM fixed effects, the results still hold but only significantly so in the tightest LLMs due to a lack of statistical power.

8.2. Local labor market tightness and the value of building

According to our model, firms' internal resources determine expected profits from entering a new sector by building versus buying. Firms choose to build when their workforce is adapted to the sector of entry, which is more profitable than buying an existing company. Therefore, our next question is whether firms that enter by building with higher internal human capital create value, as implied by our model. Are firms that do not and face greater hiring costs less productive in the short run? Based on predictions 1 and 2, we should expect that the profits firms can generate from building are positively related to internal human capital. Moreover, this relationship should be stronger when key occupations are in short supply in the external labor market. In this last section, we directly examine the relationship between the value of building and the human capital of the firm.

We proxy for the value of building with (the logarithm of) entry sales after the entry. In columns 1 and 2 of Table 12, we find that entry sales are larger when the workforce of the firm is more adapted to the sector of entry. This finding suggests that the value of building is higher when the firm has adequate internal resources. In columns 3 to 5, we run the analysis on a subsample of LLM tightness terciles and show that the positive link is entirely driven by tight labor markets (3rd tercile, column 5). This finding suggests that the value of building depends on firms' internal human capital when LLM frictions are important. In other words, firms' internal resources matter only when it is costly to obtain them outside the firm.

[Insert Table 12 here]

8.3. Build or buy and firing costs

We test whether firing costs affect firms' decision to build or buy. According to our model, when a firm buys to enter a new sector, it has to incur the cost of restructuring the target (see Appendix A.2). Therefore, our first hypothesis is that the higher the restructuring cost is, the less attractive the option to buy. Furthermore, the more adapted an acquiring firm's internal human capital, the more workers must be laid off after the acquisition because of the greater overlap of key worker occupations; therefore, the higher the restructuring cost will be after the acquisition. Therefore, our second hypothesis is that a higher firing cost makes the option to buy less attractive for firms with adapted internal human capital.

We perform two tests of these hypotheses. First, we use the fraction of permanent to temporary workers in a firm's workforce to proxy for firing costs, hence the cost at with which firms can restructure their workforce. The greater the fraction of permanent workers in a firm's workforce, the less latitude a firm has to replace temporary workers with workers who are better adapted to the sector of entry. Furthermore, in France, laying-off a permanent worker involves higher reparation costs in the case of prejudice (Fraisse, Kramarz and Prost, 2015). Therefore, expected firing costs increase in the fraction of permanent workers in the workforce.

In columns 1 and 2 of Table 13, we test whether the fraction of permanent employees in a firm's workforce is correlated with firms' decision to build or buy. In line with our first hypothesis, we find a positive relationship between the fraction of permanent workers in a firm's workforce and the decision to build, but the coefficient is not significantly different from zero. However, in line with our second hypothesis, this fraction interacts significantly with our measure of internal human capital (column 2), implying that a higher firing cost increases the importance of internal human capital in firms' decision to build or buy.

One potential problem with this first proxy for firing costs is that firms decide whether they want to hire workers under permanent or temporary contracts. Therefore, in a second set of tests, we use the geographic variation in the average length of local labor case settlements across jurisdictions as a proxy for firing costs that is exogenous to firms (see Sections 4.1.4 and 4.2.3). Higher firing costs seem to be associated with a higher probability of buying, but the coefficient is again not significant. In line with our second hypothesis, we find that internal human capital

matters significantly more in firms' decision to build or buy when the firing cost is high (column 4). Overall, the results in Table 13 show that firing costs amplify the importance of internal human capital in firms' decision to build or buy.

[Insert Table 13 here]

9. Conclusion

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? When a firm buys to enter a new sector, it has to incur both the costs of acquiring and restructuring the target, but it also secures access to the target's productive resources. When a firm builds on its existing resources to enter a new sector, it must pay the adjustment costs needed to acquire an adapted set of capabilities.

We focus on the role of labor and construct a firm-level measure of human capital based on the occupational structure of the workforce. Our main explanatory variable measures the extent to which the firm's internal human capital is adapted to the sector of entry.

We show that the vast majority of entries in a new sector consist of firms that build on their internal resources and that firms choose to buy an existing company when their human capital is not adapted to the sector of entry.

We find evidence that labor adjustment costs contribute to the importance of internal human capital in firms' choice to build or buy. On the one hand, firms are more likely to buy when it is costly to hire key workers on the external job market, i.e., when key worker occupations are in short supply. On the other hand, higher firing costs make the option to buy less attractive for firms with adapted internal human capital.

Our findings imply fundamental factors are driving firms' decision to build or buy. They are consistent with firms choosing organizational structures that best deploy the economy's pool of specialized resources (Maksimovic and Phillips, 2002). Thereby, they contribute to the literature on corporate diversification by showing that both the set of internal resources and the cost of accessing external resources play a role in explaining how firms diversify.

References

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

Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High wage workers and high wage firms." *Econometrica*, 67(2): 251–333.

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

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. 2018. "Talent in distressed firms: Investigating the labor costs of financial distress." *Working Paper*.

Bai, John Jianqiu, Douglas J Fairhurst, and Matthew Serfling. 2018. "Employment protection, investment, and firm growth." Working Paper.

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, Emily J Blanchard, Ilke Van Beveren, and Hylke Vandenbussche. 2018. "Carry-along trade." *The Review of Economic Studies*, 86(2): 526–563.

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott. 2007. "Firms in international trade." *Journal of Economic perspectives*, 21(3): 105–130.

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." *Working Paper*.

Chandler, Alfred D. 1992. "Organizational capabilities and the economic history of the industrial enterprise." *Journal of Economic Perspectives*, 6(3): 79–100.

Chen, Deqiu, Huasheng Gao, and Yujing Ma. 2018. "Human capital driven acquisition: Evidence from the inevitable disclosure doctrine." *Working Paper*.

Cheng, Wenya, and John Morrow. 2018. "Firm productivity differences from factor markets." *Journal of Industrial Economics*, 66(1): 126–171.

Davis, Steven J, and John Haltiwanger. 1992. "Gross job creation, gross job destruction, and employment reallocation." *Quarterly Journal of Economics*, 107(3): 819–863.

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

Eckbo, Espen. 1983. "Horizontal mergers, collusion, and stockholder wealth." *Journal of financial Economics*, 11(1-4): 241–273.

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

Farrell, Joseph, and Carl Shapiro. 1990. "Horizontal mergers: an equilibrium analysis." *American Economic Review*, 107–126.

Fraisse, Henri, Francis Kramarz, and Corinne Prost. 2015. "Labor disputes and job flows." *ILR Review*, 68(5): 1043–1077.

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

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

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

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

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

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

Lagaras, Spyridon. 2017. "Corporate takeovers and labor restructuring." *Working Paper*.

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, Feng Mai, Rui Shen, and Xinyan Yan. 2018. "Measuring corporate culture using machine learning." Working Paper.

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

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

Maksimovic, Vojislav, Gordon Phillips, and Nagpurnanand R Prabhala. 2011. "Post-merger restructuring and the boundaries of the firm." *Journal of Financial Economics*, 102(2): 317–343.

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.

Ma, Wenting, Paige Ouimet, and Elena Simintzi. 2016. "Mergers and acquisitions, technological change and inequality." *Working Paper*.

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

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

Moretti, Enrico. 2010. "Local labor markets." *National Bureau of Economic Research*.

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 Parker, and Rebecca Zarutskie. 2016. "Acquiring labor." Working Paper.

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

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

Rajan, Raghuram, and Luigi Zingales. 1998. "Financial Dependence and Growth." *American Economic Review*, 88(3): 559–86.

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.

Romanski, Hilton. 2017. "Cisco's 200th Acquisition - A Tradition of Advancement, Disruption and Growth." *Cisco Blog*.

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

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

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

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

Wysocki, Bernard Jr. 1997. "Many mergers driven by search for fresh talent." *Wall Street Journal*.

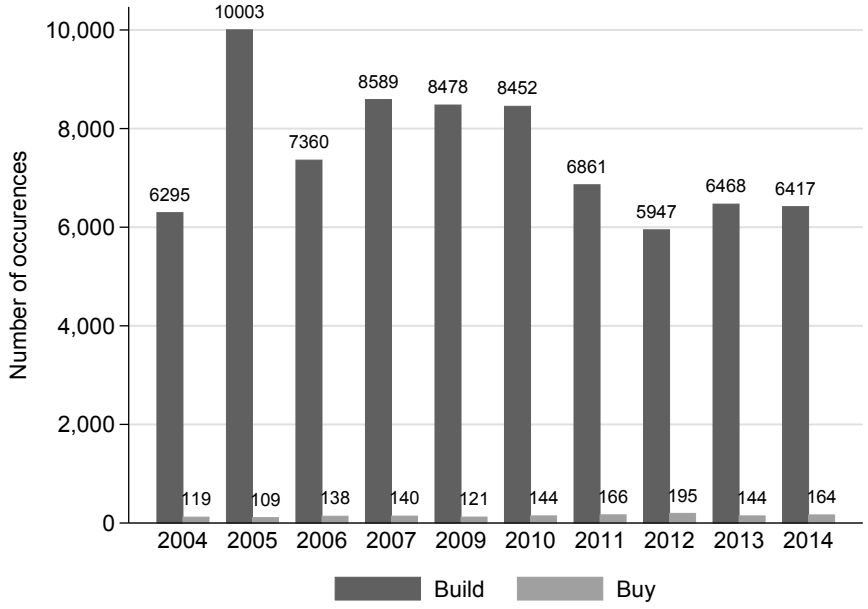
Xu, Sheng-Jun. 2018. "Skilled labor supply and corporate investment: Evidence from the H-1B visa program." *Working Paper*.

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

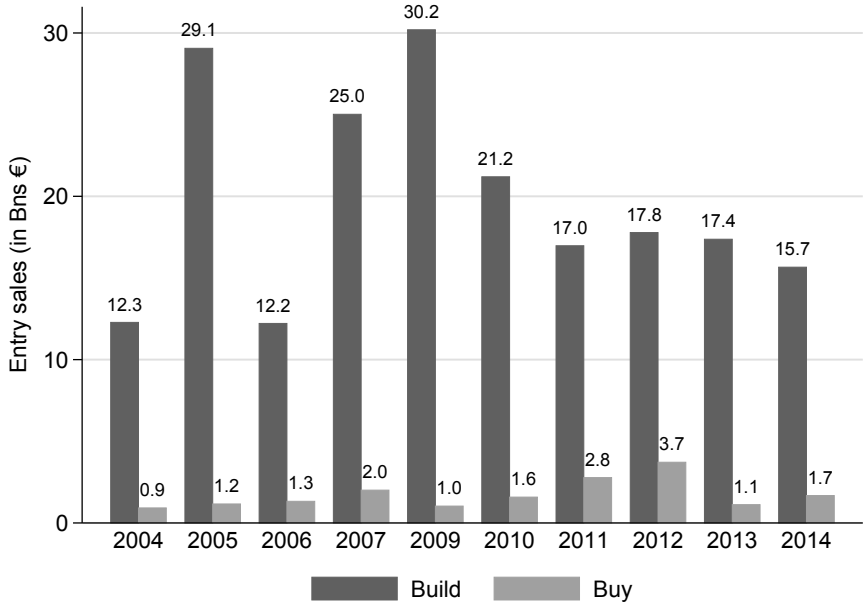
Figures and Tables

Figure 1. Number and Size of Build and Buy Entries

Source: SDC Platinum, BvD Zephyr, ESA survey. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This figure displays the number of external (Buy) and internal (Build) entries by year and the aggregate sales by type of entry. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Entries are identified with sales reported at the 5-digit level of the French SIC.



(a) Number of entries



(b) Aggregate sales by type of entry

Figure 2. Human Capital by Type of Entry

The figure displays the probability distribution function of *Internal Human capital* by mode of entry. *Source:* SDC Platinum, BvD Zephyr, EAE survey, matched employer-employee dataset. *Sample:* Firms that enter a new sector either internally (build) or externally (buy) during the periods 2003-2007 and 2009-2014. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Build entries are identified using reported sales from the ESA survey at the 5-digit level of the French SIC. *Internal Human Capital*$_{g,n,t-1}$ measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). *Internal Human Capital*$_{g,n,t-1}$ is trimmed at the 5% level (for this graph only).

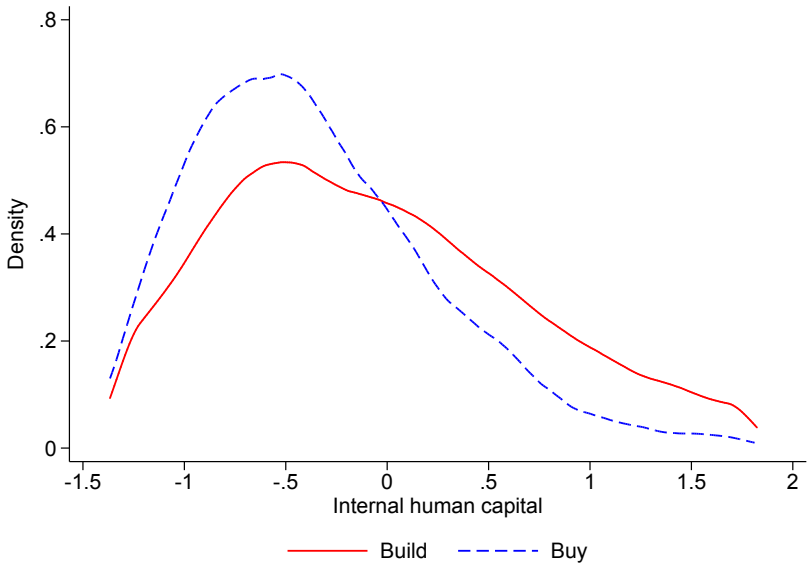
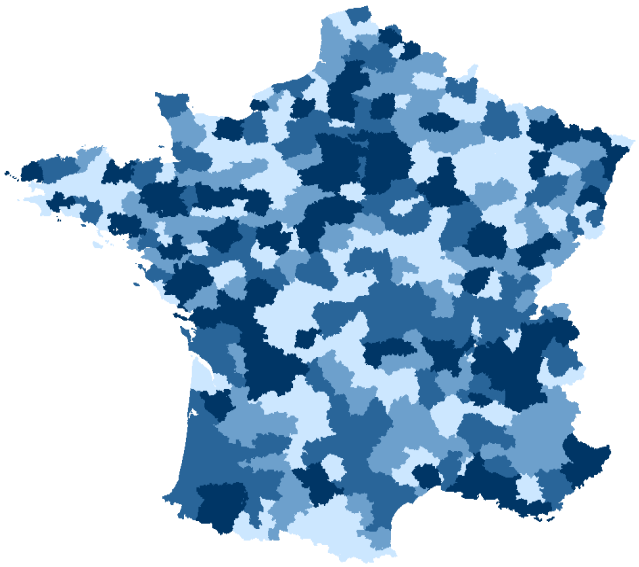


Figure 3. Occupations in Short Supply

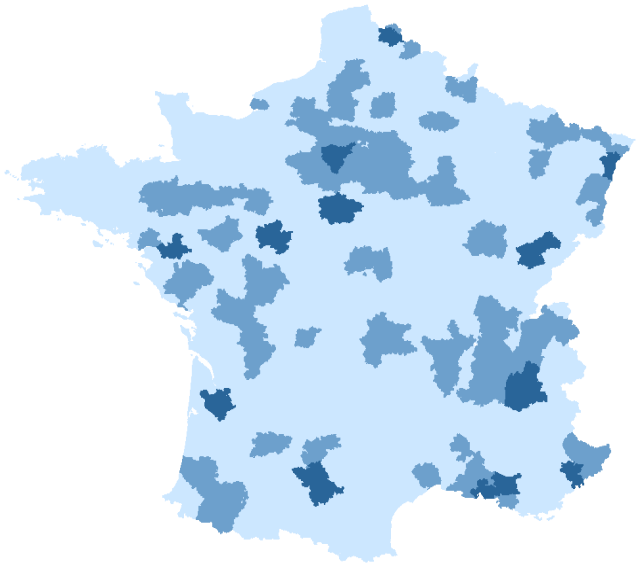
Source: French national unemployment agency. Figure 3 plots the distribution of the tightness of labor markets in 2013. Labor market tightness is measured by the number of occupations in short supply in a given local labor market. Darker shades of blue indicate a higher degree of tension in the local labor market. Occupations in short supply are identified by the French national employment agency as occupations for which (i) the ratio of job offers over job applications is high and (ii) surveyed employers forecast that it will be difficult to fill posted offers. There are 348 different local labor markets.



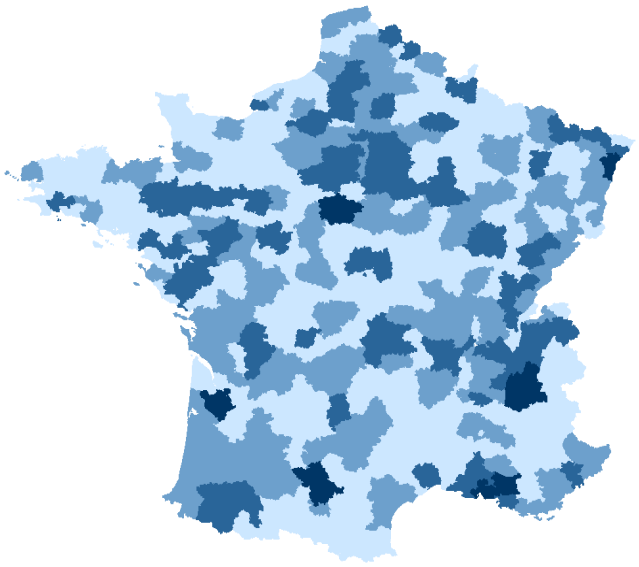
(a) Number of occupations in short supply by local labor market.

Figure 4. Local Labor Markets Tightness

Source: French national unemployment agency. The figures plot the value of $LLM\ Tightness_{n,z,t}$ of each local labor market z in $t = 2013$ for different sectors n. Figure 4a focuses on the manufacture of pharmaceutical preparations, and figure 4b focuses on the manufacture of motor vehicles. Darker shades of blue indicate local labor markets with a larger number of occupations in short supply. Occupations in short supply are identified by the French national employment 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. France is divided into 348 different local labor markets z.



(a) Manufacture of pharmaceutical preparations



(b) Manufacture of motor vehicles

Table 1. Evolution of the Numbers of Build and Buy Entries

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This table reports the ratio of Buy entries to Build entries. A firm is said to "buy" when it enters a new sector through an M&A (source: SDC Platinum and Bureau van Dijk Zephyr). A firm is said to "build" when it enters a new sector through one of its existing subsidiaries (source: ESA survey).

Panel A. Buy and Build entries by year

	2004	2005	2006	2007	2009	2010	2011	2012	2013	2014
Build (number)	6295	10003	7360	8589	8478	8452	6861	5947	6468	6417
Buy (number)	119	109	138	140	121	144	166	195	144	164
Buy (%, frequency)	1.85	1.08	1.84	1.60	1.4	1.67	2.36	3.17	2.18	2.49
Buy (%, sales)	6.98	3.85	9.79	7.44	3.31	6.95	14.07	17.30	6.06	9.68

Panel B. Buy and Build entries for varying definitions of sector

Industry level:	5 digits	4 digits	3 digits	2 digits	1 digit
Build (number)	74,870	72,250	63,176	49,570	32,431
Buy (number)	1,440	1,363	1,163	941	584
Buy (%, frequency)	1.89	1.85	1.81	1.86	1.77
Buy (%, sales)	8.05	8.55	8.31	10.57	15.72

Table 2. Summary Statistics

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This table reports descriptive statistics for firms that are identified as entering a new sector, either internally or externally, during the periods 2003-2007 and 2009-2014. Panel A reports the distribution of the main firm characteristics. Panel B compares the mean characteristics of firms that enter a new sector internally and those that enter by acquisition. A firm is said to "buy" when it enters a new sector through an M&A (source: SDC Platinum and Bureau van Dijk Zephyr). A firm is said to "build" when it enters a new sector through one of its existing subsidiaries (source: ESA survey). Sectors refer to an industry at the 5-digit level of the French SIC. Descriptions of the variables are reported in Appendix C.

Panel A. Distribution of firm characteristics

	N	Mean	St.Dev.	Percentiles				
				5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
$\mathbb{1}(\text{Buy})_{g,n,t}$	76310	0.02	0.14	0.00	0.00	0.00	0.00	0.00
Entry $\text{sales}_{g,n,t}$ (M euros)	76310	2.82	21.32	0.00	0.05	0.27	1.21	8.95
$\mathbb{1}(\text{1-year survival})_{g,n,t+1}$	76310	0.47	0.50	0.00	0.00	0.00	1.00	1.00
$\text{Investment}_{g,n,t}$ (M euros)	76310	0.96	16.24	0.00	0.03	0.10	0.35	2.44
Internal $\text{Human Capital}_{g,n,t\text{-}1}$	76310	0.14	1.04	-1.29	-0.59	-0.00	0.69	2.05
$\#\text{workers}_{g,t\text{-}1}$ (in thousands)	76310	0.63	5.16	0.02	0.04	0.07	0.19	1.29
Value added/$\text{N.workers}_{g,t\text{-}1}$	76310	0.05	0.03	0.02	0.03	0.04	0.06	0.11
Fixed assets/$\text{N.workers}_{g,t\text{-}1}$	76310	0.04	0.05	0.00	0.01	0.03	0.06	0.14
Cash holdings/$\text{N.workers}_{g,t\text{-}1}$	76310	0.02	0.02	0.00	0.00	0.01	0.02	0.06
Product $\text{market distance}_{g,n,t\text{-}1}$	76310	0.87	0.15	0.53	0.81	0.94	0.98	1.00
Vertical $\text{integration}_{g,n,t\text{-}1}$	44947	0.67	0.47	0.00	0.00	1.00	1.00	1.00

Panel B. Comparison of build and buy entries

	Build		Buy		Difference	
	Mean	St.Dev.	Mean	St.Dev.	Mean	p-value
Entry $\text{sales}_{g,n,t}$ (M euros)	2.64	20.52	12.04	45.85	-9.39^{***}	(0.00)
$\mathbb{1}(\text{1-year survival})_{g,n,t+1}$	0.47	0.50	0.59	0.49	-0.13^{***}	(0.00)
$\text{Investment}_{g,n,t}$ (M euros)	0.78	15.35	10.72	40.43	-9.95^{***}	(0.00)
Internal $\text{Human Capital}_{g,n,t\text{-}1}$	0.15	1.04	-0.21	0.77	0.36^{***}	(0.00)
$\text{N.workers}_{g,t\text{-}1}$ (in thousands)	0.59	4.82	2.37	14.33	-1.77^{***}	(0.00)
Value added/$\text{N.workers}_{g,t\text{-}1}$	0.05	0.03	0.07	0.04	-0.01^{***}	(0.00)
Fixed assets/$\text{N.workers}_{g,t\text{-}1}$	0.04	0.05	0.05	0.06	-0.00	(0.16)
Cash holdings/$\text{N.workers}_{g,t\text{-}1}$	0.02	0.02	0.02	0.02	-0.01^{***}	(0.00)
Product $\text{market distance}_{g,n,t\text{-}1}$	0.87	0.16	0.92	0.10	-0.05^{***}	(0.00)
Vertical $\text{link}_{g,n,t\text{-}1}$	0.68	0.47	0.59	0.49	0.08^{***}	(0.00)
Observations	74,870		1,440		76,310	

Table 3. Human Capital and Corporate Diversification

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and analyzes the effect of human capital on the type of diversification strategy. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). Control variables include the number of workers in logarithms as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(Buy)_{g,n,t}$				
	(1)	(2)	(3)	(4)	(5)
Internal HC$_{g,n,t-1}$	-0.015***	-0.010***			-0.005**
	(0.004)	(0.003)			(0.002)
log(#workers)$_{g,t-1}$		0.010***	0.010***	0.011***	-0.045***
		(0.002)	(0.002)	(0.002)	(0.013)
Cash holdings/#workers$_{g,t-1}$		0.044	0.043	0.054	-0.038
		(0.045)	(0.050)	(0.058)	(0.145)
Fixed assets/#workers$_{g,t-1}$		-0.021	-0.024	-0.023	-0.180*
		(0.028)	(0.031)	(0.034)	(0.106)
Value added/#workers$_{g,t-1}$		0.049	0.046	0.038	-0.295***
		(0.035)	(0.038)	(0.043)	(0.090)
Internal HC$_{g,n,t-2}$			-0.011***		
			(0.003)		
Internal HC$_{g,n,t-3}$				-0.014***	
				(0.004)	
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	Yes
R^2	0.199	0.206	0.208	0.207	0.564
Observations	76354	76296	66145	54230	57923

Table 4. The Role of Size and Financial Constraints

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of firms' financial constraints on the relationship between human capital and the type of diversification strategy. Three proxies for financial constraints are considered: size, cash holdings and public ownership status. Columns (1) and (2) include the second and third terciles of size (number of workers) interacted with *Internal Human Capital*$_{g,n,t-1}$. Columns (3) and (4) include the second and third terciles of cash holdings over workers (*Cash holdings/N. workers*) interacted with *Internal Human Capital*$_{g,n,t-1}$. In columns (5) and (6), we split the firms into publicly and privately owned firms. Public firms are those that include at least one publicly listed subsidiary within the firm. The dependent variable $\mathbb{1}(Buy)_{g,n,t}$ is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. Entries are identified at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). Control variables include the number of workers in logarithms as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(Buy)_{g,n,t}$					
	Size		Cash/N. workers		Type of firm	
					Public	Private
	(1)	(2)	(3)	(4)	(5)	(6)
Internal HC$_{g,n,t-1}$	-0.011***	-0.011***	-0.014***	-0.010***	-0.123**	-0.007***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.051)	(0.002)
2nd tercile of #workers$_{g,t-1}$	0.005**	0.005**				
	(0.002)	(0.002)				
3rd tercile of #workers$_{g,t-1}$	0.025***	0.025***				
	(0.005)	(0.006)				
2nd t. # workers$_{g,t-1}$× Int. HC$_{g,n,t-1}$	-0.001	-0.001				
	(0.001)	(0.001)				
3rd t. # workers$_{g,t-1}$× Int. HC$_{g,n,t-1}$	-0.004	-0.004				
	(0.003)	(0.003)				
2nd tercile of Cash$_{g,t-1}$			0.002	0.001		
			(0.002)	(0.001)		
3rd tercile of Cash$_{g,t-1}$			0.004**	0.003		
			(0.002)	(0.002)		
2nd t. Cash$_{g,t-1}$× Int. HC$_{g,n,t-1}$			-0.000	0.000		
			(0.001)	(0.001)		
3rd t. Cash$_{g,t-1}$× Int. HC$_{g,n,t-1}$			-0.003*	-0.002		
			(0.002)	(0.002)		
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	No	No
Origin-Entry FE	No	No	No	No	Yes	Yes
Year FE	No	No	No	No	Yes	Yes
Controls	No	Yes	No	No	Yes	Yes
R^2	0.204	0.205	0.199	0.206	0.107	0.123
Observations	76354	76296	76354	76354	1198	74204

Table 5. The Role of Sectoral and Geographical Distance

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of complementarities between the sector of origin and sector of entry on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). In addition, columns (1) and (2) include a product market distance variable adapted from Bloom, Schankerman and Van Reenen (2013). It measures the distance between the sectors in which firm g operates at $t-1$ and the sector of entry n. The distance ranges from 0 to 1 (1 being the maximum). Columns (3) and (4) include Fan and Goyal (2006)'s measure of vertical relatedness. It measures the intensity of vertical links between the main sector of activity of firm g at time $t-1$ and the sector in which g enters at time t. "Vertical" is a dummy variable that takes value 1 if the vertical relatedness exceeds 5%. Columns (5) and (6) include a dummy that indicates whether the firm enters a new geographical zone at time t. We use "region" as definition of the geographical zone (France is divided into 25 regions over the sample period). A firm enters a new geographical zone if (i) the entering subsidiary is created at time t and located in a new region, (ii) the entering subsidiary opens a plant in a new region at time t, or (iii) the entering subsidiary is acquired at time t and operates in a region in which the firm was not present at $t-1$. Control variables include the number of workers in logarithms as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$					
	Product market distance		Vertical integration		Geographical distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Internal HC$_{g,n,t-1}$	-0.010***	-0.006***	-0.019***	-0.018***	-0.015***	-0.010***
	(0.003)	(0.002)	(0.005)	(0.005)	(0.004)	(0.003)
2nd tercile of Distance$_{g,n,t-1}$	0.003	0.007**				
	(0.003)	(0.003)				
3rd tercile of Distance$_{g,n,t-1}$	0.009**	0.012***				
	(0.004)	(0.004)				
2rd t. Distance$_{g,n,t-1}$*Int. HC$_{g,n,t-1}$		-0.007***				
		(0.002)				
3rd t. Distance$_{g,n,t-1}$*Int. HC$_{g,n,t-1}$		-0.008***				
		(0.002)				
Vertical$_{g,n,t-1}$			0.001	0.000		
			(0.006)	(0.006)		
Vertical$_{g,n,t-1}$*Int. HC$_{g,n,t-1}$				-0.002		
				(0.005)		
New geographic zone$_{g,n,t}$					0.013*	0.010
					(0.007)	(0.008)
New geo. zone$_{g,n,t}$*Int. HC$_{g,n,t-1}$						-0.006
						(0.007)
Origin-Entry-Year FE	Yes	Yes	No	No	Yes	Yes
Origin FE	No	No	Yes	Yes	No	No
Entry FE	No	No	Yes	Yes	No	No
Year FE	No	No	Yes	Yes	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.208	0.208	0.102	0.102	0.201	0.208
Observations	76226	76226	68531	68531	76223	76223

Table 6. Alternative Measures of Human Capital

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and assesses the robustness of our results to the definition of human capital. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. Entries are identified at the 5-digit level of the French SIC. We test several alternative definitions of our main independent variable *Internal Human Capital*$_{g,n,t-1}$. In the baseline definition, the measure is computed as the sum of occupation × sector × year fixed effects present in the workforce of firm g at time $t-1$. The measure is scaled by the number of occupations in firm g. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). In column (1), the dependent variable is a dummy equal to one if for a sector of entry n and a time of entry t, firm g has no occupation in the top 10 of occupation-sector-year fixed effects. In column (2), we use the sum of occupation × sector × year fixed effects is weighted by the number of employees in each occupation in the firm. In column (3), CEO occupations are excluded from the sum of occupations. In column (4), occupation × sector × year fixed effects are estimated at the plant level instead of the firm level as in the baseline model. Control variables include the number of workers in logarithms as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{No Top 10})_{g,n,t-1}$	0.008***			
	(0.003)			
Internal $\text{HC}_{g,n,t-1}$ (weighted)		-0.009***		
		(0.002)		
Internal $\text{HC}_{g,n,t-1}$ (no CEO)			-0.010***	
			(0.003)	
Internal $\text{HC}_{g,n,t-1}$ (plant-level)				-0.009***
				(0.003)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R^2	0.205	0.206	0.205	0.206
Observations	76296	75614	76270	76094

Table 7. Selection into Diversification

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests two potential self-selection scenarios that could drive the main results. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). In columns (1) to (3), firms that decrease their activity in a preexisting sector while entering sector n (*Shifting firms*) are excluded. *Shifting firms* are firms for which the minimum growth rate of sales is negative and greater than 100%, 50% and 25% in absolute value. We compute the growth rate of sales between $t-1$ and t in each sector in which firms were operating at $t-1$ and take the firm-level minimum of sectoral growth rates. In column (4), *Serial acquirers* are excluded. Serial acquirers are firms that enter more than one sector by acquisition during the time period. Control variables include the number of workers in logarithms as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
Excluding:	Shifting firms			Serial acquirers
	100%	50%	25%	
	(1)	(2)	(3)	(4)
Internal HC$_{g,n,t-1}$	-0.010***	-0.012***	-0.012***	-0.003***
	(0.003)	(0.003)	(0.004)	(0.001)
log(#workers)$_{g,t-1}$	0.010***	0.010***	0.011***	0.003***
	(0.002)	(0.002)	(0.003)	(0.001)
Cash holdings/#workers$_{g,t-1}$	0.056	0.072	0.073	0.001
	(0.046)	(0.058)	(0.057)	(0.014)
Fixed assets/#workers$_{g,t-1}$	-0.010	0.007	0.041	-0.011
	(0.029)	(0.032)	(0.036)	(0.010)
Value added/#workers$_{g,t-1}$	0.043	0.039	0.024	0.016
	(0.033)	(0.043)	(0.050)	(0.016)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
R^2	0.205	0.222	0.243	0.157
Observations	74510	54072	38847	72681

Table 8. Human Capital and the Scale of the New Entry

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and give the results of the baseline regression when requiring that firms realize similar entry sales or capital expenditures during the year of entry. In columns (1) and (2), we rank entry sales into ten deciles and run our baseline regressions with interacted sector of origin × sector of entry × year × sales decile fixed effects. Sales are defined as the total amount of sales realized in the sector of entry n by firm g at time t. In columns (3) and (4), we replace sales with investment deciles. In the case of build entries, investment is measured as the total amount of capital expenditures realized by firms that entered sector n at time t. In the case of buy entries, investment is the amount of acquired fixed physical assets. The dependent variable is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). Control variables include the number of workers in logarithms as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
	Sales		Investment	
	(1)	(2)	(3)	(4)
Internal HC$_{g,n,t-1}$	-0.009***	-0.007**	-0.004*	-0.003*
	(0.003)	(0.003)	(0.002)	(0.002)
Origin-Entry-Year × Sales bucket FE	Yes	Yes	No	No
Origin-Entry-Year × Inv. bucket FE	No	No	Yes	Yes
Controls	No	Yes	No	Yes
R^2	0.232	0.235	0.267	0.267
Observations	45959	45959	31735	31735

Table 9. Reallocation Costs in the Internal Labor Market

Source: EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Subsidiaries of firms that enter a new sector internally during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests whether subsidiaries that enter the new sector have high human capital relative to the other subsidiaries of the same firm. The dependent variable is a dummy variable that takes value one if the entry in sector n at time t is made through subsidiary f and zero if the entry is not made through subsidiary f within firm g. Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a subsidiary-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the subsidiary is adapted to the sector of entry (see Section 3.2). Columns (1) and (2) are the baseline specifications. In columns (3) and (4), the measure of internal human capital is calculated at $t-2$ and $t-3$, respectively. Columns (2) to (4) include the following set of control variables: the number of workers, the subsidiary's cash holdings, amount of tangible assets and value added, with the last three variables being scaled by the number of workers in the subsidiary. All models are estimated with firm × sector of entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}$(Build)$_{f,n,t}$			
	(1)	(2)	(3)	(4)
Internal HC$_{f,n,t-1}$	0.021***	0.024***		
	(0.004)	(0.004)		
log(Number of employees)$_{f,t-1}$		0.018***	0.017***	0.017***
		(0.003)	(0.003)	(0.003)
Cash holdings/#workers$_{f,t-1}$		0.308**	0.320**	0.308*
		(0.153)	(0.161)	(0.159)
Fixed assets/#workers$_{f,t-1}$		0.078***	0.079***	0.086***
		(0.026)	(0.027)	(0.029)
Value added/#workers$_{f,t-1}$		0.080**	0.071**	0.062*
		(0.032)	(0.034)	(0.035)
Internal HC$_{f,n,t-2}$			0.024***	
			(0.004)	
Internal HC$_{f,n,t-3}$				0.023***
				(0.004)
Firm-Entry-Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	No
R^2	0.229	0.235	0.236	0.234
Observations	362089	362089	316760	267060

Table 10. Human Capital and Workforce Adjustment

Source: matched employer-employee dataset, ownership links dataset, EAE survey. *Sample:* Subsidiaries through which internal entries are realized during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the link between internal human capital and the variation in the number of workers. The main variable in the first three columns is the growth rate of the number of workers (computed as in Davis and Haltiwanger (1992)). The variation is computed within the intervals[t-1;t+1], [t-1;t], and [t;t+1] with t being the year of entry in the new sector. In the last column, the main variable is the simple difference in *Internal Human Capital*$_{g,n,t'}$ taken between t and $t+1$. The main independent variable is a subsidiary-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. This variable measures the extent to which the workforce of the subsidiary is adapted to the sector of entry (see Section 3.2). We include the following set of control variables: the number of workers, the subsidiary's cash holdings and amount of tangible assets and value added, with these three variables being scaled by the number of workers in the subsidiary. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	Δ Number of workers$_f$			Δ Internal HC$_f$
	[t-1 ; t+1]	[t-1 ; t]	[t ; t+1]	[t ; t+1]
	(1)	(2)	(3)	(4)
Internal HC$_{f,n,t\text{-}1}$	-0.013*	0.002	-0.016**	-0.105***
	(0.007)	(0.001)	(0.008)	(0.032)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R^2	0.639	0.102	0.659	0.422
Observations	19219	26010	19219	16091

Table 11. Human Capital, Diversification, and Local Labor Market Tightness

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the period 2010-2014 (Pole Emploi began collecting the list of occupations in short supply in 2010). The table reports OLS estimates and tests the effect of tight local labor markets for key occupations on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). *LLM Tightness*$_{n,z,t}$ is the sum of occupations in short supply in the local labor market z at time $t-1$, weighted by the occupation $\times$ sector $\times$ year fixed effects, scaled by the number of occupations present in sector n at (see Equation (4.2.3)). Terciles of LLM tightness are included in columns (1) and (2). The models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. Control variables include the number of workers in logarithms as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector of origin $\times$ sector of entry $\times$ year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
Tercile of LLM Tightness$_{n,z,t-1}$:	All		$\leq$ P33	$\geq$ P33 and $\leq$ P66	$\geq$ P66
	(1)	(2)	(3)	(4)	(5)
Internal HC$_{g,n,t-1}$	-0.016***	-0.008**	-0.003	-0.013**	-0.016***
	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)
2nd tercile of Tightness$_{n,z,t-1}$	0.007**	0.006**			
	(0.003)	(0.003)			
3rd tercile of Tightness$_{n,z,t-1}$	0.014***	0.012***			
	(0.005)	(0.004)			
2nd t. Tightness$_{n,z,t-1}$ $\times$ Int. HC$_{g,n,t-1}$	-0.005**	-0.004**			
	(0.002)	(0.002)			
3rd t. Tightness$_{n,z,t-1}$ $\times$ Int. HC$_{g,n,t-1}$	-0.005**	-0.004*			
	(0.002)	(0.002)			
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
R^2	0.418	0.424	0.525	0.501	0.377
Observations	28598	28957	7953	7796	8067

Table 12. Local Labor Market Tightness and the Value of Building

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the period 2010-2014 (Pole Emploi began collecting the list of occupations in short supply in 2010). The table reports OLS estimates and tests the effect of tight local labor markets for key occupations on the relationship between human capital and performance in the sector of entry. The dependent variable $log(Sales)_{g,n,z,t}$ is the logarithm of sales realized by firm g in sector n the year of entry t. The variable is defined only for build entries, i.e., when $1(Buy)_{g,n,z,t} = 0$. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). *LLM Tightness*$_{n,z,t}$ is the sum of occupations in short supply in local labor market z at time $t-1$, weighted by the occupation $\times$ sector $\times$ year fixed effects, scaled by the number of occupations present in sector n at (see Equation (4.2.3)). Terciles of LLM tightness are included in columns (1) and (2). The models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. Control variables include the number of workers in logarithms as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector of origin $\times$ sector of entry $\times$ year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$log(Sales)_{g,n,t}$				
Tercile of LLM Tightness$_{n,z,t-1}$:	All		$\leq$ P33	$\geq$ P33 and $\leq$ P66	$\geq$ P66
	(1)	(2)	(3)	(4)	(5)
Internal HC$_{g,n,t-1}$	0.101***	0.094***	0.076	0.048	0.116*
	(0.021)	(0.028)	(0.052)	(0.051)	(0.060)
2nd tercile of Tightness$_{n,z,t-1}$		0.004			
		(0.036)			
3rd tercile of Tightness$_{n,z,t-1}$		-0.009			
		(0.049)			
2nd t. Tightness$_{n,z,t-1}$ $\times$ Int. HC$_{g,n,t-1}$		0.008			
		(0.029)			
3rd t. Tightness$_{n,z,t-1}$ $\times$ Int. HC$_{g,n,t-1}$		0.012			
		(0.030)			
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.498	0.498	0.569	0.505	0.488
Observations	27760	27760	7635	7512	7786

Table 13. Diversification, Human Capital and Firing Costs

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of firing costs on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). *Permanent workers (%)*$_{g,t-1}$ measures the ratio of permanent to temporary workers in firm g's workforce at $t-1$. Terciles of *Permanent workers (%)*$_{g,t-1}$ are included in columns (1) and (2). *Length case settlements*$_{z,t-1}$ measures the average length of local labor case settlements in jurisdiction z at time $t-1$. There are 140 jurisdictions over the sample period. Terciles of *Length case settlements*$_{z,t-1}$ are included in columns (3) and (4). Control variables include the number of workers in logarithms as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector of origin × sector of entry × year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Build})_{f,n,t}$			
	(1)	(2)	(3)	(4)
Internal HC$_{g,n,t-1}$	-0.003**	-0.002	-0.010***	-0.010***
	(0.002)	(0.001)	(0.003)	(0.003)
2nd tercile of Permanent workers (%)$_{g,t-1}$	0.001	0.002		
	(0.002)	(0.002)		
3rd tercile of Permanent workers (%)$_{g,t-1}$	0.002	0.003		
	(0.002)	(0.002)		
2nd t. Permanent workers (%)$_{g,t-1}$ × Int. HC$_{g,n,t-1}$		-0.003***		
		(0.001)		
3rd t. Permanent workers (%)$_{g,t-1}$ × Int. HC$_{g,n,t-1}$		-0.003**		
		(0.001)		
2nd tercile of Length case settlements$_{z,t-1}$			-0.001	-0.000
			(0.002)	(0.002)
3rd tercile of Length case settlements$_{z,t-1}$			-0.003	-0.002
			(0.003)	(0.003)
2nd t. Length case settlements$_{z,t-1}$ × Int. HC$_{g,n,t-1}$				-0.002**
				(0.001)
3rd t. Length case settlements$_{z,t-1}$ × Int. HC$_{g,n,t-1}$				-0.004**
				(0.002)
Controls	Yes	Yes	Yes	Yes
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
Jurisdiction FE	No	No	Yes	Yes
R^2	0.420	0.421	0.515	0.517
Observations	63722	63722	87341	89650

Internet Appendix

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

Appendix A provides proofs of the predictions and a micro-foundation for the firing costs.

Appendix B presents additional analyses to accompany our main empirical results.

Table B1 shows the explanatory power of various fixed effects combinations to define our measure of internal human capital.

Table B2 presents estimated fixed effects for top 5 occupations in three 5-digit SIC sectors.

Table B3 presents the top combinations of sector of origin and sector of entry in our sample.

Table B4 presents the top 10 occupation in short supply in 348 local labor markets.

Table B5 reruns our main results in table 3 at a 3-digit sector level (instead of 5-digit).

Table B6 reruns the results in table 11 using an alternative definition of LLM tightness.

Table B7 tests whether public acquirers behave differently in their diversification choices.

Appendix C provides definition of variables used in the empirical analysis.

Appendix A Theoretical Appendix

A.1 Predictions 1 and 2

When $\gamma > (\sigma - 1)/\sigma$, the difference $\Delta(c_I, c_E, c_A) = \Pi^{Build}(c_I, c_E) - \Pi^{Buy}(c_I, c_E, c_A)$ is a decreasing function of c_I. Moreover, $\lim_{c_I \to 0} \Delta(.) = F > 0$, and

$$\lim_{c_I \to \infty} \Delta(c_I) = F + K \left(c_E^{-\frac{\gamma}{\gamma-1}} \right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} - K \left(c_E^{-\frac{\gamma}{\gamma-1}} + c_A^{-\frac{\gamma}{\gamma-1}} \right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}}, \tag{16}$$

which is negative if F is small enough relative to the marginal labor costs of workers in the target firm. This ensures the unique existence of $c_I^* > 0$ such that for any c_E and c_A, $c_I > c_I^*$ implies $\Delta(.) = 0$. This proves prediction 1. Moreover, $\Delta(.)$ is a decreasing function of c_E; thus, $\Delta(.) = 0$ for a lower threshold c_I^* when c_E is large, implying prediction 2.

A.2 Micro-foundation of the fixed cost: Endogenous restructuring costs

We propose a micro-foundation for F, based on the cost of laying off some workers in the acquired firm. We denote ϕ as the marginal cost of a layoff, L_A^{before} as the labor input of the acquired firm *before* the acquisition, and L_A^{after} as the labor input *after* the merger has occurred. We assume that the layoff cost is proportional to the distance between L_A^{before} and L_A^{after}:

$$F(L_A^{before}, L_A^{after}) = \phi(L_A^{before} - L_A^{after}). \tag{17}$$

Given the standard monopolistic competition framework, output Y is equal to $K(\sigma - 1)c^{-\sigma}$. Since $Y = \mathcal{L}$, we can write the labor input of the acquired firm before the acquisition as

$$L_A^{before} = c_A^{-\sigma} K(\sigma - 1). \tag{18}$$

After the target is acquired, the acquiring firm minimizes the total labor cost across the three worker pools $i \in \{I, E, A\}$, equal to $\sum_i c_i L_i^{after}$, subject to the production function (1). The first-order conditions of this problem can be rewritten as

$$L_i^{after} = \frac{c_i^{-\frac{1}{1-\gamma}}}{\sum_i c_i^{-\frac{\gamma}{1-\gamma}}} \cdot \sum_i c_i L_i^{after}, \tag{19}$$

so that after the acquisition, the labor input coming from the acquired firm is

$$L_A^{after} = c_A^{-\sigma} K(\sigma-1)\left(1+\left(\frac{c_E}{c_A}\right)^{-\frac{\gamma}{1-\gamma}}+\left(\frac{c_I}{c_A}\right)^{-\frac{\gamma}{1-\gamma}}\right)^{(\frac{1-\gamma}{\gamma})(\sigma-1)-1}. \tag{20}$$

Comparing (18) and (20), we have $L_A^{after} \leq L_A^{before}$. It is clear from (20) that $F(L_A^{before}, L_A^{after})$ is decreasing in c_I and c_E. Recall that $\Delta(c_I, c_E, c_A)$ is also a decreasing function of c_I and c_E. We now have $\lim_{c_I \to 0} \Delta(c_I) = \phi L_A^{before} > 0$, and

$$\begin{aligned}\lim_{c_I\to\infty}\Delta(c_I) =& K\left(c_E^{-\frac{\gamma}{\gamma-1}}\right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} - K\left(c_E^{-\frac{\gamma}{\gamma-1}}+c_A^{-\frac{\gamma}{\gamma-1}}\right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} \\ &+ \phi c_A^{-\sigma}K(\sigma-1)\left(1-\left(1+\left(\frac{c_E}{c_A}\right)^{-\frac{\gamma}{1-\gamma}}\right)^{(\frac{1-\gamma}{\gamma})(\sigma-1)-1}\right),\end{aligned} \tag{21}$$

$\Delta(.)$ retains the same properties as before. It is strictly decreasing, positive when $c_I = 0$, and negative when $c_I \to \infty$ if (21) is negative, i.e., if ϕ is not too large. Therefore, the testable predictions stated in the main text remain unchanged.

A.3 Optimal marginal costs (Equation (8))

Given a minimum threshold $\underline{h}_{io}$ and the random match quality $h \geq 1$ following a Pareto distribution with cf $\Psi(h)$, each worker occupation $o \in \mathcal{O}$ supplies the following amount of human capital:

$$H_{io} \equiv N_i a_{io} m_o \int_{\underline{h}_{io}}^{\infty} h d\Psi(h) = \frac{N a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}, \tag{22}$$

with $i \in \{I, E, A\}$.

We write the Lagrangian of the minimization of (6) subject to (7) and (22):

$$\begin{aligned}\mathcal{L} =& N_i\left[\sum_{o\in\mathcal{O}} a_{io} w_o \underline{h}_{io}^{-k} + f c_i\right] \\ &+ \sum_{o\in\mathcal{O}} \mu_o\left[H_{io} - \frac{N_i a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}\right] \\ &+ \lambda_{io}\left[L_i - \left(\sum_{o\in\mathcal{O}} H_{io}^{\theta}\right)^{1/\theta}\right].\end{aligned}$$

We obtain the first-order conditions

$$\mu_o = \frac{w_o}{m_o \underline{h}_{io}} \qquad \text{(wrt } \underline{h}_{io}\text{)} \tag{23}$$

$$C_i = \sum_{o \in \mathcal{O}} \mu_o H_{io} \qquad \text{(wrt } N_i\text{)} \tag{24}$$

$$\frac{w_o}{m_o \underline{h}_{io}} = \lambda H_{io}^{\theta-1} \left(\sum_{o \in \mathcal{O}} H_{io}^{\theta} \right)^{\frac{1}{\theta}-1} \qquad \text{(wrt } H_{io}\text{)}. \tag{25}$$

Plugging (23) into (25), then rearranging and summing over o, we obtain

$$C_i = \left[\sum_{o \in \mathcal{O}} \left(\frac{a_{io} m_o^k w_o^{1-k}}{L_i^k (N_i k)^{-1} (k-1)} \right)^{\frac{\theta}{\beta}} \right]^{\frac{\beta}{\theta(1-k)}}. \tag{26}$$

We guess that $C_i = c_i L_i$. Using the definition of total costs C_i given by (6) we have that

$$\begin{aligned} c_i L_i &= N_i \left(\sum_{o \in \mathcal{O}} a_{io} w_o \underline{h}_{io}^{-k} + f c_i \right) \\ &= \sum_{o \in \mathcal{O}} \left(\frac{k-1}{k} \right) \frac{m_o}{w_o \underline{h}_{io}} H_{io} + N_i f c_i \\ &= \left(\frac{k-1}{k} \right) c_i L_i + N_i f c_i, \end{aligned} \tag{27}$$

where we use Equation (22) in the second row and Equation (25) in the third row. It follows naturally from (27) that $N_i = \frac{L_i}{fk}$. Now, plugging $C_i = c_i L_i$ into (26) and using $N_i = \frac{L_i}{fk}$, we find that the optimal unit labor cost function for each labor market $i \in \{I, E, A\}$ is:

$$c_i = \left[\sum_{o \in \mathcal{O}} \left(\frac{a_{io} m_o^k w_o^{1-k}}{f(k-1)} \right)^{\frac{\theta}{\beta}} \right]^{\frac{\beta}{\theta(1-k)}}. \tag{28}$$

A.4 The relative wage share for each worker occupation (Equation (9))

Combining the first-order conditions (24) and (25) and taking the sum over $o \in \mathcal{O}$, we obtain

$$\frac{\frac{w_o}{m_o \underline{h}_{io}} H_{io}}{C_i} = \frac{H_{io}^{\theta}}{\sum_{o \in \mathcal{O}} H_{io}^{\theta}}, \tag{29}$$

so that the share of each worker occupation in total costs is equal to the share of human capital supplied by that worker occupation.

We normalize the input from each labor market at $L_i = 1$, which implies $\sum_{o \in \mathcal{O}} H_{io}^{\theta} = 1$. Plugging (22) into (29) and using $N_i = \frac{1}{fk}$, we obtain

$$\underline{h}_{io} = \left(\frac{a_{io}}{f(k-1)} \right)^{\frac{1-\theta}{\beta}} w_o^{\frac{1}{\beta}} m_0^{-\frac{\theta}{\beta}} C_i^{-\frac{1}{\beta}}, \tag{30}$$

where $\beta \equiv \theta + \theta\,(1-k)$.

From Equation (27), we have that the costs of conducting interviews, expressed in labor units, are equal to $N_i f c_i = \frac{c_i}{k}$. This implies that for any number of workers $\tilde{A_{io}}$ to work in an occupation, firms must hire a larger number of workers $A_{io} = \frac{k}{k-1}\tilde{A_{io}}$ because some workers will be conducting interviews. By assumption, the number of workers in occupation $o \in \mathcal{O}$ hired is written as $\tilde{A_{io}} = N_i a_{io}(1 - \Psi(\underline{h}_{io})) = N_i a_{io} \underline{h}_{io}^{-k}$. Finally, using (30) and $N_i = \frac{1}{fk}$, we find that the total number of workers in occupation o hired to produce one unit of output is

$$A_{io} = \left(\frac{a_{io}}{f(k-1)} \right)^{\frac{\theta}{\beta}} w_o^{\frac{-k}{\beta}} m_0^{\frac{k\theta}{\beta}} C_i^{\frac{k}{\beta}}. \tag{31}$$

It follows that relative share of the wage bill that goes to a given occupation o can be expressed as in Equation (9).

Appendix B Additional Tables for Internet Appendix

Table B1. Explaining Occupational Wage Shares using Fixed Effects Decomposition

Source: Matched employer-employee dataset. *Sample:* Firms with more than 20 employees included in the matched employer-employee dataset over the 2003-2007 and 2009-2014 periods. This table presents the adjusted R^2 of the regression of the logarithm of the share of the total wage bill by occupation for a given firm in a given year on various fixed effects. Column 1 corresponds to the estimation that we use to rank occupations by sector and year (i.e., Equation (11)).

	(1)	(2)	(3)	(4)	(5)	(6)
$Adjusted R^2$	0.661	0.503	0.448	0.178	0.475	0.005
Occupation-Sector-Year FE	Yes	Yes	No	No	No	No
Firm-Year FE	Yes	No	Yes	No	No	No
Occupation FE	No	No	No	Yes	No	No
Firm FE	No	No	No	No	Yes	No
Year FE	No	No	No	No	No	Yes
Observations	6262541	6308510	6483176	6526045	6524067	6526045

Table B2. Top-5 Occupations within Sector

Source: matched employer-employee dataset for the period 2003-2014. *Sample:* Universe of full-time French workers who earn at least 1,000 euros a year. This table reports the top-five occupation × sector × year fixed effects estimated following Equation (10). Occupation × sector × year fixed effects capture the average share of the wage bill that goes to an occupation within a sector in a given year. The higher the fixed effects are, the more important the occupation is for the production process of a given sector.

Panel A. Pharmaceutical preparations (2013)

Occupation	Estimated fixed effects
Technicians in production and control quality	0.99
Chemists, operators and skilled workers	0.96
R&D engineers and executives	0.61
Sales managers in SMEs	0.31
Sales representatives and technicians	0.29

Panel B. IT consultancy activities (2013)

Occupation	Estimated fixed effects
Computer science R&D engineers and executives	2.34
IT project manager	1.63
CEOs of service companies (1-49 workers)	0.99
CEOs of commercial companies (1-49 workers)	0.92
IT support engineers and executives	0.82

Panel C. Manufacture of motor vehicles (2013)

Occupation	Estimated fixed effects
Mechanical qualified assemblers in series	0.99
Technicians specialized in mechanics and metal work manufacturing and quality control	0.85
CEOs of companies (50-499 workers)	0.40
R&D engineers specialized in mechanics and metal work	0.40
Unskilled workers in assembly lines of metal work	0.14

Table B3. Top 10 Sectors of Entry in 2013

The table reports the number of entries for sectors with the largest number of entries by entry type in 2013. Build entries (Panel A) are identified as sales reported in a new sector (source: ESA survey). Buy entries (Panel B) are identified when firms realize diversifying acquisitions (source: SDC Platinum and Bureau van Dijk Zephyr). Sectors refer to an industry at the 4-digit level of the French SIC.

Panel A. Build entries

Sector of origin	Sector of entry	N. Pairs origin × entry
Retail sale in stores with food or tobacco	Manufacture of bread, pastry and cakes	880
Retail sale in stores with food or tobacco	Renting and leasing of cars and light motor vehicles	565
Retail sale in stores with food or tobacco	Agents involved in the sale of food	488
Retail sale in stores with food or tobacco	Retail sale of automotive fuel in specialized stores	395
Retail sale in stores with food or tobacco	Retail sale via mail order or Internet	338
Sale of cars and light motor vehicles	Maintenance and repair of motor vehicles	332
Retail sale in stores with food or tobacco	Renting or leased real estate	296
Freight transport by road	Other transportation support activities	295
Freight transport by road	Warehousing and storage	278
Freight transport by road	Maintenance and repair of motor vehicles	272

Panel B. Buy entries

Sector of origin	Sector of entry	N. Pairs origin × entry
Computer consultancy activities	Computer programming activities	14
Activities of head offices	Wholesale of machinery and equipment	8
Software publishing	Computer programming activities	8
Computer consultancy activities	Software publishing	7
Computer consultancy activities	Business and other management consultancy activities	6
Activities of head offices	Wholesale of other household goods	6
Engineering activities	Programming activities	6
Wholesale of machinery and equipment	Manufacture of lifting and handling equipment	5
Computer consultancy activities	Wholesale of computers, peripheral equipment and software	5
Computer consultancy activities	Other business support service activities	5

Table B4. Top 10 Occupations in Short Supply

Source: French unemployment agency. *Sample:* Occupations' supply and demand for 348 local labor markets in 2013. This table reports the top-10 occupations that are the reported as being in short supply the most often among local labor markets. The last column provides the percentage of local labor markets in which the occupation is reported as being in short supply. Occupations in short supply are identified by the French national employment 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.

Rank	Occupation	% of local labor markets
1	Kitchen staff	77.0
2	Machining equipment operators	58.3
3	Butchers	51.7
4	Metal workers	50.0
5	Nurses	48.9
6	Technical and commercial relation managers	48.3
7	Bakers	48.0
8	Car mechanicians	48.0
9	Catering staff	48.0
10	Machining equipment maintenance workers	46.2

Table B5. Robustness Check: Main Results with 3-Digit Sector Level

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and analyzes the role of human capital on the type of diversification strategy. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. Entries are identified with sales reported at the 3-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. This variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). Control variables include the number of workers in logarithms as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(Buy)_{g,n,t}$				
	(1)	(2)	(3)	(4)	(5)
Internal HC$_{g,n,t-1}$	-0.018***	-0.013***			-0.003*
	(0.004)	(0.004)			(0.002)
log(#workers)$_{g,t-1}$		0.010***	0.010***	0.011***	-0.052***
		(0.002)	(0.002)	(0.002)	(0.012)
Cash holdings/#workers$_{g,t-1}$		0.091*	0.090	0.093	0.162
		(0.053)	(0.062)	(0.066)	(0.137)
Fixed assets/#workers$_{g,t-1}$		-0.048	-0.056	-0.049	-0.195**
		(0.034)	(0.036)	(0.036)	(0.089)
Value added/#workers$_{g,t-1}$		0.052	0.040	0.028	-0.421***
		(0.053)	(0.053)	(0.060)	(0.116)
Internal HC$_{g,n,t-2}$			-0.015***		
			(0.004)		
Internal HC$_{g,n,t-3}$				-0.017***	
				(0.005)	
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	Yes
R^2	0.175	0.182	0.184	0.184	0.545
Observations	81180	81117	70964	59206	63750

Table B6. Robustness Check: Alternative Measure of Local Labor Market Tightness

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the period 2010-2014 (Pole Emploi began collecting the list of occupations in short supply in 2010). The table reports OLS estimates and tests the effect of tight local labor markets on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). *LLM Tightness*$_{n,z,t}$ is the sum of occupations in short supply in local labor market z at time $t-1$, scaled by the number of occupations present in sector n (see Equation (34)). Terciles of LLM tightness are included in columns (1) and (2). The models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. The set of control variables includes the number of workers, total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector of origin × sector of entry × year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
Level of LLM Tightness$_{n,z,t-1}$:	All		$\leq$ P33	$\geq$ P33 and $\leq$ P66	$\geq$ P66
	(1)	(2)	(3)	(4)	(5)
Internal HC$_{g,n,t-1}$	-0.015***	-0.009**	-0.000	-0.011**	-0.019***
	(0.004)	(0.004)	(0.002)	(0.005)	(0.007)
2nd tercile of Tightness$_{n,z,t-1}$	0.007*	0.006*			
	(0.004)	(0.004)			
3rd tercile of Tightness$_{n,z,t-1}$	0.012**	0.012**			
	(0.005)	(0.005)			
2nd t. Tightness$_{n,z,t-1}$ × Int. HC$_{g,n,t-1}$	-0.003**	-0.003			
	(0.002)	(0.002)			
3rd t. Tightness$_{n,z,t-1}$ × Int. HC$_{g,n,t-1}$	-0.007**	-0.005*			
	(0.003)	(0.003)			
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
R^2	0.418	0.424	0.498	0.543	0.400
Observations	28957	28957	7897	7778	7882

Table B7. Robustness Check: Are Public Acquirers Special?

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests whether the coefficients of the different control variables depend on whether the firm is public. Public firms are those that include at least one publicly listed subsidiary within the firm. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes value one if the entry of firm g in sector n at time t is made through an acquisition and zero if the entry is made internally. Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see Section 3.2). Control variables include the number of workers in logarithms as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at the sector of origin and sector of entry levels and reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Internal HC$_{g,n,t-1}$	-0.010***	-0.010***	-0.011***	-0.010***	-0.011***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Cash holdings/#workers$_{g,t-1}$	0.044	-0.008	0.037	0.022	0.019	0.041
	(0.045)	(0.046)	(0.045)	(0.045)	(0.045)	(0.046)
Fixed assets/#workers$_{g,t-1}$	-0.022	-0.022	-0.029	-0.019	-0.021	0.002
	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)
Value added/#workers$_{g,t-1}$	0.049	0.050	0.047	0.012	0.044	0.006
	(0.035)	(0.035)	(0.035)	(0.034)	(0.035)	(0.035)
log(#workers)$_{g,t-1}$	0.010***	0.010***	0.010***	0.009***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Public× Int. HC$_{g,n,t-1}$	-0.009					-0.004
	(0.017)					(0.018)
Public× Cash/#workers$_{g,t-1}$		0.945***				-0.454
		(0.242)				(0.359)
Public× Fixed A./#workers$_{g,t-1}$			0.250*			-0.655***
			(0.129)			(0.214)
Public× VA/#workers$_{g,t-1}$				0.809***		0.957***
				(0.129)		(0.301)
Public× log(#workers)$_{g,t-1}$					0.009***	0.006*
					(0.002)	(0.004)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.206	0.208	0.207	0.211	0.210	0.212
Observations	76296	76296	76296	76296	76296	76296

Appendix C Description of Variables

Variables	Description
Dependent variables	
$\mathbb{1}(\text{Build})_{f,n,t}$	Dummy variable that takes value one if the entry into sector n is made through a firm f within firm g and zero if the entry is *not* made through firm f. This variable is constructed only for the subsample of build entries for which we can precisely identify the entity responsible for the entry. *Source:* ownership links dataset, ESA survey.
$\mathbb{1}(\text{Buy})_{g,n,t}$	Dummy variable that takes value one if firm g enters sector n in year t through an acquisition and zero if the entry is made internally by building on preexisting resources. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Build entries are identified using reported sales from the ESA survey at the 4-digit level of the French SIC. *Source:* SDC Platinum, BvD Zephyr, Tax files, Ownership links dataset, ESA survey.
Build entries	Number of occurrences in which new sales are reported in a new sector. *Source:* ESA survey.
Buy entries	Number of occurrences in which firms realize diversifying acquisitions. *Source:* SDC Platinum and Bureau van Dijk Zephyr.
Δ workers$_f$ [t-i;t+j]	Growth rate of the number of workers computed following Davis and Haltiwanger (1992): $$\Delta\text{workers}_{t-i,t+j} = \frac{\#\ Workers_{t-i,t+j} - \#\ Workers_{t-i,t+j-1}}{0.5 * (\#\ Workers_{t-i,t+j} + \#\ Workers_{t-i,t+j-1})}$$. The growth rate ranges between -2 and 2. The variation is computed between [t-1;t+1], [t-1;t] and [t;t+1] with t being the year of entry in the new sector. *Source:* Matched employer-employee dataset.
Δ Internal Human Capital$_{g,n}$ [t;t+1]	Simple difference in Internal Human Capital$_{g,n,t}$ between t and $t+1$ with t being the year of entry in the new sector. *Source:* Matched employer-employee dataset.
log(Sales)$_{g,n,t}$	logarithm of sales realized by firm g in sector n the year of entry t. *Source:* Tax files.
Independent variables	
$\mathbb{1}(\text{No Top 10})_{g,n,t\text{-}1}$	Dummy variable that takes value 1 if none of the 10 most important occupations for sector n of entry are present in the workforce of firm g at time $t-1$ and zero if there are. *Source:* Matched employer-employee dataset.
Internal Human Capital$_{g,n,t\text{-}1}$	Sum of (exponentiated) occupation $\times$ sector $\times$ year fixed effects present both in the workforce of firm g at time $t-1$ and in the sector of entry n. The measure is scaled by the number of occupations in firm g. Fixed effects are estimated on the full French matched employer-employee dataset and are retrieved from a regression that takes the share of the wage bill that goes to a given occupation within a given sector as the dependent variable. It can be interpreted as the fit of a firm's workforce to a sector prior to diversification. *Source:* Matched employer-employee dataset.
Internal Human Capital$_{g,n,t\text{-}i}$	Lagged measure of *Internal Human Capital* at time $t-2$ or $t-3$. The variable is constructed using the same method as Internal Human Capital$_{g,n,t\text{-}1}$ *Source:* Matched employer-employee dataset.
Internal Human Capital (weighted)$_{g,n,t\text{-}1}$	Variable constructed using the same method as Internal Human Capital$_{g,n,t\text{-}1}$, except the sum of occupation $\times$ sector $\times$ year fixed effects is weighted by the number of employees by occupation in the firm. *Source:* Matched employer-employee dataset.
Internal Human Capital (without CEO)$_{g,n,t\text{-}1}$	Variable constructed using the same method as Internal Human Capital$_{g,n,t\text{-}1}$, except that CEOs are excluded from the sum of occupations. *Source:* Matched employer-employee dataset.
Internal Human Capital (plant-level)$_{g,n,t\text{-}1}$	Variable constructed using the same method as Internal Human Capital$_{g,n,t\text{-}1}$, except the fixed effects are estimated at the plant level instead of the firm level. *Source:* Matched employer-employee dataset.

Continued next page

Description of Variables (continued)

Variables	Description
Control variables	
Cash holdings/N.workers$_{g,t\text{-}1}$	Total cash holdings of firm g at time $t-1$, scaled by the number of workers in the firm. *Source:* Tax files, Ownership links dataset, Matched employer-employee dataset.
Fixed assets/N.workers$_{g,t\text{-}1}$	Total value of fixed assets held by firm g at time $t-1$, scaled by the number of workers in the firm. *Source:* Tax files, Ownership links dataset, Matched employer-employee dataset.
Length case settlements$_{z,t\text{-}1}$	Average length of local labor case settlements in jurisdiction z at time $t-1$. There are 140 jurisdictions in France over the sample period. *Source:* Matched employer-employee dataset, French Ministry of Justice.
LLM tightness$_{n,z,t\text{-}1}$	Sum over all occupations in short supply in local labor market z at time t of the exponentiated occupation × sector × year fixed effects for sector n, scaled by the number of occupations observed in sector n at time t (see Equation 4.2.3). Occupations in short supply in an LLM are identified by the French national employment agency Pole Emploi 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. France is divided into 348 different local labor markets. *Source:* Matched employer-employee dataset, Pole Emploi.
log(N.workers)$_{g,t-1}$	Logarithm of the number of workers in firm g at time $t-1$. *Source:* Matched employer-employee dataset, Ownership links dataset.
New geographical zone$_{g,t-1}$	A dummy that indicates whether the firm enters a new geographical zone at time t. We use the "region" as definition of the geographical zone (France is divided into 25 regions over the sample period). A firm enters a new geographical zone if (i) the entering subsidiary is created at time t and located in a new region, (ii) the entering subsidiary opens a plant in a new region at time t, or (iii) the entering subsidiary is acquired at time t and operates in a region in which the firm was not present at $t-1$. *Source:* Matched employer-employee dataset, Ownership links dataset, SIRENE.
Value added/N.workers$_{g,t\text{-}1}$	Total value added generated by firm g at time $t-1$, scaled by the number of workers in the firm. *Source:* Tax files, Ownership links dataset, Matched employer-employee dataset.
Permanent workers (%)$_{g,t\text{-}1}$	Fraction of workers employed in permanent contracts in firm g at time $t-1$; *Source:* Matched employer-employee dataset.
Product market distance$_{g,n,t\text{-}1}$	Distance between the sectors in which firm g operates at $t-1$ and the sector of entry n. The distance ranges from 0 to 1. A value of 0 (resp. 1) indicates that the sales of firm g are perfectly correlated (resp not correlated) at time $t-1$ to the sales of the firms operating in the sector of entry n. The product market distance is adapted from Bloom, Schankerman and Van Reenen (2013). See Section 4.2.2 for further details. *Source:* Ownership links dataset, Tax files, ESA survey.
Vertical integration$_{g,n,t\text{-}1}$	Dummy variable that is equal to one if the sector of origin of the firm (as measured in the I-O OECD classification) sources more than 5% of its inputs from the sector of entry (as measured in the I-O OECD classification). The variable is not defined when the sector of origin and the sector of entry belong to the same item of the I-O OECD classification. The method follows Fan and Goyal (2006)'s measure of vertical relatedness. *Source:* Ownership links dataset, Tax files, OECD 1999 French Input-Output table.
Other variables	
Entry sales$_{g,n,t}$	Total sales reported by firm g in sector n at time t. *Source:* Tax files
$\mathbb{1}$(1-year survival)$_{g,n,t+1}$	Dummy equal to one if firm g reports sales in sector of entry n one year after the entry n at time t. *Source:* Tax files
Investment bucket$_{g,n,t}$	Decile of total capital expenditures realized by firm g to enter sector n. Investment is measured in the case of build entries as the total amount of investment realized by the firm(s) of the firm that entered in sector n at time t. In the case of acquisitions, investment is set equal to the amount of fixed assets of the firm that has been acquired. The investment buckets are used as fixed effects to compare firms within the same investment bucket. *Source:* Tax files
Sales bucket$_{g,n,t\text{-}1}$	Deciles of realized sales by the firm g in sector n at time t. The sales buckets are used as fixed effects to compare firms within the same entry sale bucket. *Source:* Tax files

Continued next page

Description of Variables (continued)

Variables	Description
Serial acquirers$_g$	Serial acquirers are firms that enter more than one sector by acquisition during the time period. *Source:* SDC Platinum, BvD Zephyr, Ownership links dataset, ESA survey
Shifting firms$_{g,t}$	Shifting firms are firms that decrease their activity in a preexisting sector while entering sector n. They are identified with the growth rate of sales between $t-1$ and t in each sector in which firms were operating at $t-1$ that is negative and greater than 100%, 50% and 25% in absolute value. We take the firm-level minimum of sectoral growth rates. *Source:* ESA survey, Ownership links dataset.
Public firm$_g$	Dummy variable that is equal to one if at least one entity within the firm is a publicly listed company, zero otherwise. *Source:* Bureau van Dijk Amadeus