CRESTＣENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

WORKING PAPER SERIES

Matching Workers' Skills and Firms' Technologies: From Bundling to Unbundling

Philippe Choné, Francis Kramarz

N°10 / July 2021

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

Matching Workers' Skills and Firms' Technologies: From Bundling to Unbundling*

Philippe Choné† and Francis Kramarz†⋆

†*CREST-ENSAE, Institut Polytechnique de Paris, France*

⋆*Department of Economics, Uppsala University, Sweden*

July 20, 2021

Abstract

How are workers matched to their employing firms when workers have multidimensional skills and firms differ in the importance of each such skill for their production function?

When workers' skills cannot be unpacked and sold separately on skill-specific markets, the implicit price of each skill varies across firms. The wage function is shown to be log-additive in worker's quality and a firm-specific effect that reflects the firm's chosen aggregate mix of skills and the associated equilibrium matching.

When individual skills can be purchased thanks to new technologies and increasing access to outsourcing, temp agencies and other pro-market institutions, firms reinforce their hires of skills in which they have a comparative advantage yielding a more polarized matching equilibrium. Generalist workers – endowed with a balanced set of skills – are shown to benefit whereas specialists are negatively affected by markets opening. We also examine the case when workers or firms pay a fee to an unbundling platform. Then we discuss the empirical content of our model and present some empirical evidence based on this content, using Swedish data sources on workers' skills and their employing firm and occupation. We conclude by pointing connections between our contribution and various literatures.

JEL Codes: D20, D40, D51, J20, J24, J30
Keywords: bundling; multidimensional skills; matching; sorting; heterogeneous firms; polarization

*We are grateful to Marco Cuturi, Clémence Lenoir, Simon Mongey, Bérengère Patault, Oskar Skans for extremely helpful conversations. Choné gratefully acknowledges support from the French National Research Agency (ANR), (Labex ECODEC No. ANR-11-LABEX-0047) and Kramarz from the European Research Council (ERC), advanced grant FIRMNET (project ID 741467).

1 Introduction

Uberization, the Gig Economy ... These are buzzwords that attract lots of attention but generate also lots of fears. Despite important work by Acemoglu and his co-authors on robots, see Acemoglu and Restrepo (2018), or by Autor (with co-authors) on tasks and technology, see Autor (2015) and references therein, clear definitions and a convincing theoretical framework to think about jobs in this new world and, consequently the future of jobs, appear to be missing.

To understand how jobs are evolving in this new world, we start from a modelling of the old ones, the jobs of the seventies. In those years, workers typically had full-time positions at firms, over long periods of time. The jobs involved multiple skills and varied tasks, sometimes repetitive and with little cognitive content, while others were more complex, sometimes having important non-cognitive elements.

How were workers matched to their employing firms? And how has this matching changed in recent years? What is the role of various forces such as Uberization and new technologies? What is the impact of the opening of markets, potentially within global value chains, where individual skills can be purchased thanks to new technologies and increasing access to outsourcing, temp agencies and other pro-market institutions?

To capture the essence of the old world which, we believe, was characterized by firms hiring persons and their entire skill-set rather than buying each skill on a market, we follow some particularly perceptive and early scholars who reflected on the specific nature of workers' endowments. Mandelbrot (1962) is the first to note

> "the impossibility of renting the different factors to the different employers" (page 61).

Heckman and Scheinkman (1987), from which the above citation is taken, go a step further in their analysis of the impact of **Bundling** – i.e. when a worker's package of skills cannot be unpacked – onto wages and more precisely *the (differential) payment of similar skills across sectors.*

In this article, we start from Heckman and Scheinkman's theoretical insight and build on their study of bundling. A **Bundle** will denote a set of skills *when it cannot be unpacked.* This bundle of skills is what the employing firm may use when it hires a worker. There are k skills (used to produce a set of k tasks by the firm) and a worker's endowment is denoted by the skill vector $x = (x_1, \ldots, x_j, \ldots, x_k)$, with j being the index for the skill-type.

In a bundled world where skills cannot be unbundled, i.e. sold separately, there is no market for each type of skill *separately* as well as no market for each task *separately.* Put differently, an employing firm has access to all skill components a person is endowed with but cannot untie these components to purchase them because there exists *no market* for such untied skills or tasks. We also follow Heckman and Scheinkman in assuming that **each firm's production function depends on its workers' (bundled) skills aggregated by skill-types**, $X = (X_1, \ldots, X_k)$, to produce a bundle of k tasks[1] *rather than each worker's (job) production aggregated over workers (jobs) employed at the firm.*[2]

Importantly, in the world we study, both firms and workers display rich multi-dimensional heterogeneity, allowing us to examine the *matching* of workers to firms and the induced *sorting*.[3] Indeed, even though we follow Heckman and Scheinkman in having a continuum of multi-dimensional skills on the worker side, we strongly depart from them by having full heterogeneity on the firm side rather than a n-sector setup (with a continuum of identical firms within each sector). We also escape from the linear characteristics approach of Lancaster (1966) that Heckman and Scheinkman follow (as their title attests) since in our approach the wage an employee receives is allowed to be a non-linear function of her bundle x.[4] Indeed, we derive the wage schedule that prevails at the general competitive equilibrium of this economy and show that it is a) homogenous of degree one in the "quality" of the worker; b) convex in the bundle. Hence, *there is more than one price per type of skill*, potentially an infinite number of such prices. Indeed, in equilibrium, the implicit price of each skill-type varies across firms and the law of one price does not apply. This result is a direct consequence of the inefficiency – *constrained efficiency* – induced by bundling: the impossibility of unpacking a worker's multi-dimensional skills. Crucially, we exhibit the allocation of workers to firms and the *sorting* patterns displayed at this equilibrium. More precisely, under usual single-crossing conditions of their technology, sorting obtains and firms hire their unique preferred mix of skill-types, say the ratio X_2/X_1 in a two-skills world in the spirit of the Roy model, a phenomenon that we label "sorting in the horizontal dimension". Depending on the skills supply prevailing in the economy, this preferred mix is obtained by hiring workers with exactly that preferred mix or by hiring a combination of workers delivering this exact preferred mix. To give an intuition of this last result, let us consider a world with two skills(-types), x_1 and x_2.[5] In this world, let us assume

[1]Following Acemoglu and Autor (2011), or recently Edmond and Mongey (2020), skills get transformed into tasks and become inputs of sectors' production functions in their approach, when they are intermediary inputs of heterogeneous firms' production functions in this model.

[2]This assumption is found in a fraction of the literature, see in particular Lindenlaub (2017), or Lindenlaub and Postel-Vinay (2020) discussed below. See also Eeckhout and Kircher (2018), discussed extensively below, for a very interesting contribution to this problem.

[3]The sorting literature is vast, starting with Becker (1973). We discuss some recent contributions below.

[4]In Heckman and Scheinkman (1987), wages are linear in skills and the returns are allowed to differ in the two sectors of their "Roy-style" economy.

[5]When there is no ambiguity, we will use skill and skill-type interchangeably in what follows.

that the supply is restricted to two types of workers with exactly $(x_1, 0)$ for type 1 and $(0, x_2)$ for type 2. A firm that needs both skills to produce, with an optimal mix equal to α between skill 1 and 2, will hire a mixture of workers of type 1 and type 2 so as to obtain this optimal mix X_2/X_1. In this example, no worker in the firm will be endowed with the optimal mix. However, when most of the supply is situated away from the axes and closer to the 45 degree line of the (x_1, x_2) quadrant, at the equilibrium all workers in the firm will be endowed with their employing firm's optimal mix. Furthermore, this optimal mix does not imply that a given firm employs workers of the same quality. For instance, when supply is located away from the axes and the production function is CES, a firm hires workers heterogeneous in their quality λ endowed with this firm's optimal mix X_2/X_1; $\{x = (\lambda X_1, \lambda X_2)$ with λ in a subset of $\mathbb{R}_+\}$.

The model not only delivers sorting in the horizontal dimension but also sorting in the vertical one. High-productivity firms are shown to be also employing a high-quality labor force (endowed with a high total amount of the different skills). A high-quality labor force, a well-defined firm-level concept, may stem from hiring many good workers, hence by increasing the size of the firm, or from hiring a smaller number of excellent workers. And, conditional on employment, high-productivity firms employ high-quality individual workers. Hence, sorting in the vertical dimension is never strict and workers can be skills-heterogeneous within their employing firm in this vertical dimension but share the same skills ratio in the horizontal one. [6]

As mentioned just above, supply together with demand conditions may yield an equilibrium in which firms, to hire their optimal combination of skills, must mix workers with skills that differ from the optimal mix. As a result, at the equilibrium, identical workers will be hired by different firms. Because this situation is reminiscent of tax (or consumer) theory, in which heterogeneous agents make an identical choice, we use in what follows the term *Bunching*.

Another consequence of our results in this bundled world (assuming homogeneity of the production function) is the log-additivity of the wage function in worker's quality and in a firm-specific effect. This last effect reflects the firm's production technology with the associated optimal mix derived from the sorting of those skills central to the firm-specific production function. This result exactly holds *in the convex portions of the wage schedule*. Bunching is shown to induce linear faces in this wage schedule. When those faces are "small" enough, the wage function is close to such log-additivity. Hence, our bundled world – with multi-dimensional skills and firms with heterogeneous production functions – delivers a wage equation of the type studied in Abowd, Kramarz, and Margolis (1999), in which the log-wage is the sum of a person-effect and of a firm-effect, the latter coming from technology rather than profit-sharing or monopsony as

[6] In the absence of bunching, see just below for an exact definition.

in recent contributions. However, since workers sort perfectly, the firm-effect cannot be separately identified from the person-effect by using workers' firm-to-firm mobility as the literature does routinely.

The new world, with Uber and its likes, is one of unfettered markets thanks to the advances of technology, globalization, and various pro-market institutions (such as temp agencies). To analyze this new world, we examine how the matching of workers to firms is altered when opening all markets for skill-types and tasks simultaneously. *Full unbundling* restores unconstrained efficiency. In a bundled world, workers must supply all their labor to their employing firm. In the unbundled world, workers' labor supply becomes endogenous: workers can choose how much skill to supply to their firm and how much skill to supply to the market. Wages become linear combinations of workers' skills endowments. In this world, the one studied by most of the previous literature, a market exists for each skill. The first characterization of these changes (going from a world with bundled skills to one were they are unbundled), directly related to our initial question about the Uberization of the economy, is obtained by identifying those workers benefitting from unbundling and those harmed by it. Indeed, again to use our two-skills example, we demonstrate that generalists – endowed with a balanced set of skills – benefit whereas specialists are negatively affected by markets opening. The intuition for this result is straightforward: workers most constrained by bundling are those who possess both skills in close quantities and are shown to be "underpaid" under bundling. Importantly, the same style of comparative advantage sorting continues to hold after unbundling, even though the exact allocation of workers to firms changes: firms reinforce their hiring in skills in which they have a comparative advantage yielding a more *polarized* sorting equilibrium.

We then examine the case when workers or firms pay a fee to the unbundling platform. We show how firms with different technologies behave differently, some complementing their workforce with skills purchased on the market. In this latter case, a firm may well pay two different prices for the same skill, one for its employees, one for its contract workers (workers supplied by the platform). Going from an infinite cost (equivalent to full bundling) to a zero cost (full unbundling) allows us to see the widening of polarization and the flattening of the equilibrium wage schedule in detail.

We then discuss the empirical content of our model and briefly present some evidence relating this empirical content to Swedish data on workers' skills, employers, and occupations. We conclude by emphasizing how our contribution helps connect various literatures that have examined related but diverse questions.

In the next Section, we present our model of bundling. In Section 3, we examine what happens when skills can be unbundled. We then discuss the empirical consequences of our model before confronting these consequences with data evidence.[7] We conclude by examining how our contribution helps connect various literatures.

2 Bundling

Workers are heterogeneous in their skill endowments. Each worker's endowment is given by a skill vector $x = (x_1, \ldots, x_j, \ldots, x_k)$, where x_j represents worker's endowment level in skill j. Throughout the paper, we refer to $\lambda = |x|$ as the overall quality of a worker of type x and to $\tilde{x} = x/|x|$ as her skill profile.[8] Skill profiles and quality can be thought of as horizontal and vertical dimensions of worker heterogeneity. Worker skills are distributed according to a positive probability measure $\mathrm{d}H^w(x)$ on $\mathcal{X} = \mathbb{R}^k_+$. The measure does not necessarily have full support.

The firm's production process involves k tasks, $k \geq 2$. Following Acemoglu and Autor (2011), task j, $j = 1, \ldots, k$, is produced by (linearly) aggregating workers' endowments in skill j, for those workers employed at the firm

$$X_j = \int x_j \, \mathrm{d}N^d(x; \phi), \tag{1}$$

where $\mathrm{d}N^d(x; \phi)$ is the number of workers of type x hired by the firm with type ϕ.[9]

The production functions $F(X; \phi)$ are concave in the firm aggregate skill vector X. Firms are heterogeneous, with their type being of the form $\phi = (\alpha, z)$, where z captures total factor productivity, i.e., $F(X; \alpha, z) = zF(X; \alpha, 1)$. We assume that the worker and firm heterogeneities have the same dimension, in other words that α lies in a space of dimension $k - 1$. Firms' types are distributed according to a probability measure $\mathrm{d}H^f(\phi)$ on a set Φ.[10]

The output in equation (1) is an aggregation of workers' skills for each skill-type j used to produce an intermediary input, task j, that enters the firm's production function $F(X; \alpha, z)$. In this respect, we can and will often use tasks (both an input of the firm's production function and an output of skill aggregation) and skills (an input to produce tasks) interchangeably in what follows.

[7] This Section contains chosen elements of a paper, co-written with Oskar Skans Nordström, Skans, Choné, and Kramarz (2021), in which we study some aspects of the empirics of bundling and unbundling using Swedish data.

[8] The underlying metric will be made precise later in the paper.

[9] We discuss other aggregation schemes later in the text. These different schemes do not affect our analysis.

[10] By changing the scales of x and z, we normalize the numbers of firms and workers to one.

A matching between workers and firms is characterized by a probability measure $\pi(x, \phi) = N^d(x; \phi)\, H^f(\phi)$ on $\mathcal{X} \times \Phi$ that satisfies

$$\mathrm{d}H^w(x) = \int \mathrm{d}N^d(x; \phi)\, \mathrm{d}H^f(\phi), \tag{2}$$

for H^w-almost $x \in \mathcal{X}$. The above equation expresses that the supply and demand for skills must coincide for all worker-types. In Section 6, we clarify the relationship of our framework with optimal transport theory. At this point, we only emphasize two distinctive features of our environment. First, whenever the production function F is nonlinear in the firm-aggregate skill X, the surplus to be shared between firms and workers , i.e., the total output in the economy

$$\text{Total Output} = \int F\left(\int x\, \mathrm{d}N^d(x; \phi); \phi\right) \mathrm{d}H^f(\phi), \tag{3}$$

is a nonlinear function of the coupling π. Hence, in contrast to the optimal transport setting, maximizing total output under the equilibrium condition (2) is *not* a linear programming problem.[11] In this respect, our framework is related to the class of weak optimal transport problems introduced by Gozlan, Roberto, Samson, and Tetali (2017). Second, the firm' demand of skill, $\mathrm{d}N^d(x; \phi)$, is a positive measure that is not normalized, i.e., that is not necessarily a *probability* measure.[12] We thus allow the size of firms to be endogenous. In Choné, Gozlan, and Kramarz (2021), we extend the existence and duality results of Gozlan, Roberto, Samson, and Tetali (2017) to this case.

The example of CES production functions: Our leading example is the CES production function that exhibits constant elasticity of substitution and decreasing returns to scale:

$$F(X; z, \alpha) = (z/\eta) \left[\sum_{j=1}^{k} \alpha_j X_j^\sigma\right]^{\eta/\sigma}, \tag{4}$$

with $\sum_{j=1}^{k} \alpha_j = 1$, $\eta < 1$, $\sigma \neq 0$, and $\sigma < 1$. The marginal rate of technical substitution for the CES function is[13]

$$\frac{F_j}{F_k} = \frac{\alpha_j}{\alpha_k} \left(\frac{X_k}{X_j}\right)^{1-\sigma}. \tag{5}$$

The parameter α_j reflects the intensity of the firm's demand for skill-type j.

[11] If firms simply aggregated the output produced by their employees, *which is not what we intend to do here*, the total surplus $\iint F(x; \phi)\, \mathrm{d}\pi(x; \phi)$ would be linear in π.

[12] The total mass of $\mathrm{d}N^d$, namely $\mathrm{d}N^d(\mathcal{X}; \phi)$, is not necessarily one.

[13] Unless specified otherwise, we use subscripts to represent partial derivatives. For instance, the marginal productivity $\partial F/\partial x_j$ is denoted by F_j.

Existence of competitive equilibria under bundling: Under bundling, the workers' sets of skills cannot be untied, there are no markets for skills. Firms are restricted to purchase packages $x = (x_1, \dots, x_k)$. The worker skills are observed by the firm and are contractible. The wage of a worker of type x is denoted by $w(x)$. We rule out agency problems: a firm that pays $w(x)$ for x gets x. The wage schedule $w(.)$ is therefore a map: $\mathbb{R}_+^k \to \mathbb{R}_+$.

Given the wage schedule $w(x)$, i.e., the wage to be paid to any worker of type x, the skill demand of a firm ϕ is a positive measure $\mathrm{d}N^d(x;\phi)$ that maximizes its profit

$$\Pi(\phi; w) = \max_{N^d(.;\phi)} F\left(\int x \, \mathrm{d}N^d(x;\phi); \phi\right) - \int w(x) \, \mathrm{d}N^d(x;\phi). \tag{6}$$

The objective of the firm depends only on the aggregate skill $X^d(\phi) = \int x \, \mathrm{d}N^d(x;\phi)$ and on the associated wage bill $\int w(x) \, \mathrm{d}N^d(x;\phi)$. The existence of equilibria is proved in Lemma A.1. As already mentioned, the connection with the weak optimal transport literature is presented in Section 6. The envelope theorem implies that

$$\nabla_\phi \Pi(\phi; w) = \nabla_\phi F(X^d(\phi); \phi), \tag{7}$$

where ∇_ϕ is the operator $(\partial/\partial z, \partial/\partial \alpha_1, \dots, \partial/\partial \alpha_{k-1})$.[14] We assume throughout the paper that for any ϕ, the map $x \to \nabla_\phi F(x;\phi)$ is injective:

Assumption 1 (Twist conditions). *For any ϕ, $x \neq y$ implies $\nabla_\phi F(x;\phi) \neq \nabla_\phi F(y;\phi)$.*

We check in the Appendix that the CES production functions satisfy the twist conditions.

2.1 Equilibrium properties of the wage schedule

We establish below necessary properties that a wage schedule must satisfy at an equilibrium. These properties come from the linear aggregation of skills within firms, given by equation (1). They guarantee the absence of arbitrage opportunities for firms. If these properties did not hold, firms could reduce their wage bill by replacing some workers with combinations of workers yielding the same aggregate skill.

Lemma 1. *In equilibrium, the wage schedule is convex and homogenous of degree one.*

The homogeneity property implies that in equilibrium a worker's wage is linear in her quality:

$$w(x) = w(\lambda \tilde{x}) = \lambda w(\tilde{x}).$$

[14] Recall that $\alpha_k = 1 - \alpha_1 - \cdots - \alpha_{k-1}$.

Hereafter, we make extensive use of the first-order derivatives $w_i(x) = \partial w/\partial w_i$, which are homogenous of degree zero, and as such depend on the worker's skill profile $\tilde{x}$ but not on her quality λ. We refer to $w_i(x)$ as the implicit price of skill i for workers of type x. As explained below, these prices are relevant for the firms that consider hiring workers of type x.

Wage structure: The homogeneous wage schedule is entirely determined by the iso-wage surface $w(x) = 1$, which is made of the set of skill types that firms can obtain in return for one dollar.[15] The set $\mathcal{W}$ of worker types paid less than one dollar

$$\mathcal{W} = \{\ x \in \mathbb{R}_+^k \mid w(x) \leq 1\ \} \tag{8}$$

is convex. Hereafter we denote the iso-wage surface by $\partial_+\mathcal{W} = \{\tilde{x} \in \mathbb{R}_+^k \mid w(\tilde{x}) = 1\}$. For any skill vector $x = \lambda\tilde{x}$ with $\tilde{x} \in \partial_+\mathcal{W}$, we refer to $\tilde{x}$ and λ as the worker's skill profile and quality respectively. As commonly observed in the nonlinear pricing literature (see, e.g., Wilson (1993) and Laffont and Martimort (2009)), a convex schedule $w(x)$ can be implemented using the menu of its tangents. In Appendix, we check that the convex and homogenous wage schedule satisfies

$$w(x) = \sum_{i=1}^k w_i(x)x_i = \max_{y \in \mathbb{R}_+^k} \sum_{i=1}^k w_i(y)x_i. \tag{9}$$

It follows from (9) that the isoquant $w(x) = 1$ is the envelope of the family of hyperplanes indexed by y with equations $\sum_i w_i(y)x_i = 1$. The literature that deals with multi-dimensional skills, Heckman and Scheinkman (1987), Edmond and Mongey (2020), assumes special forms for the family of linear tariffs. For instance, in the case of two skills, Edmond and Mongey (2020) assume two sectors with homogenous firms within each sector and a sector-specific wage schedule, in other word they assume a two-part wage schedule such as the one represented on Figure A.1. If on the contrary w is strictly convex, all points of the iso-wage surface $\partial\mathcal{W}$ are extremal points of $\mathcal{W}$.

Figure 1 shows the case of two skills. With $k = 2$, a worker's skill profile can be represented as $\tilde{x} = (\cos\theta, \sin\theta)$ and the implicit prices $w_1(\theta)$ and $w_2(\theta)$ can be parameterized by the argument θ in polar coordinates. Equation (9) can be rewritten here as

$$\tilde{w}(\theta) \stackrel{\mathrm{d}}{=} w(\cos\theta, \sin\theta) = \max_{\theta'} w_1(\theta')\cos\theta + w_2(\theta')\sin\theta, \tag{10}$$

with the maximum being achieved for $\theta' = \theta$. The iso-wage curve is the envelope of the family of straight lines $w_1(\theta')x_1 + w_2(\theta')x_2 = 1$ indexed by θ'. If the wage schedule

[15]Because the wage schedule is homogenous of degree one, it has homothetic isoquants.

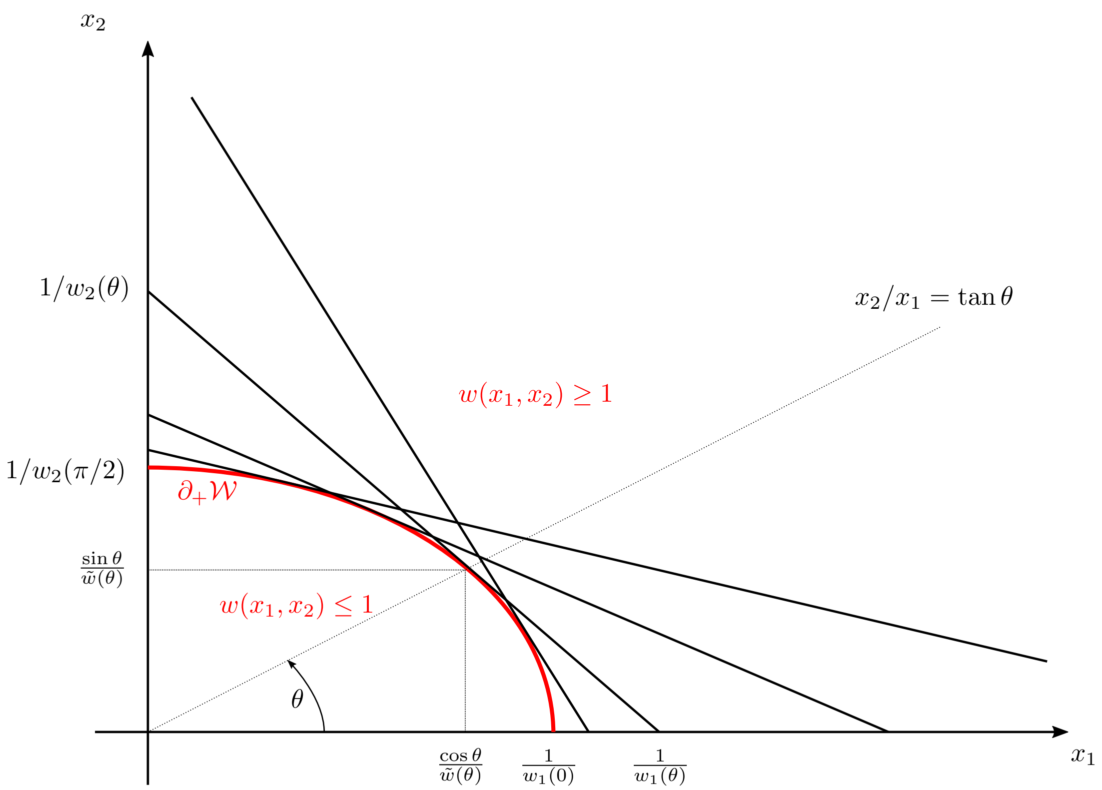


Figure 1: Iso-wage surface $\partial_+\mathcal{W} = \{ \ x \in \mathbb{R}^k_+ \mid w(x) = 1 \ \}$. The implicit prices of skills 1 and 2 for workers with skill profile θ are $w_1(\theta)$ and $w_2(\theta)$

is locally linear, the iso-wage curve coincides with one of the straight lines along a non-degenerate face, see the segment $[AB]$ on Figure 4.

Wilson (1993) interprets non-linear pricing (even for the one-dimensional case) as bundling: the charge for the purchase of a basket comprising different units is less than the sum of the charges for the components. However, in this paper, bundling is not a mere interpretation, it is at the heart of the economic environment. Let (e_i) be the canonical basis of $\mathbb{R}^k$, i.e., $e_i = (0, \dots, 1, \dots, 0)$, with 1 in the ith coordinate. Because w is convex and homogenous of degree one, it is sub-additive, hence

$$w(x) = w\left(\sum_{i=1}^{k} x_i e_i\right) \leq \sum_{i=1}^{k} w(e_i x_i) = \sum_{i=1}^{k} w_i(e_i) x_i.$$

Workers endowed with a single skill are called specialists. It is less costly to hire a worker with skill mix $x = (x_1, \dots, x_k)$ than k specialist workers endowed with the corresponding amount x_i of skill in each dimension i. We can check directly on Figure 1 that it is less costly for firms to purchase the bundle $(x_1(\theta), x_2(\theta))$ from a generalist worker (a worker endowed with both skills in sufficient quantities) than to purchase $x_1(\theta)$ units of skill

1 and $x_2(\theta)$ units of skill 2 separately from specialist workers (workers endowed with essentially one of the two skills).

2.2 The demand for skills

We now study the properties of firms' wage bill and aggregate demand for skills.

Lemma 2. *The wage bill of firm ϕ is $w(X^d(\phi))$, where $X^d(\phi)$ is its aggregate skill demand.*

Proof. The sub-additivity of the wage schedule yields

$$w(X^d(\phi)) = w\left(\int x \,\mathrm{d}N^d(x;\phi)\right) \leq \int w(x)\,\mathrm{d}N^d(x;\phi). \tag{11}$$

In other words, the lowest possible wage cost for the aggregate skill $X^d(\phi)$ is $w(X^d(\phi))$. The firm can achieve the lowest bill by hiring only workers who have the same skill profile as X^d, i.e., that are all proportional to the aggregate skill. When the wage schedule is locally linear, other ways to achieve the lowest wage bill exist, and will indeed be observed in equilibrium, as shown in our analysis in Section 2.4. □

We now characterize the firm's aggregate skill demand $X^d(\phi)$. Then we study the exact composition of the workforce within each firm, i.e. we determine the measure $\mathrm{d}N^d(.;\phi)$.

Lemma 3 (Firm aggregate skill)**.** *At the firm ϕ's aggregate skill X^d, the productivity of each skill equals its marginal price:*

$$F_j(X^d(\phi);\phi) = w_j(X^d(\phi)). \tag{12}$$

Proof. The aggregate skill maximizes $F(X^d;\phi) - w(X^d)$. Since F is concave and w is convex, the problem is well-posed, with a unique solution characterized by the first-order conditions (12). Geometrically, as shown on Figures 2 and 4, the firm's production isoquant is tangent to the iso-wage surface at X^d. □

The first-order condition (12) generalizes the standard condition that wage equals marginal productivity at a competitive equilibrium. When the wage schedule is locally linear, i.e., is of the form $< \bar{p}, x >$, we are back to $F_j(X^d(\phi);\phi) = \bar{p}$, i.e., price equals marginal productivity. Otherwise, the *implicit* price of skill i in the neighborhood of the aggregate skill X^d is the partial derivative $w_i = \partial w/\partial x_i$ evaluated at that point.

From (12), the technical rate of substitution equals the ratio of implicit prices across skills:

$$\frac{F_j(X^d(\phi);\alpha,z)}{F_j(X^d(\phi);\alpha,z)} = \frac{w_j(X^d(\phi))}{w_k(X^d(\phi))}$$

and, after eliminating total productivity factor z:

$$\frac{F_j(X^d(\phi);\alpha,1)}{F_k(X^d(\phi);\alpha,1)} = \frac{w_j(X^d(\phi))}{w_k(X^d(\phi))}.$$

In what follows, we distinguish the (quality-adjusted) size of a firm and the aggregate profile of its employees. Specifically, we write the aggregate skill demand by firm ϕ as $X^d(\phi) = \Lambda^d(\phi)\tilde{X}^d(\phi)$, where $\Lambda^d(\phi) = |X^d(\phi)|$ is the total quality of the firm's employees and $\tilde{X}^d(\phi)$ is their average profile. Using the homogeneity of the wage schedule, we eliminate $\Lambda^d(\phi)$ from the right-hand side of equation:

$$\frac{F_j(X^d(\phi);\alpha,1)}{F_k(X^d(\phi);\alpha,1)} = \frac{w_j(\tilde{X}^d(\phi))}{w_k(\tilde{X}^d(\phi))}. \tag{13}$$

Homogenous production functions: If the production functions have homothetic isoquants, the slopes F_j/F_k evaluated at $X^d = \Lambda^d\tilde{X}^d$ do not depend on Λ^d, and we can further simplify the above equalities as follows:

$$\frac{F_j(\tilde{X}^d(\phi);\alpha,1)}{F_k(\tilde{X}^d(\phi);\alpha,1)} = \frac{w_j(\tilde{X}^d(\phi))}{w_k(\tilde{X}^d(\phi))}. \tag{14}$$

The profile of the workers employed by a firm therefore depends on the intensity parameters α, but not on the total factor productivity parameter z.

When the production function $F(x;\phi)$ is homogenous of degree $\eta < 1$, two firms that differ only in their size parameter z (total factor productivity) have proportional aggregate skills. Noticing that w_j and F_j are homogenous of degree 0 and $\eta - 1$ respectively, we get from (12) that

$$X^d(\alpha,z) = z^{1/1-\eta}\, X^d(\alpha,1).$$

The total quality of a firm's employees is determined by maximizing its profit:

$$\Pi(\phi;w) = \max_{\Lambda} z\, F(\Lambda\tilde{X}^d(\alpha);\alpha,1) - \Lambda w(\tilde{X}^d(\alpha)).$$

Using that F is homogenous of degree $\eta < 1$, we find that the wage bill of firm $\phi = (\alpha,z)$ is given by

$$w(X^d(\phi)) = \left[\eta\, z\, F\left(\tilde{X}^d(\alpha);\alpha,1\right)\right]^{\frac{1}{1-\eta}}, \tag{15}$$

where $\tilde{X}^d(\alpha) \in \partial_+\mathcal{W}$ represents its aggregate skill profile.

Assumption 2. *The parameter $\alpha = (\alpha_1, \ldots, \alpha_k)$ belongs to the $(k-1)$-dimensional surface $\sum_{j=1}^k \alpha_j = 1$ in $\mathbb{R}_+^k$. The marginal rate of technical substitution F_j/F_k increases with α_j and decreases with α_k.*

Proposition 1 (Matching of aggregate skill profiles)**.** *Under Assumption 2 and with homothetic isoquant production functions, if a firm's technology is more intensive in skill j (i.e., α_j is higher), then it uses relatively more of that skill.*

Proof. As the production function F and the wage schedule w are respectively concave and convex, the left-hand side and the right-hand side of (14) respectively increases and decreases with the ratio $\tilde{X}_k^d(\alpha)/\tilde{X}_j^d(\alpha)$, where $\tilde{X}_j^d$ and $\tilde{X}_k^d$ are the j-th and k-th components of the skill mix $\tilde{X}^d$. By Assumption 2, the left-hand side decreases with α_k/α_j. The demand of skill k relative to skill j, $\tilde{X}_k^d(\alpha)/\tilde{X}_j^d(\alpha)$, increases with α_k/α_j. $\square$

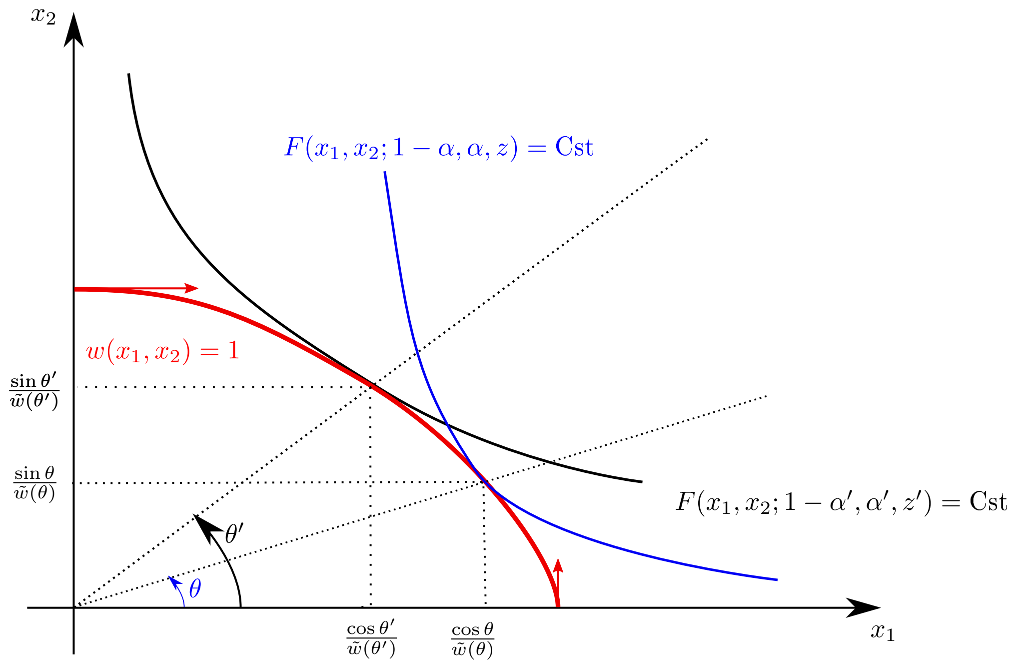


Figure 2: Matching in the skill dimension: Firm $(1-\alpha, \alpha, z)$ is more intensive in skill 1 than firm $1-\alpha', \alpha', z$.

Two tasks: As explained above, when $k = 2$, we may parameterize skill profiles as $\tilde{X} = (\cos\theta, \sin\theta)$ and represent the aggregate demand $X^d = (\Lambda^d \cos\theta^d, \Lambda^d \sin\theta^d)$ in polar coordinates, where Λ^d is the total quality of workers employed at firm ϕ. We now show that the aggregate workers-to-firms matching pattern exhibits positive assortative matching (PAM), in the sense that the Jacobian $D_{(\alpha_2,z)}(\theta^d, \Lambda^d)$ is a P-matrix, i.e., all the principal minors of the Jacobian are positive.

Lemma 4. *Under Assumption 2 and with homothetic isoquant production functions, the firm-aggregate matching pattern* $(\theta^d(\alpha_2, z), \Lambda^d(\alpha_2, z))$ *is PAM.*

In contrast to Lindenlaub (2017), the sorting pattern highlighted above pertains to the horizontal and vertical dimensions of the workers' skills (skill profile and total quality) rather than to each of the two skills.[16] Even more importantly, the above PAM property applies in our context to firms' *aggregates* rather than to individual workers' characteristics. At the individual level, two points are worth mentioning. First, even though the workers-to-firms matching is arbitrary in the vertical dimension (worker qualities), we explain in Section 4.1 that the monotonicity of the total quality of employees with the firms' total factor productivity does have testable implications. Second, regarding the horizontal dimension (worker profiles), workers' sorting patterns may be blurred by bunching, see Section 2.4.

CES with two tasks example: We consider the production function (4) with two skills ($k = 2$):

$$F(X_1, X_2; \alpha_1, \alpha_2, z) = \frac{z}{\eta} \left(\alpha_1 X_1^\sigma + \alpha_2 X_2^\sigma\right)^{\eta/\sigma}.$$

This technology satisfies Assumption 2.

With the parametrization $\tilde{X}^d = (\cos\theta^d, \sin\theta^d)$, the general workers-to-firms matching condition (14) writes

$$\left(\tan\theta^d(\alpha_2)\right)^{1-\sigma} = \frac{\alpha_2}{1-\alpha_2} \frac{w_1\left(\theta^d(\alpha_2)\right)}{w_2\left(\theta^d(\alpha_2)\right)}. \tag{16}$$

The matching between workers and firms is represented by the increasing function $\theta^d(\alpha_2)$ implicitly defined by (16). The relative skill endowment in skill 2 of the workers, $\theta^d(\alpha_2)$, increases with the demand intensity in skill 2, α_2, as illustrated on Figures 2 and 4.

Non-homothetic isoquant production functions: If the production functions have non-homothetic isoquants, equation (13) does not simplify into (14). The aggregate profile $\tilde{X}^d(\alpha_2, z)$ depends on total factor productivity z because the ratio $F_j(\Lambda^d\tilde{X}^d)/F_k(\Lambda^d\tilde{X}^d)$ depends on total worker quality Λ^d, which itself depends on z. In the Appendix, we prove the following

Lemma 5. *With two tasks ($k = 2$), the total quality of the workers employed by firm,* $\Lambda^d(\alpha_2, z)$*, increases with firm's total factor productivity* z*.*

[16] In technical terms, we use polar coordinates rather than Cartesian coordinates. Moreover, contrary to Lindenlaub (2017), our technology does not exhibit a diagonal structure.

In the Appendix, we provide a technical condition (inequality (A.5)) guaranteeing that PAM holds in the sense of Lemma 4 even when production isoquants are non homothetic.[17]

2.3 Equilibrium without bunching

In this section, we assume that the wage schedule is strictly convex or equivalently that the iso-wage surface $\partial_+\mathcal{W}$ is strictly concave. Under this circumstance, the minimization of the wage bill at given aggregate skill X^d, i.e. the equality in (11), imposes that firm ϕ hires only workers with skill profile $\tilde{X}^d(\phi)$. It follows that the support of the matching transport π is included in the graph of $\tilde{X}^d$.

Because the wage schedule w is homogenous of degree one and skill-types are aggregated additively within the firm, workers with proportional skills are perfectly substitutable, up to a multiplicative factor reflecting their overall quality. Sorting is therefore arbitrary in this vertical dimension.

For any skill vector x, the wage earned by worker with type $x/w(x)$ is one, or equivalently $x/w(x)$ belongs to the iso-wage surface $\partial_+\mathcal{W} = \{x \in \mathbb{R}^k_+ \mid w(x) = 1\}$. For any distribution H on $\mathcal{X}$, we define the distribution $W_\# H$ as the push-forward of the positive measure $w(x)H(x)$ by the projection $x/w(x)$ onto the iso-wage surface $\partial_+\mathcal{W}$:

$$W_\# H = \left(\frac{x}{w(x)}\right)_\# w(x)H.$$

Formally, for any test function h, we have[18]

$$< W_\# H, h >= \int h\left(\frac{x}{w(x)}\right) w(x)\,\mathrm{d}H(x) \tag{18}$$

The distribution $W_\# H$ is supported on the iso-wage surface $\partial_+\mathcal{W}$ and places the mass $\int_0^\infty \lambda\,\mathrm{d}H(\lambda|\tilde{x})$ on any point $\tilde{x} \in \partial_+\mathcal{W}$. That mass is nothing but the sum of the wages received by all the workers with skill profile $\tilde{x}$.

[17] The condition is a consequence of Assumption 2 when isoquants are homothetic.

[18] For any measurable map $T : \mathcal{X} \to \mathcal{Y}$, the push-forward of a positive measure μ on $\mathcal{X}$ by T is the positive measure $T_\#\mu$ on $\mathcal{Y}$ that satisfies, for all continuous function h on $\mathcal{Y}$

$$(T_\#\mu)h = \int_{\mathcal{X}} h(T(x))\,\mathrm{d}\mu(x). \tag{17}$$

Proposition 2. *When the equilibrium wage schedule is strictly convex, the matching is pure in the horizontal dimension*

$$\textit{Support } \pi \subset \{(\tilde{X}^d(\phi) \times \mathbb{R}_+ \,,\, \phi) \mid \phi \in \varPhi\}. \tag{19}$$

In this case, the equilibrium condition (2) *is equivalent to*

$$W_\# H^w = W_\# X^d_\# H^f \tag{20}$$

where the operator W *is given by* (18).

As sorting is arbitrary in the vertical dimension, the equilibrium condition collapses onto $\partial_+\mathcal{W}$. When the wage is strictly convex, firm ϕ picks all its employees from the ray $\tilde{X}^d(\phi) \times \mathbb{R}_+$ in $\mathcal{X}$, and the equilibrium condition holds pointwise on the iso-wage surface, i.e., separately for each ray. The measure $X^d_\# H^f$ is the push-forward of the distribution of the firms' technological parameters by their skill aggregate demand X^d, which we have examined in Section 2.2. Condition (20) expresses that in equilibrium the total value of efficiency units of labor offered by workers and demanded by firms coincide for each skill profile separately.

When the aggregate skill profile $\tilde{X}^d$ is increasing in its arguments, for instance under the assumptions of Proposition 1, the equilibrium condition $W_\# H^w = W_\# X^d_\# H^f$ translates into an ordinary differential equation for the matching map, as we illustrate below in the case of two tasks.

Back to the two tasks example: Assume that the production function is homogenous of degree $\eta < 1$.[19] The aggregate skill demand by firm $\phi = (\alpha, z)$ is represented as $X^d(\phi)$, where the employees' skill profile $\tilde{X}^d(\alpha) = (\cos\theta^d(\alpha_2), \sin\theta^d(\alpha_2))$ satisfies the workers-to-firms matching condition

$$\frac{F_1\left(\cos\theta^d(\alpha_2), \sin\theta^d(\alpha_2); \alpha, 1\right)}{F_2\left(\cos\theta^d(\alpha_2), \sin\theta^d(\alpha_2); \alpha, 1\right)} = \frac{w_1\left(\theta^d(\alpha_2)\right)}{w_2\left(\theta^d(\alpha_2)\right)}. \tag{21}$$

[19] In the following equations, it is important to note that homogeneity is the only requirement. Hence, these results apply broadly and not only to the CES with two tasks case.

Setting $\tilde{w}(\theta) = w(\cos\theta, \sin\theta)$ as in (10), and using expression (15) for the wage bill of firm $\phi = (\alpha, z)$, we can write the equilibrium condition (20) for any α_2 as

$$\int_0^{\theta^d(\alpha_2)} \Lambda^w(\theta)\tilde{w}(\theta)\,\mathrm{d}H^w(\theta) = \int_0^{\alpha_2} Z^f(\alpha)\left[F\left(\frac{\cos\theta^d(\alpha)}{\tilde{w}(\theta^d(\alpha))}, \frac{\sin\theta^d(\alpha)}{\tilde{w}(\theta^d(\alpha))}; \alpha, 1\right)\right]^{1/(1-\eta)} \mathrm{d}H^f(\alpha), \tag{22}$$

where $\Lambda^w(\theta) = \int_z \lambda\,\mathrm{d}H^w(\lambda|\theta)$ and $Z^f(\alpha) = \int_z (\eta z)^{1/(1-\eta)}\,\mathrm{d}H^f(z|\alpha)$ are exogenous quantities that depend on the primitive distributions H^f and H^w. The left-hand side of (22) represents the total wages earned by workers with skill profile θ below $\theta^d(\alpha_2)$. According to (15), the right-hand side represents the total wage bill paid by the employing firms of those workers, namely those with α below α_2.

Differentiating with respect to α_2 yields:

$$\Lambda^w(\theta^d)\,\tilde{w}(\theta^d)\,h^w(\theta^d)\,\frac{\mathrm{d}\theta^d}{\mathrm{d}\alpha_2} = Z^f(\alpha_2)h^f(\alpha_2)\left[F\left(\frac{\cos\theta^d}{\tilde{w}(\theta^d)}, \frac{\sin\theta^d}{\tilde{w}(\theta^d)}; \alpha, 1\right)\right]^{1/(1-\eta)}, \tag{23}$$

where θ^d stands for $\theta^d(\alpha_2)$.

Equation (23) relates the matching map $\theta^d(\alpha_2)$ implicitly given by (21) and its derivative $\mathrm{d}\theta^d/\mathrm{d}\alpha_2$ to the distributions of workers' skills and firms' technologies. It follows that for any strictly convex wage schedule $w(x)$, any homogenous production functions $F(.;\phi)$ satisfying Assumption 2, and any workers' distribution H^w, there exist distributions of the firms' technological parameters ϕ for which w is the equilibrium wage. Such distributions H^f are not uniquely identified as Equation (23) only determines (for any α_2) the quantity $Z^f(\alpha_2)h^f(\alpha_2)$ that drives the demand for workers with skill profile $\theta^d(\alpha_2)$ by firms with intensity α_2 in skill 2.

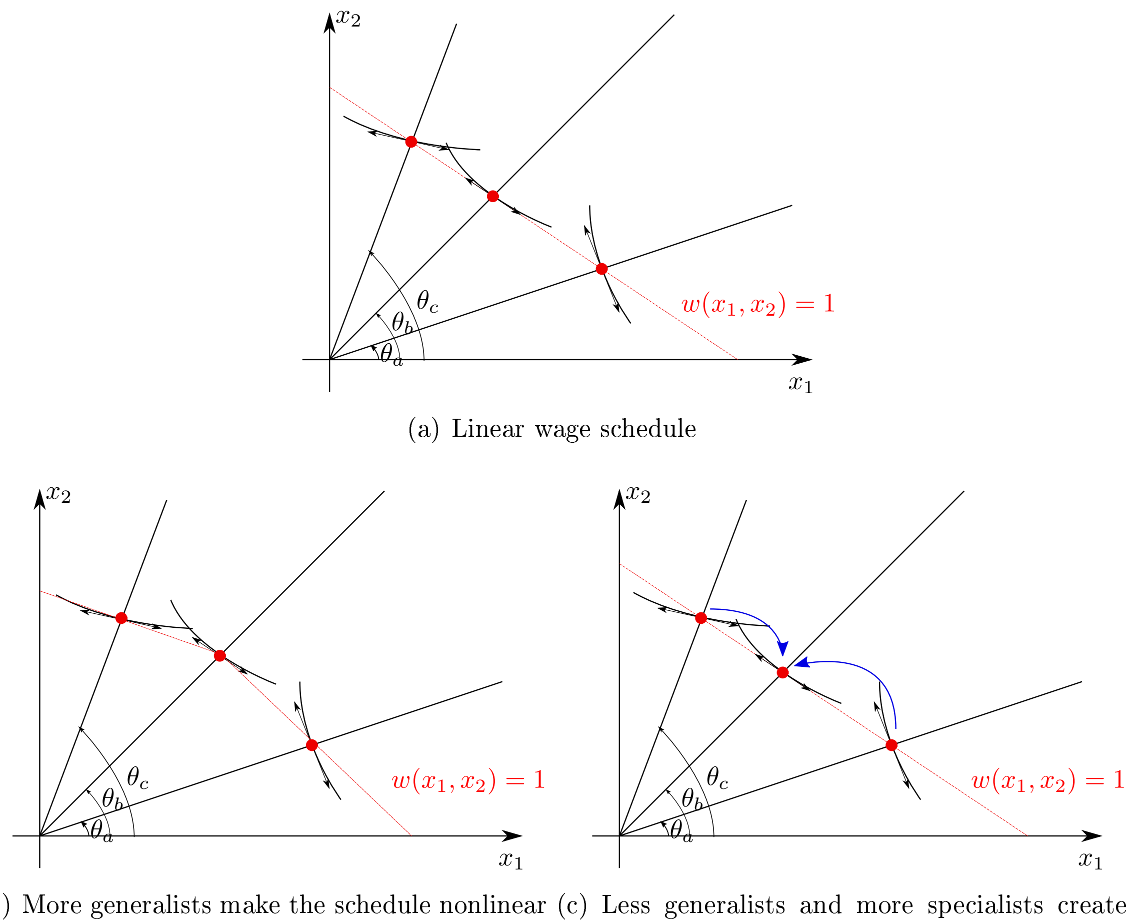
2.4 Equilibrium with bunching

We now turn to situation where bunching prevails, i.e., different firm-types hire workers with similar skill-types (albeit never using the same combination because of aggregate workers-to-firms matching). First, we explain intuitively how bunching can arise in equilibrium, and how it is connected to the heterogeneity of skill profiles within firms. Next, we formally characterize equilibria with bunching.

A simple economy with three types of skills: We start from an initial equilibrium without bunching for which the price schedule is linear, and from this equilibrium we change the distribution of skills in the economy. We first show that if we increase the relative number of "generalists" (workers with a balanced set of skills), their price falls and the wage schedule becomes nonlinear. We then show that if we decrease the relative number of generalists starting from this initial equilibrium, the wage schedule remains linear, leading to bunching in which heterogeneity of skill profiles *within firms* emerges.

We illustrate the mechanism in a setting with two tasks and three skill profiles $\theta_a < \theta_b < \theta_c$, see Figure 3. Recall $\tan\theta_i = x_{i2}/x_{i1}$ is the endowment of workers $i \in \{a, b, c\}$ in skill 2 relative to skill 1. We choose $p_1 > 0$ and $p_2 > 0$ and construct distributions H^w and H_f for which the linear wage schedule $w(x_1, x_2) = p_1x_1 + p_2x_2$ prevails in equilibrium. We choose three values for the skill intensities $\alpha_k = (\alpha_{1k}, \alpha_{2k})$, $k \in \{a, b, c\}$ such that

$$\frac{\alpha_{a1}}{\alpha_{a2}}(\tan\theta_a)^{1-\sigma} > \frac{p_1}{p_2}, \quad \frac{\alpha_{b1}}{\alpha_{b2}}(\tan\theta_b)^{1-\sigma} = \frac{p_1}{p_2}, \quad \text{and} \quad \frac{\alpha_{c1}}{\alpha_{c2}}(\tan\theta_c)^{1-\sigma} < \frac{p_1}{p_2}.$$



(a) Linear wage schedule

(b) More generalists make the schedule nonlinear (c) Less generalists and more specialists create bunching

Figure 3: Equilibrium with three relative skill endowments in the economy

Firms with intensity α_k hire workers with profile θ_k. Firms α_a would prefer workers with more skill 1 relative to skill 2, but no such workers are available in the economy.

In this discrete setting, the equilibrium is achieved separately on each ray, i.e. for θ_a, θ_b and θ_c separately. Equation (23) takes the form

$$\Lambda^w(\theta_i)\, h^w(\theta_i) = Z^f(\alpha_{2i}) h^f(\alpha_{2i}) \left[\frac{F(\cos\theta_i, \sin\theta_i; \alpha_{2i}, 1)}{\tilde{w}(\theta_i)} \right]^{1/(1-\eta)}.$$

We choose $\Lambda^w(\theta_i)\, h^w(\theta_i)$ and $Z^f(\alpha_{2i})h^f(\alpha_{2i})$ so that the above equation holds for $i = a, b, c$, i.e. so that Figure 3(a) represents the equilibrium configuration.

We now slightly increase the number of generalists $\Lambda^w(\theta_b)\, h^w(\theta_b)$. To equalize the demand and the supply of generalists, we need to reduce their wage. The equilibrium configuration is modified as shown on Figure 3(b). The wages of the two specialist types a and c remain unchanged, and firms a and c do not change their behavior. The wage schedule has become nonlinear.

To generate bunching, we on the contrary decrease the number of generalists relative to the equilibrium of Figure 3(a). Specifically, we reduce $\Lambda^w(\theta_b)\, h^w(\theta_b)$ by $\nu_b > 0$ and we define $\nu_a > 0$ and $\nu_c > 0$ by

$$\nu_b(\cos\theta_b, \sin\theta_b) = \nu_a(\cos\theta_a, \sin\theta_a) + \nu_c(\cos\theta_c, \sin\theta_c).$$

We raise the number of specialist workers $\Lambda^w(\theta_a)\, h^w(\theta_a)$ and $\Lambda^w(\theta_c)\, h^w(\theta_c)$ by ν_a and ν_c respectively. Figure 3(c) shows the new equilibrium configuration. Firms α_a and α_c do not change their behavior. Firms α_b keep the same aggregate skill $X^d(\phi)$ but obtain such an aggregate skill using a different composition of their workforce. They hire all workers with relative skill endowment θ_b, but also some workers of type θ_a and θ_c workers, specifically ν_a and ν_c efficiency units, respectively. Hence in equilibrium firms α_a and α_b both hire some θ_a workers, and firms α_b and α_c both hire some θ_c workers. In the extreme case where $\nu_b = \Lambda^w(\theta_b)\, h^w(\theta_b)$, there are no more θ_b workers in the economy, and firms α_b achieve their optimal aggregate skill θ_b by mixing θ_a and θ_c workers.

Remark: Our previous example should have made clear how we use the term *bunching*. Because there is always perfect separation in terms of the firm's aggregate skill mix – θ always increases with α – there is no bunching of the sort studied in goods consumption since there is full sorting. On the other hand, there is bunching in the sense that firms with different skills intensities, different α's, may hire workers of the same type to construct their optimal mix of skills, α. In models with goods and consumers, this would correspond to the following situation: two different consumers may buy the same product (bunching), but these consumers indeed buy many different products to form their optimal mix of products, an optimal mix that differs across consumers endowed with different types.

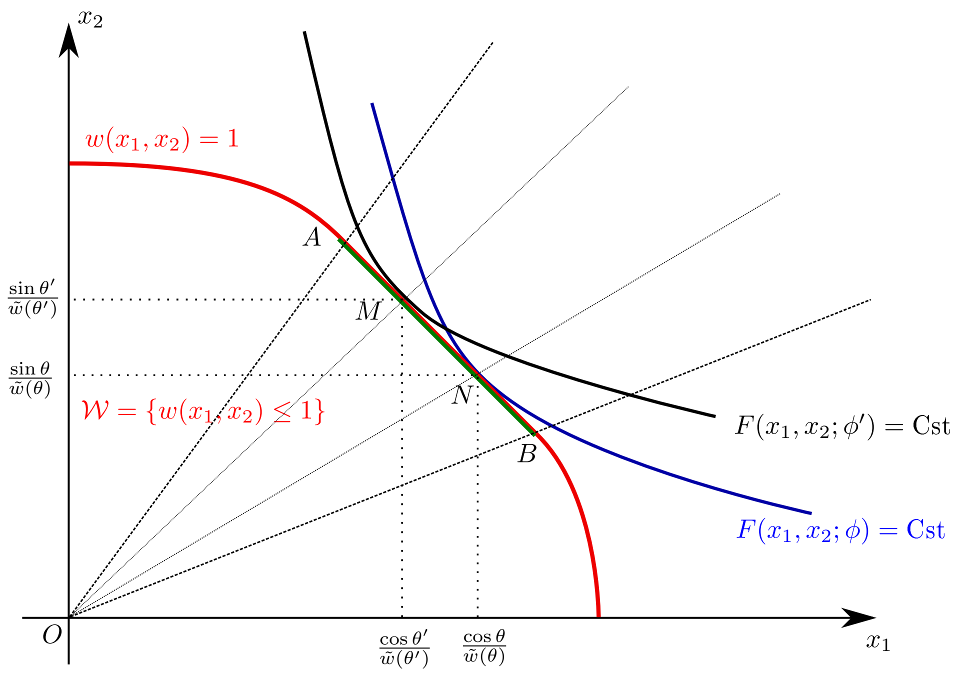


Figure 4: Matching is not pure. Firms $\phi=(\alpha_1,\alpha_2,z)$ and $\phi'=(\alpha_1',\alpha_2',z')$, pick their employees in the cone generated by the face $[AB]$ of $\mathcal{W}$ in $\mathbb{R}_+^2$. Firm ϕ' is more intensive in skill 2: $\alpha_2'>\alpha_2$ and $\theta^d(\alpha_2')>\theta^d(\alpha_2)$.

Characterization of equilibrium under bunching: If the wage schedule w is locally linear, the minimization of the wage bill, problem (11), allows a firm to hire employees with different skill profiles. In this case, the set $\mathcal{W}$ of worker types paid less than one dollar has faces of positive dimension, such as $[A,B]$ on Figure 4.[20]

Rockafellar (1970), Theorem 18.2., states that any convex set can be written as a disjoint union of relative interiors of different faces. For any X, let $\mathcal{F}(X)$ be the (unique) face of $\mathcal{W}$ such that $X/w(X)$ belongs to the relative interior of $\mathcal{F}(X)$. The cone

$$\mathcal{C}\left(X^d(\phi)\right)=\mathcal{F}\left(X^d(\phi)\right)\times\mathbb{R}_+ \tag{24}$$

is the largest set $\mathcal{C}$ in $\mathcal{X}$ such that $\tilde{X}^d(\phi)$ belongs to the relative interior of $\mathcal{C}$ and the wage w is linear on $\mathcal{C}$, see Lemma A.2. If $X/w(X)$ is an extremal point of $\mathcal{W}$ (such as point A on the figure), then $\mathcal{F}(X)$ is the singleton $\{X/w(X)\}$ and the cone is reduced to a ray such as (OA). For the firms ϕ and ϕ', $\mathcal{F}\left(X^d(\phi)\right)$ and $\mathcal{F}\left(X^d(\phi')\right)$ are equal to the segment $[AB]$, which generates the cone lying between the rays (OA) and (OB).

A firm ϕ may obtain its aggregate skill $X^d(\phi)$ by hiring workers with different individual profiles. To minimize the firm's wage bill, the support of $\mathrm{d}N^d(x;\phi)$ must be

[20] A face $\mathcal{F}$ of a convex set $\mathcal{W}$ is a convex subset $\mathcal{F}\subset\mathcal{W}$ such that $W\setminus\mathcal{F}$ is convex.

included in $\mathcal{C}\left(X^d(\phi)\right)$. Because the wage schedule w is linear on that cone, we have

$$\int w(x)\,\mathrm{d}N^d(x;\phi) = w\left(\int x\,\mathrm{d}N^d(x;\phi)\right) = w(X^d(\phi)).$$

For instance, firm ϕ on Figure 4, rather than picking employees with skills proportional to $\tilde{X}^d(\phi)$, i.e., along the half line $[OM)$, can use skills located in the entire cone AOB.

Proposition 3. *When the equilibrium wage schedule is locally linear, the matching is not pure in the horizontal dimension*

$$\text{Support}\ \pi \subset \{\,\mathcal{C}\left(X^d(\phi)\right)\ ,\ \phi) \mid \phi \in \varPhi\,\}, \tag{25}$$

where $\mathcal{C}\left(X^d(\phi)\right)$ is the cone given by (24). The equilibrium condition (2) is equivalent to the measure $W_\# X^d_\# H^f$ being dominated by $W_\# H^w$ in the convex order, i.e.,

$$W_\# H^w \succeq_C W_\# X^d_\# H^f \tag{26}$$

where the operator W is given by (18).

When there is bunching, it is no longer true that the total value of efficiency units of labor offered by workers and demanded by firms coincide for each skill profile, i.e., that the distributions $W_\# X^d_\# H^f$ and $W_\# H^w$ are equal. Equality (20), which does not hold under bunching, must be replaced with the weaker condition (26). The convex order generalizes the notion of mean-preserving spread to multidimensional settings. A measure μ_1 is said to be dominated by a measure μ_2 if and only if $\mu_2 h \geq \mu_1 h$ for all convex functions h. It means that μ_2 is "riskier" than μ_1.[21]

Bunching in the horizontal dimension leads to many-to-many matching as illustrated on Figure (5). Firms with different types hire workers with the same skill profile, and accordingly workers with the same type are employed in different firms. For instance, firms F and F' on the figure, which have different technologies, both hire workers with skills in the cone (AOB). In the extreme case where workers' skill are located only along the two rays (OA) and (OB), firms F and F' both hire workers with skill profiles A and B, but in different proportions to achieve their aggregate demand.[22]

The equilibrium condition (26) states that, locally, there is excess supply for specialist workers and excess demand for generalist ones. In terms of efficiency units of labor (as valued by the wage schedule), the distribution of workers' skills H^w lies closer to the boundary of the cone than the demand distribution $X^d_\# H^f$. The supply is more

[21] For one-dimensional distributions, μ_2 is a mean-preserving spread of μ_1.

[22] In the absence of bunching, when the equilibrium wage schedule is strictly convex, cones are degenerated, i.e., coincide with rays.

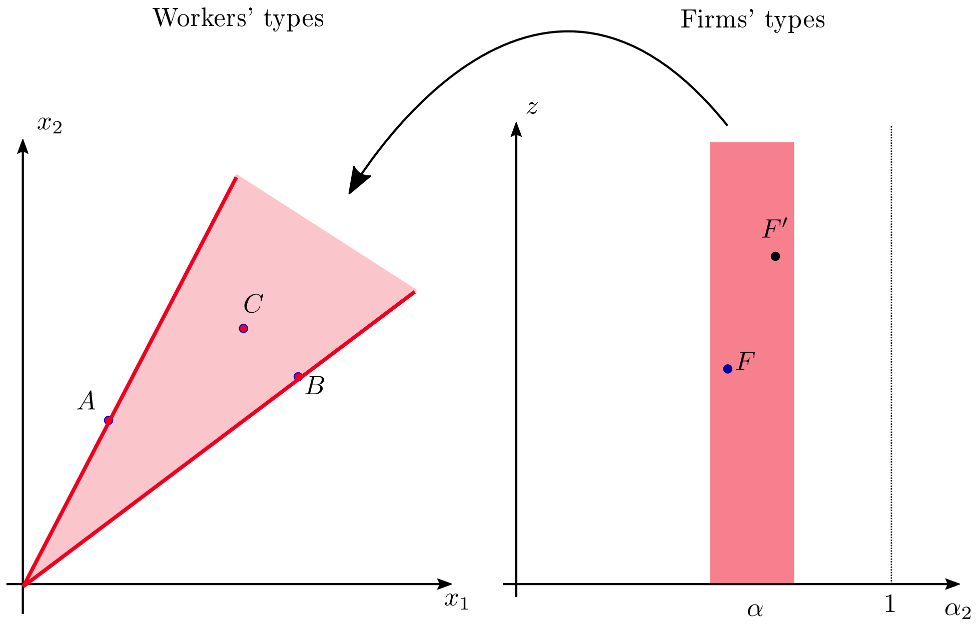


Figure 5: Sorting with bunching: Within-firm heterogeneity in skill profiles

concentrated along the rays OA and OB, while the demand is more concentrated in the interior of the cone.

Because workers' sorting is arbitrary in the vertical dimension, the equilibrium condition collapses onto the iso-wage surface: $W_\# H^w = W_\# N^d_\# H^f$. In the absence of bunching, firms' demand $N^d(.;\phi)$ may be replaced with the firms' aggregate skill X^d to get the equilibrium condition (20). The aggregate skill $X^d(\phi)$ is a deterministic transport that can be expressed under Assumption 1 as

$$X^d(\phi) = (\nabla_\phi F)^{-1} \nabla_\phi \Pi(\phi; w), \tag{27}$$

where $(\nabla_\phi F)^{-1}$ is the inverse of the function $x \to \nabla_\phi F(x;\phi)$. In the presence of bunching, Strassen et al. (1965) has demonstrated that condition (26) is equivalent to the existence of a martingale coupling between $W_\# H^w$ and $W_\# X^d_\# H^f$, i.e., the existence of a transport kernel $s_{\tilde{x}}(\tilde{y})$ that spreads the mass of $W_\# X^d_\# H^f$ over the facets of $\partial_+ \mathcal{W}$ to get $W_\# N^d_\# H^f$

$$W_\# H^w(\tilde{y}) = W_\# N^d_\# H^f(\tilde{y}) = \int \mathrm{d}s_{\tilde{x}}(\tilde{y})\, \mathrm{d}W_\# X^d_\# H^f(\tilde{x}).$$

The support of $s_{\tilde{x}}$ is the facet $\mathcal{F}(\tilde{x})$ of the iso-wage surface $\partial_+\mathcal{W}$. The mass placed by $s_{\tilde{x}}$ on $\tilde{y}$ is the share the total wage bill of firms ϕ with aggregate profile $\tilde{X}^d(\phi) = \tilde{x} \in \partial_+\mathcal{W}$ spent on workers with individual profile $\tilde{y} \in \partial_+\mathcal{W}$.

2.5 Robustness to alternative aggregation schemes

Aggregation of skills: Our model allows for additive aggregations of any transformation of the skill vector x. Let $g : \mathcal{X} \to \mathcal{X}$ be any one-to-one transformation of the skills. Let $\mathcal{M}$ denote the set of couplings of μ and ν. The primal problem

$$\max_{\pi \in \mathcal{M}(H^w, H^f)} \int F\left(\int g(x)\, \mathrm{d}\pi(x|\phi); \phi\right) \mathrm{d}H^f(\phi)$$

is equivalent to

$$\max_{\pi' \in \mathcal{M}(g_\# H^w, H^f)} \int F\left(\int y\, \mathrm{d}\pi'(y|\phi); \phi\right) \mathrm{d}H^f(\phi)$$

In other words, the change of variables $y = g(x)$ gets us back to the baseline formulation, with the transport plans being linked through $\pi'(y|\phi) = g_\#\pi(x|\phi)$. The wage schedule $\tilde{w}(y) = w(g^{-1}(y))$ has the same properties as those of $w(x)$ in the current setting. For instance, if $g(x) = (x_1^\gamma, \ldots, x_k^\gamma)$, $\tilde{w}(y)$ is homogenous of degree one in y, hence w is homogenous of degree γ in x: $w(\lambda x) = \lambda^\gamma w(x)$. Hence the fundamental requirement we impose is that skill aggregation is separable from the technology, i.e., formally that the transformation g is independent of the type ϕ of the firm.

Non-separable aggregation: Assuming such a sharp assortative matching to be a desirable feature, one can think of the following extension of our model and let us aggregate skills as follows:

$$F\left(\iint A(\lambda, z)\, \tilde{x}\, \mathrm{d}N(x)\right), \tag{28}$$

where $A(\lambda, z)$ is supermodular, i.e., $\partial^2 A/\partial\lambda\partial z > 0$. Supermodularity, however, is not enough when there are quantities (here represented by the labor demand $\mathrm{d}N$), as the example $A(\lambda, z) = \lambda z$ shows: supermodularity holds, but we are back to efficiency units (λ is "absorbed" into N). As shown by Eeckhout and Kircher (2018), $A(\lambda, z)$ needs to be non-linear in λ, for instance CES (CES functions are supermodular). Even assuming that λ and $\tilde{x}$ are independent across workers and that z and α are independent across firms, some of the properties derived in Section 2 are likely to be lost. For instance, can we preserve additive separability of $w(\lambda\tilde{x})$, convexity of the wage schedule, or can the sorting in the vertical dimension (i.e., in the workers' qualities λ) be orthogonal to the sorting in the horizontal dimension (i.e., in the skill profiles $\tilde{x}$)? Such questions will be left for future research because, at this stage, we view them as less pressing.

3 Unbundling

We now assume the availability of a technology that enables workers to unbundle their skills and allows workers and firms to trade skills as a commodity. In a first step, we assume that this unbundling technology is costless for all market participants. Then we assume that it entails some costs incurred by workers and/or firms.

3.1 Costless unbundling

In this Section, we first recall that full efficiency prevails when competitive markets for skills do exist. Then, we explain how the opening of such markets affects generalist and specialist workers. Finally, we show that bundling causes firms to specialize and the fit between employees skills and technological parameters to improve – a phenomenon we refer to as *polarization*.

Suppose there exists a central planner that can untie skills from workers and allocate them freely to firms in order to maximize output in the economy, something we label *full unbundling*. The planner would choose the amount of skills $X(\phi)$ allocated to each firm to maximize

$$\int F(X(\phi);\phi)\,\mathrm{d}H^f(\phi)$$

subject to the k feasibility constraints

$$\int X(\phi)\,\mathrm{d}H^f(\phi) = \bar{X}^w,$$

where $\bar{X}^w = \int x\,\mathrm{d}H^w(x) \in \mathbb{R}_+^k$ is the total amount of available skills for each of the k tasks. Unconstrained efficiency therefore requires that the marginal productivities are constant across firms, i.e., for any $j = 1, \ldots, k$, there exists μ_j such that

$$F_j(X^*(\phi);\phi) = \mu_j$$

for all firms ϕ. Full efficiency holds at the competitive equilibrium where k markets are opened. Because there are k markets, one for each skill, there are k prices. Let $p \in \mathbb{R}_+^k$ denote this vector of prices for each skill. Firm ϕ's demand for skills is given by maximizing $F(X;\phi) - p'X$ On the supply side, the total supply of skills is unchanged. However, each worker can split her entire supply of skills between an employing firm and the market, making individual labor supply *endogenous*.

The full unbundling equilibrium coincide with the bundling equilibrium studied in Section 2 if and only if there is complete bunching under that equilibrium, i.e., the bunching set consists of the whole population of workers. In this case, the wage schedule is fully linear. This happens trivially in the special case where all workers are specialists, i.e., when the distribution of workers is concentrated on the axes. Then, in effect, markets for workers and market for skills are one and the same. More generally, it follows from Proposition 3 that the equilibria under bundling and full unbundling coincide if and only if the distribution of workers' skills is a generalized mean-preserving spread of that of aggregate demand under unbundling, i.e., $W_{\#}H^w \succeq_C W_{\#}X^*_{\#}H^f$.

Assuming two tasks and a CES technology, we now characterize those workers benefiting from full unbundling and those harmed in the process.

Proposition 4. *Assume that the production function is given by (4) with $k = 2$. Except in the case where the wage schedule is linear under bundling (i.e., there is full bunching), some generalist workers ($0 < \theta < \pi/2$) are strictly better off after unbundling. If skills are complements ($\sigma < \eta$), some specialist workers ($\theta = 0$ and/or $\theta = \pi/2$) are strictly worse off.*

Proof. Let $w_i^b(\alpha) = w_i(\theta^b(\alpha))$ denote the implicit price for skill i perceived by the firms with factor intensity α in the bundling environment. The workers hired by those firms earn $w^b(\alpha) = w_1^b(\alpha)\cos\theta^b(\alpha) + w_2^b(\alpha)\sin\theta^b(\alpha)$ under bundling and $p_1\cos\theta^b(\alpha) + p_2\sin\theta^b(\alpha)$ under unbundling.[23] We start by studying the ratio of the wages under bundling and unbundling, $r(\alpha)$,

$$r(\alpha) = \frac{p_1\cos\theta^b(\alpha) + p_2\sin\theta^b(\alpha)}{w^b(\alpha)} = \frac{p_1 + p_2 t^b(\alpha)}{w_1^b(\alpha) + w_2^b(\alpha)t^b(\alpha)}, \tag{29}$$

where $t^b(\alpha) = \tan\theta^b(\alpha)$. If the ratio is greater than one, the workers with skill mix $t^b(\alpha)$ are better off after unbundling. To compute the derivative of r, we observe that from (10) and the envelope theorem

$$w'(\theta) = -w_1(\theta)\sin\theta + w_2(\theta)\cos\theta.$$

It follows that the derivative of r is given by

$$r'(\alpha) = (\theta^b)'(\alpha)\ \frac{w_1^b(\alpha)p_2 - w_2^b(\theta)p_1}{w^b(\alpha)^2}.$$

As $w_1^b(\alpha)$, and $w_2^b(\alpha)$ are respectively decreasing and increasing in α, the numerator of the above fraction decreases with α. It is zero for $\hat{\alpha}$ such that $w_2^b(\hat{\alpha})/w_1^b(\hat{\alpha}) = p_2/p_1$,

[23] The workers are not employed by the same firm in the two environments.

i.e., for the firms that hire workers with the same skill mix $\hat{\theta} = \theta^b(\hat{\alpha})$ under bundling and unbundling. This skill mix $\hat{\theta}$ is represented on Figure 6. It follows that the function $r(\alpha)$ is quasi-concave in α and achieves its maximum at $\hat{\alpha}$. Workers with skill mix $\theta^b(\hat{\alpha})$ are those who benefit the most (or suffer the least) from the unbundling of skills. The ratio $r(\alpha)$ is represented on Figure 7.

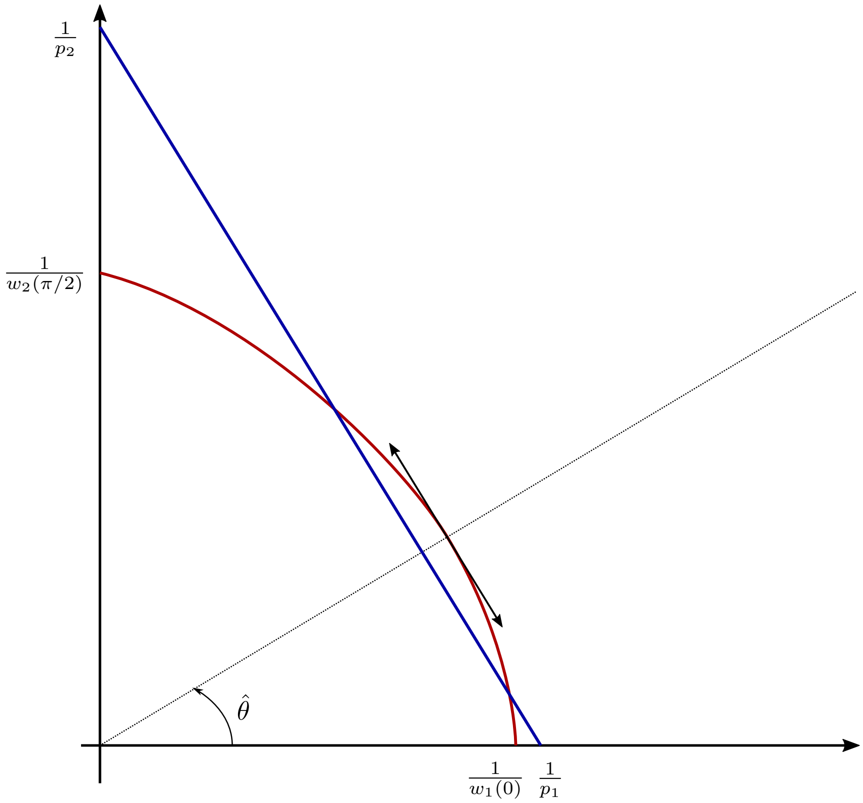


Figure 6: The iso-wage curves under bundling and unbundling are parallel for $\theta = \hat{\theta}$

In Appendix we show that a weighted average of the ratio $r(\alpha)$ is larger than one. Given the shape of $r(\alpha)$, this property guarantees that the workers of type $\hat{\theta}$ are indeed strictly better off under unbundling.

Finally, we prove the second part of the Proposition, assuming that skills are complements ($\sigma < \eta$). Suppose by contradiction that the two types of specialist workers are better off: $p_1 > w_1^b(0)$ and $p_2 > w_2^b(\pi/2)$. This would imply that $p_1 > w_1^b(\alpha)$ and $p_2 > w_2^b(\alpha)$ for all α. As by complementarity the demands for the two skills are both decreasing in p_1 and in p_2, this would imply that all firms would reduce their demand for both skills, which is impossible as the aggregate supply of skills remains unchanged.

In conclusion, except if full bunching prevails under bundling, at least some generalist workers are strictly better off and at least one type of specialist workers is worse off after unbundling. $\square$

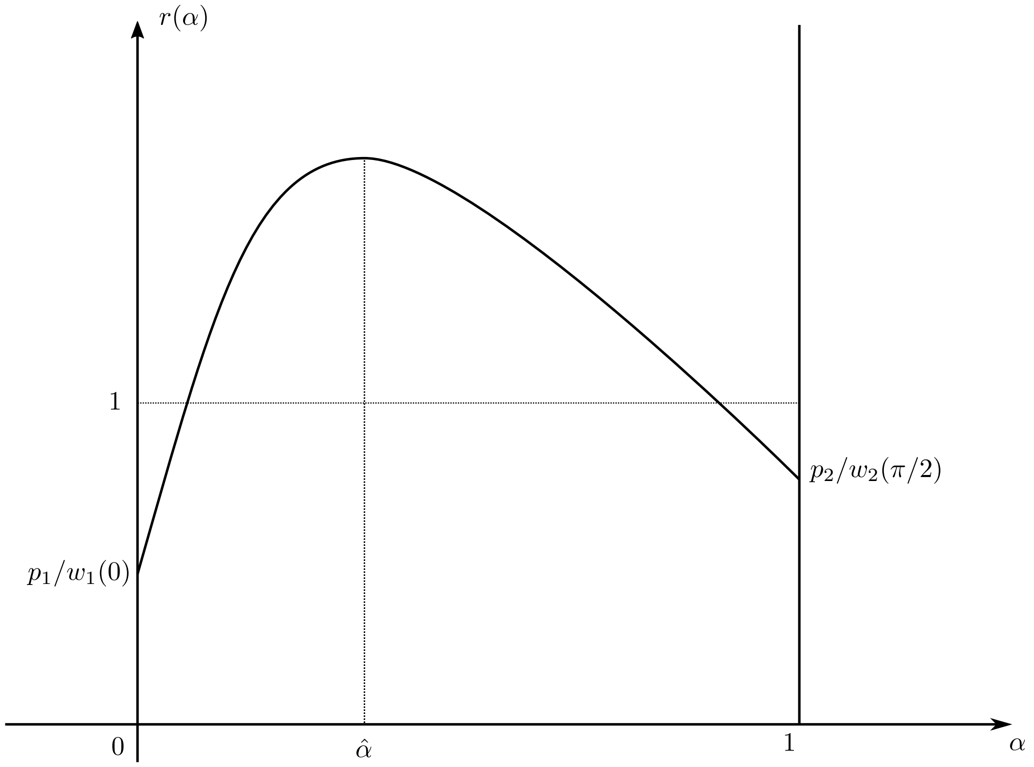


Figure 7: Ratio of the wages earned under bundling and unbundling by the workers who are hired by firm α under bundling

Proposition 5 (Polarization)**.** *After unbundling, specialized firms tend to specialize further, with their skill mixes being better aligned with their technologies, see Figure 8. Formally, there exists $\hat{\alpha}$ such that*

$$\begin{aligned} \theta^u(\alpha) &\le \theta^b(\alpha) \quad \textit{if } \alpha \le \hat{\alpha} \\ \theta^u(\alpha) &\le \theta^b(\alpha) \quad \textit{if } \alpha \ge \hat{\alpha}. \end{aligned} \tag{30}$$

Proof. Let $\hat{\alpha}$ be such that $\theta^b(\hat{\alpha}) = \hat{\theta}$. For $\alpha \le \hat{\alpha}$, we have, using (16)

$$\frac{1-\alpha}{\alpha}\left[\tan\theta^b(\alpha)\right]^{1-\sigma} = \frac{w_1(\theta^b(\alpha))}{w_2(\theta^b(\alpha))} \ge \frac{p_1}{p_2} = \frac{1-\alpha}{\alpha}\left[\tan\theta^u(\alpha)\right]^{1-\sigma}$$

which implies $\theta^b(\alpha) \ge \theta^u(\alpha)$, and proves (30). Figure 8 shows how sorting is affected by unbundling. Firms with a high relative intensity in a skill use relatively more of that skill after unbundling than in the bundling equilibrium. $\square$

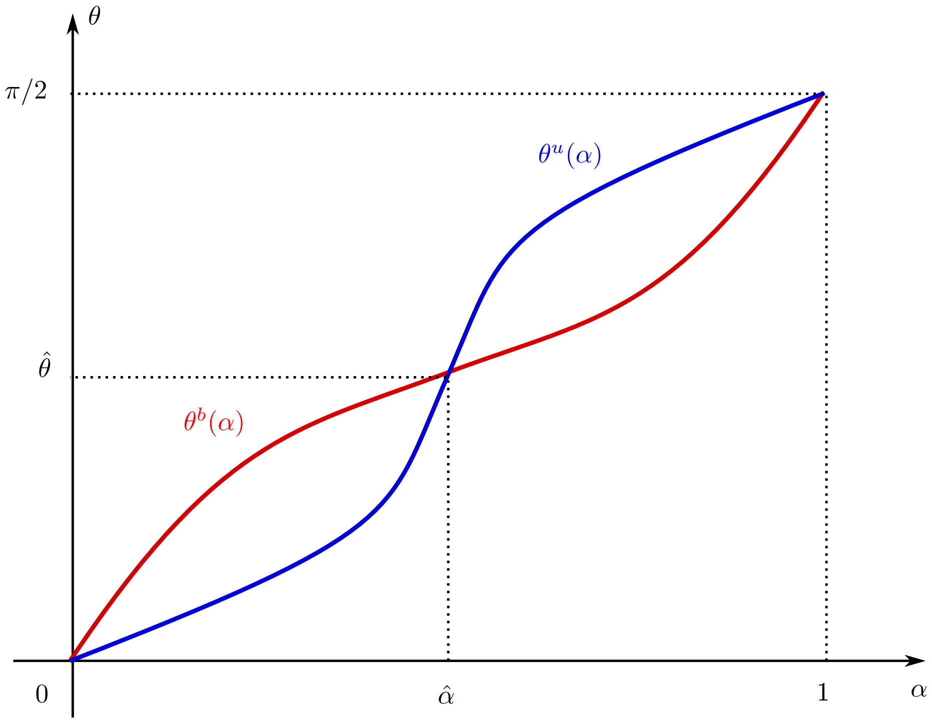


Figure 8: Polarization: Skill profile of labor force better aligned with the firm's core business. Core business of firm α is task 1 if $\alpha < \hat{\alpha}$, task 2 if $\alpha > \hat{\alpha}$. Workers-to-firms matching under bundling $\theta^b(\alpha)$ and unbundling $\theta^u(\alpha)$. Symmetric configuration $\hat{\alpha} = 1/2$, $\hat{\theta} = \pi/4$.

3.2 Costly unbundling

So far, we have assumed that the unbundling of skills is a costless process. However, if unbundling comes from an innovation (such as Uber which creates a market for driving skills), workers are likely to have to pay a fee or, more generally, incur a cost to have their skills unbundled. We therefore introduce wedges between the market wages paid to workers and prices paid by firms. Two interpretations for these wedges are possible:

1. There is one market price p_i^f for skill i, but workers incur a cost c_i per unit of unbundled skill i;

2. The platform(s) purchase(s) skill i from workers at price p_i^w and resell(s) it to firms at price p_i^f, with a margin c_i.

Consider first the bundling environment. When the difference between the maximum (implicit) price for a skill i and the minimum (implicit) price for the same skill **under bundling** is greater than the cost of unbundling skill i ($\max w_i^b(.) - \min w_i^b(.) > c_i$ for some skill i), then those workers employed by firms paying the (implicit) price $\min w_i^b(.)$

are paid "too little" for that skill. Indeed, they have an incentive to deviate and sell their skill i to those firms that use it intensively and are therefore ready to pay the most for it, namely the firms paying the (implicit) price $\max w_i^b(.)$. This arbitrage opportunity for workers employed in these low-paying firms generates a potential deviation that breaks the bundling equilibrium.

When all markets for skills are open (unbundling), with the associated wedge vector $\mathbf{c} = (c_1, \ldots, c_n)$, we denote by $w^u(x; \mathbf{c})$ the equilibrium wage schedule and by $w_i^u(\theta; \mathbf{c})$ the associated implicit price for skill $i = 1, \ldots, n$. For the reason explained just above, the maximal gap between two implicit prices for skill i cannot exceed c_i:

$$\max w_i^u(.; \mathbf{c}) - \min w_i^u(.; \mathbf{c}) \leq c_i.$$

If the range is exactly c_i, the market for skill i is active. The market price is $p_i^f = \max w_i^b(.; \mathbf{c})$ for firms and the market wage is $p_i^w = \min w_i^b(.; \mathbf{c})$ for workers. Workers implicitly paid p_i^w for skill i by their employer may indifferently sell all or part of their skill endowment on the market. All other workers, those who face an implicit price greater than p_1^w, choose to keep their skills for their firm and, therefore, choose to avoid the market.

By symmetry, the firms that face the implicit price p_i^f for skill i have an incentive to purchase skill i on the market. All other firms, those that face a lower implicit price for skill i, have zero demand in the market for skill i.

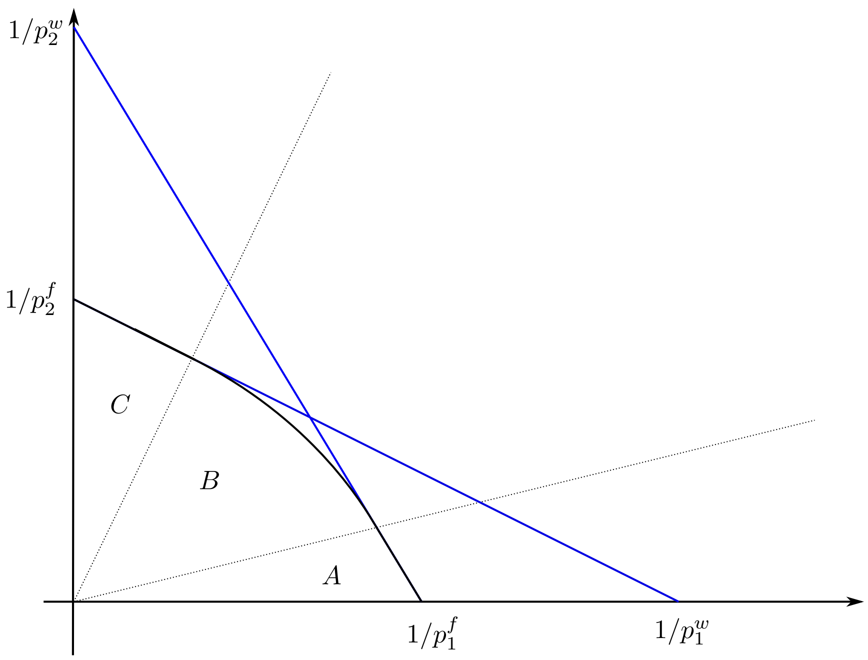


Figure 9: Unbundling equilibrium with wedges c_i and market prices $p_i^f = p_i^w + c_i$

The resulting wage schedule and implicit prices are shown on Figure 9 and 10 in the case of two skills. In region B, there is no arbitrage opportunity for workers, and in the absence of bunching in that region the implicit price equates demand and supply for each skill mix θ, exactly as in the case under bundling. By contrast, there is excess demand for skill 1 and excess supply for skill 2 in region A (see the structure of implicit prices). Workers in that region, being relatively underpaid for their skill 2 by their employing firms, supply skill 2 on the market. Whereas those employing firms have more demand for skill 1 than what their workers can offer, hence they purchase additional skill 1 on the corresponding market. The reverse is true in region C. Firms need more of skill 2. They buy it on the market using the supply coming from workers employed by firms in region A (see just above). And workers from region C sell their "unused" (by their employer) skill 1 on the market for that skill. The excess demand for skill 1 in region A is exactly matched by the excess supply for that skill in region C. The same holds for skill 2 between regions C and A.

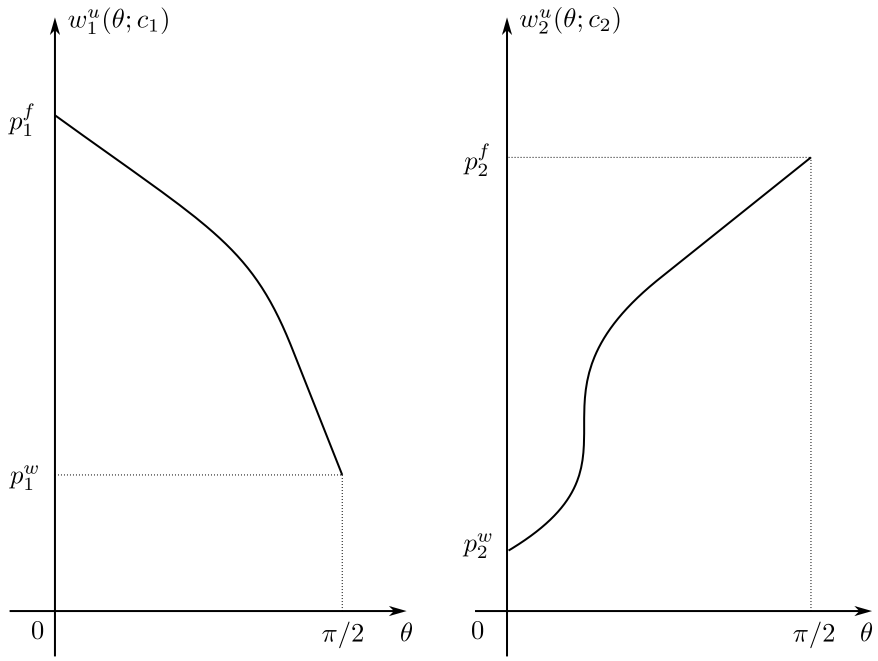


Figure 10: Implicit prices under unbundling with wedges c_i and market prices $p_i^f = p_i^w + c_i$

Proposition 6. *Under costly unbundling, the range of implicit prices for each skill satisfies:*

$$\max w_i^u - \min w_i^u \leq c_i, \tag{31}$$

where c_i is the cost incurred per unit of unbundled skill i. If a positive amount of skill i is traded on the market, then equality prevails in (31), with $p_i^f = \max w_i^u$ and $p_i^w = \min w_i^u$ being respectively the firm price and the worker price for that skill.

Only a subset of skills may be traded on markets: The configuration shown on Figure 9 is compatible with only one market being active. Suppose for instance that the ranges of the implicit prices for skill 1 and 2 satisfy $\max w_1^u(.;\mathbf{c}) - \min w_1^u(.;\mathbf{c}) = c_1$ and $\max w_2^b(.;\mathbf{c}) - \min w_2^b(.;\mathbf{c}) < c_2$. In this case, no market for skill 2 will open. The prices p_2^f and p_2^w, which do not exist, must simply be replaced with $\max w_2^b(.;\mathbf{c})$ and $\min w_2^b(.;\mathbf{c})$ on Figures 9 and 10. The workers in region A do not supply skill 2 on an external market. So the demand for skill 2 from firms hiring in region C must be covered by the supply of that skill from workers in the same region. In region C, however, the workers do supply skill 1 to region A firms, which hire workers with θ within A. In other words, a positive amount of skill 1 is transferred from region C to region A, but no transfer of skill 2 occurs in the opposite direction.

Same skill paid differently within a firm: The presence of wedges between firm and worker prices implies that contracted workers – those who supply one of their skill through the market – and employed workers – those who supply their skills bundle to a firm – are paid different prices for the same skill used at the same firm. Specifically, the workers whose types lie in Region A are "employed" and, hence, implicitly paid p_1^f for their skill 1 by their employers. The contracted workers with type in Region C, who supply some of their skill 1 to those firms through the market, are paid p_1^w, which is lower than p_1^f. The reverse is true in Region C for skill 2.

From bundling to unbundling: Considering a symmetric environment with two skills, we now explain how the wage schedule and the workers-to-firms matching pattern evolve as the unbundling cost falls from a high value (bundling world) to zero (full unbundling).

To run this comparative static exercise, we assume that the distributions $H^f(\alpha, z)$ and $H^w(\theta, \lambda)$ are symmetric with respect to $\alpha = 1/2$ and $\theta = \pi/2$ respectively. The unbundling costs are the same for the two skills. On Figure 11, we represent the wage schedules and the workers-to-firms matching curves for five values of that common cost: $c_0 = 0 < c_1 < c_2 < c_3 < c_\infty = \infty$. The polar cases $c = c_0$ and $c = c_\infty$ correspond to full unbundling and full bundling respectively. The iso-wage curves $w(\theta; c_j) = 1$ are represented on Figure 11(a), see A_jB_j for $j = 0, 1, 2, 3, \infty$. The sorting curves $\theta = \theta(\alpha; c_j)$ are represented on Figure 11(b).

Each iso-wage curve has linear parts (dashed lines) and a non-linear, strictly concave, part (solid line). There is full employment in regions where the curve is non-linear. In such regions, the equilibrium equation (23) holds. Firms do not use contracted workers because the marginal productivity of each skill is below the firm market price, $F_i < p_i^f$.

Employees do not sell skills on markets because the implicit wage (i.e. the productivity of the skill at their employing firm) is above the worker market price: $F_i > p_i^w$.

When the unbundling cost for the two skills c_3 is very large (schedule A_3B_3), a fraction of the (generalist) workers M_3N_3 and their employing firms are unaffected by unbundling. The schedules $A_\infty B_\infty$ and A_3B_3 coincide along M_3N_3. The linear parts of A_3B_3 do not intersect $A_\infty B_\infty$, i.e., the non-linear portion of A_3B_3 is *larger* than M_3N_3. As the unbundling costs decrease, the linear parts of the wage schedule become larger and the non-linear, strictly concave, part shrinks. The schedules A_2B_2 and $A_\infty B_\infty$ coincide and are tangent to each other at P_∞.

Under full unbundling, $c_0 = 0$, we have $p_1 = p_2$ by symmetry and from (16) the workers-to-firms matching is defined by

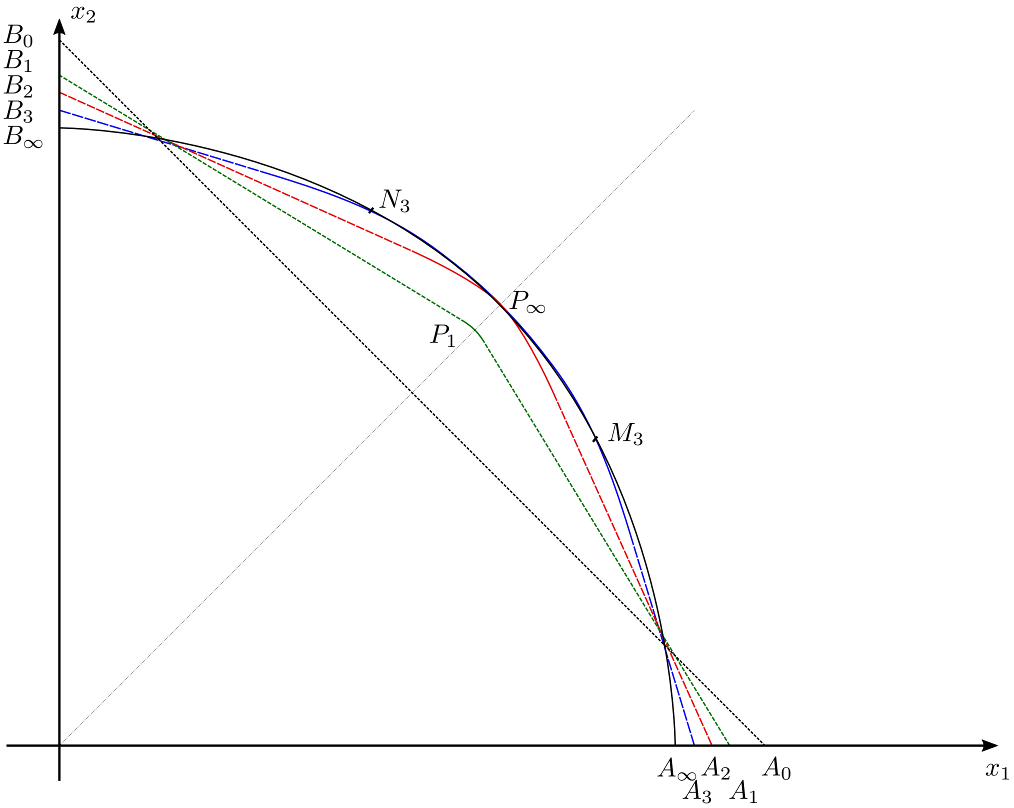
$$(\tan\theta(\alpha; c_0))^{1-\sigma} = \frac{\alpha}{1-\alpha},$$

see Figure 11(b).

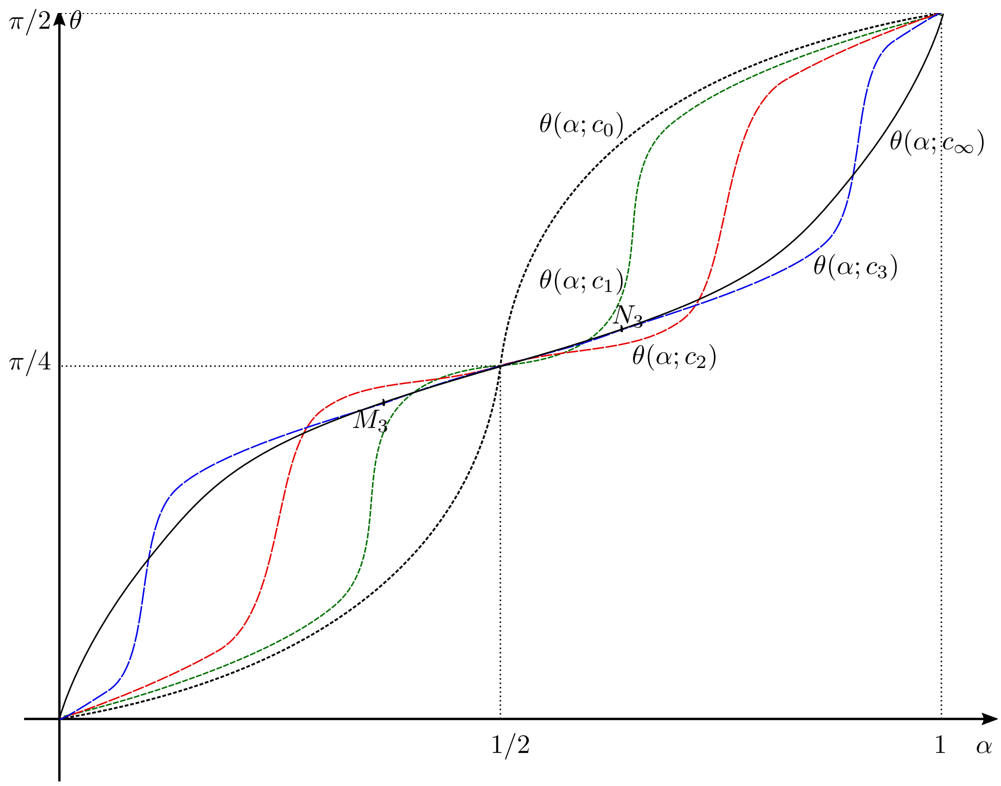
Consider the equilibrium equation (23) for $\alpha = 1/2$. By symmetry, the matching satisfies $\theta(1/2; c) = \pi/4$ for all values of the unbundling cost. On the left-hand side of (23), the supply-side factors $\Lambda^w(\theta(\alpha; c_j))h^w(\alpha)$ are the same for all those values. On the right-hand side, the demand factors $Z^f(\alpha)h^f(\alpha)$ do not depend on c_j either. The equilibrium equation thus depends on c_j through the slope of the matching $\theta'(1/2; c_j)$ and the wage $w(\pi/4; c_j)$.

When the cost is small, say $c = c_1$, the wage $w_1(\pi/4; c_1)$ is larger than the wage under bundling $w_\infty(\pi/4; \infty)$, and hence the matching is less steep than under bundling: $\theta'(1/2; c_1) < \theta'(1/2; \infty)$. This means that the fraction of firms with types α demanding generalist workers with profiles in $[\pi/4 - \mathrm{d}\theta/2, \pi/4 + \mathrm{d}\theta/2]$, namely $\mathrm{d}\alpha = \mathrm{d}\theta/\theta'(1/2; c)$, is higher under unbundling with cost c_1 than under bundling (infinite cost), hence generating a stronger demand for generalist workers under unbundling leading to a higher wage for these workers.

When the cost reaches c_2, the sorting curves is tangent to its bundling counterpart: $\theta'(1/2; c_2) = \theta'(1/2; \infty)$ and the wages of generalist workers are the same: $w(\pi/4; c_2) = w(\pi/4; \infty)$. When the cost exceeds c_2, say $c = c_3$, then the matching and the wage schedule coincide on a nontrivial interval M_3N_3.



(a) Iso-wage lines (dashed parts are linear): $w(\pi/4; c_0) > w(\pi/4; c_1) > w(\pi/4; c_2) = w(\pi/4; c_3) = w(\pi/4; c_\infty)$



(b) Matching: $\theta'(1/2; c_1) < \theta'(1/2; c_2) = \theta'(1/2; c_3) = \theta'(1/2; c_\infty)$

Figure 11: From bundling to unbundling (symmetric skills): Unbundling costs $c_\infty = \infty > c_3 > c_2 > c_1 > c_0 = 0$

4 The Empirical Content of Bundling and Unbundling

In this Section, we discuss the empirical content of our theoretical model.

4.1 Matching under Bundling: Within-Firm Heterogeneity

The two equations that define the equilibrium are (21):

$$\frac{F_1(\cos\theta^d(\alpha_2),\sin\theta^d(\alpha_2);\alpha,1)}{F_2(\cos\theta^d(\alpha_2),\sin\theta^d(\alpha_2);\alpha,1)}=\frac{w_1(\theta^d(\alpha_2))}{w_2(\theta^d(\alpha_2))} \tag{32}$$

and (23):

$$\Lambda^w(\theta^d(\alpha_2))\,h^w(\theta^d(\alpha_2))\,\frac{\mathrm{d}\theta^d}{\mathrm{d}\alpha_2}=Z^f(\alpha_2)h^f(\alpha_2)[\frac{F(\cos\theta^d(\alpha_2),\sin\theta^d(\alpha_2);\alpha,1)}{\tilde{w}(\theta(\alpha))}]^{1/(1-\eta)}. \tag{33}$$

where $\tilde{w}(\theta)=w(\cos\theta,\sin\theta)$.

Equation (23) relates the matching map $\theta^d(\alpha_2)$ implicitly given by (21) and its derivative $\mathrm{d}\theta^d/\,\mathrm{d}\alpha_2$ to the distributions of workers' skills and firms' technologies.

Hence, when the wage schedule w and the distribution of skills in the economy H^w are *known*, equation (23) defines the identification condition of the distribution of αs. When the supply of workers is given, i.e. the distribution $H^w(.)$, $\theta(\alpha)$ is known. However, the left-hand side comprises a complex combination of $H^f(z|\alpha)$ and $H^f(\alpha)$. Hence, the distribution H^f is not identified, only the left-hand side combination. As already mentioned, the above results apply to all homogenous production functions, in particular the last identification result. However, a more precise workers-to-firms matching condition can be obtained in the CES case (from (21)):

$$\left(\tan\theta^d(\alpha_2)\right)^{1-\sigma}=\frac{\alpha_2}{1-\alpha_2}\,\frac{w_1\left(\theta^d(\alpha_2)\right)}{w_2\left(\theta^d(\alpha_2)\right)}. \tag{34}$$

The matching between workers and firms is implicitly defined by the increasing function $\theta^d(\alpha_2)$. This function underlines how workers with different qualities but similar skill profiles may be employed within the same firm. Hence, as mentioned multiple times, this result implies some within-firm workers' heterogeneity. This is true in absence of bunching. This is even more valid when bunching obtains. We will come back to this point. But, first let us see how this result contrasts with other approaches and their associated results from the literature.

Absolute advantage ...: Indeed, recent work focusing on sorting (Lindenlaub (2017), Eeckhout and Kircher (2018)) predict perfect matching of high-quality workers to high-quality jobs (for the former) or firms (with quality defined in various ways in the latter). Perfect matching implies *no within-firm heterogeneity*: all workers employed in similar jobs or the same firm are identical. This sharp prediction has direct and important policy consequences in terms of productivity of an economy. Even Lindenlaub and Postel-Vinay (2020) who exhibit skill-specific ladders across jobs, having no firms, cannot talk to this question.

Indeed, despite recent path-breaking advances in the identification and estimation of "sorting" patterns in job-search models, the empirical results confirm the existence of *some* positive sorting (workers-to-firms matching, more precisely). But, estimated matching patterns are never as sharp as those predicted in the above models. Multiple reasons are likely to explain this absence of a strict matching: asymmetric information on workers' quality at entry in a job, imperfect monitoring of productivity on the job ... (see for instance Fredriksson, Hensvik, and Skans (2018) and their study of mismatch).

... Or comparative advantage: We take stock of these results and the associated explanations of deviations from perfect/absolute workers-to-firms matching. However, we also believe that there are deeper reasons (than imperfect or asymmetric information) for the observed dispersion of matching' quality or skill-set *within* a firm and occupation. Our model provides two such reasons. *First*, the equilibrium structure of matching in a bundling environment allocates workers to firms because of the workers' comparative advantages in a type of skill fitting the comparative advantage of the firm in a similar skill rather than the workers' absolute advantage and the firm's absolute advantage. Hence, in a two-types of skills environment, workers with an identical skills profiles x_2/x_1 but endowed with different quality levels (λ) may well work with the same employer. *Second*, when specific supply conditions prevail, "bunching" may occur. In this situation, a firm in order to achieve its optimal mix of skill types will hire workers situated between the two edges of the face that includes this optimal mix. Again, this equilibrium behavior generates within-firm and occupations workers' heterogeneity in skill-types and quality.

In absence of bunching, to test our model of comparative advantage against one of absolute advantage, and relying on data on two types of skills (typically, cognitive (C) and non-cognitive (N)), we must examine how workers' sorting between firms operates. In this world of comparative advantage, the ratio of skill C to that of skill N (measured in some non-parametric format) for any given worker i employed in firm j should be a constant, specific to the firm. Hence workers with different quality-levels but similar ratios should be co-workers, conditional on the occupation.

By contrast, when absolute advantage holds, matching on some non-parametric measure of quality λ should hold. And, there should be nothing left across firms explaining the matching pattern. Hence, co-workers should be endowed with similar quality levels, again conditional on the occupation.

Comparative advantage and bunching: When firms bunch skills to obtain their preferred mix, the firm employs workers with ratios of skills that may differ from that of their co-workers as well as from the firm j's aggregate ratio as long as such workers have skills that, once aggregated with those of co-workers, fit firm j's optimal matching $\theta(\alpha_j)$. Hence, i's co-workers' skills ratio will not be aligned with that of i. Bunching implies other restrictions on workers' wages that are discussed in the next subsection.

Link between z, firm's total factor productivity, and average quality of workers: Let us write Λ as $\Lambda = L\bar{\lambda}$ with $\bar{\lambda}$ the average quality of the firm's workers. Λ increases with z. High-z firms, which are also high-Λ, can achieve this high total quality through a large number of employees or/and a large average quality of its bundled workers. For a given size L of the firm, our model implies a positive association between high-zs and high-$\bar{\lambda}$. Again, our model does not imply homogenous quality of workers within a firm.

4.2 Wages under Bundling and Unbundling

Wage equation under bundling: Assume again that, when $k = 2$, skill 1 comprises all *Cognitive* skills, C, and skill 2 comprises all *Non-Cognitive* skills, N, as will be measured in the Swedish data. From Subsection 2.1, we know that the log-wage is the sum of a person component, $\ln \lambda$, and a firm component that reflects the equilibrium workers-to-firms matching pattern as a function of the firm's technology, $\theta(\alpha)$.

As noted above, this result is reminiscent of the additive decomposition of the log-wage into a person and a firm effect contained in Abowd, Kramarz, and Margolis (1999):

$$\ln w(x_C, x_N) = \ln \lambda(x_C, x_N) + \ln(w_C(\theta)\cos\theta + w_N(\theta)\sin\theta),$$

with the implicit prices of the two skills, $w_C(\theta)$ and $w_N(\theta)$, respectively decreasing and increasing in θ.

With homogenous production functions, the above firm-effect is independent of z, the firm's total factor productivity. However, in the general case of a non-homogenous production function, the profile of workers, θ, and the size of the firm, Λ, are given by the system (A.1 and A.2):

$$zF_C(\Lambda\cos\theta,\Lambda\sin\theta;\alpha,1)-w_C(\theta) \;=\; 0 \tag{35}$$
$$zF_N(\Lambda\cos\theta,\Lambda\sin\theta;\alpha,1)-w_N(\theta) \;=\; 0. \tag{36}$$

Hence, production isoquants are non-homothetic and F_C/F_N depends on Λ. Lemma 5 shows that total quality of the workers employed by firm (α, z), $\Lambda(\alpha, z)$ increases with z, firm's total factor productivity. Assuming that the marginal rate of technical substitution F_C/F_N increases with Λ, the equality $F_C/F_N = w_C/w_N$ implies that θ decreases with z. Put differently, the marginal productivity of *Cognitive* skills relative to that of *Non-Cognitive* skills increases with the size of firms; big firms use relative more *Cognitive* skills, implying that θ decreases with z.[24] Hence, the firm-effect now depends on z.

Identification of the wage equation is not straightforward. First, the usual strategy used to estimate the AKM decomposition is based on workers' mobility. However, *using workers' mobility to identify the firm-effect separately from the person-effect has no foundation* here since workers' matching to firms is perfect in the absence of bunching. The way to identify the two components is first to control for worker's quality (in some non-parametric format) and then identify the firm component across firms, hence by using the cross-sectional dimension. Notice though that, in contrast to the classical interpretation of a rent-sharing parameter (see Card, Cardoso, Heining, and Kline (2018), the firm component in the above equation does not necessarily capture value-added or sales or profits. It always captures a component of the firm's technology – the firm's reliance on *Cognitive* skills w.r.t. *Non-Cognitive* skills in its production technique – and may capture firm's total factor productivity z *when the production function is non-homothetic*. Hence, in the latter case, this dependence on z may induce a correlation with profits or value-added.

Finally, in zones where bunching takes place, the wage is linear in skills along the face. The firm's optimal mix is comprised between the two extremal points of the cone (see . Assuming that the face is "small" enough, then the difference between worker's individual (log-) wage and her (log-) quality will be close to the (log-) firm-effect as measured at the optimal mix. However, when the (linear) face of the equilibrium wage schedule is large enough, the AKM property is likely to be lost.

[24] See also Appendix A.9 when the iso-wage curve is an ellipse with $b > a$.

4.3 Polarization after Unbundling

Full unbundling: From an empirical perspective, there are two main messages from our analysis of full unbundling. First, generalists benefit from full unbundling when specialists are harmed. Furthermore, firms employing the former are hurt when firms employing the latter benefit from this opening of markets. Second, because firms can use all skills freely, they tend to increase their specialization in the direction of their comparative advantage, their preferred technology.

Figure 8 presents this tendency to specialization, which is akin to a *polarization*. Generalists were constrained by bundling in their ability to sell their skills, not any more in full unbundling whereas firms who employed generalists need to pay more for such workers in the unbundled world. All in all, firms become more polarized in their technological choice. In addition, when markets for skills open, the change in the equilibrium matching implies a change in the equilibrium composition of workers. Hence, a fraction of workers have to move to a new firm in which their comparative advantage fits that of firm's technology better under the new workers-to-firms matching equilibrium than under the old one.

Furthermore, with costly unbundling firms may employ workers endowed with an amount of, say, skill 1 and, at the same time, hire on the market for the same skill. The marginal price for this skill of those employed by the firm should be larger than the price paid, for the same skill purchased on the market, to the contracted workers.

5 Some Empirical Evidence

In this Section, we provide *preliminary* empirical evidence of our theory directly taken from Skans, Choné, and Kramarz (2021). It is not a full testing of its various components, both descriptive and structural, left for future research

5.1 The Data

5.1.1 Data overview

We use a data set measuring multidimensional skills of a large fraction of Swedish male workers. The data originate from the Swedish military conscription tests taken by most males born between 1952 and 1981.[25] The tests were taken at age 18 and the data should therefore be understood as capturing pre-market abilities. There are two main components; *cognitive abilities*, henceforth denoted as C, measured through a set of

[25] Although the share of test takers is lower in the final year, we have no reason to believe that this will interfere with our analysis. Our focus is not to compare workers across cohorts.

written tests and *non-cognitive abilities*, henceforth denoted as N, measured during a structured interview with a specialized psychologist. As noted in the introduction, the data have been used to assess labor market sorting in previous work, most notably by Fredriksson, Hensvik, and Skans (2018) and Håkanson, Lindqvist, and Vlachos (2020). Our definitions and set-up draw heavily on Fredriksson, Hensvik, and Skans (2018) (FHS, hereafter) in several dimensions.

Our data on employment cover the period 1996 to 2013. We include all workers with measured test results in ages 20 to 64. A large fraction of our analysis will be centered on *sorting*, hence on the allocation of workers, and *not* on the matching of workers to establishments. As a consequence, we will examine each worker's co-workers rather than each worker's employing establishment and its characteristics (productivity for instance). Furthermore, we include all workers in their main job in November as long as we measure the identifier of this establishment.[26] Our data on wages and occupations come from a firm-based sample which heavily over-samples large firms. These data cover 30 percent of private sector employees and all public sector employees. For the same set of workers, we also observe occupations. We can verify that our main wage results are insensitive to this sampling by using average monthly earnings, which we observe for all. For all observations, we only use one job per year.[27]

Our sorting analysis examines how workers are "grouped" across *Establishments*. But we also present results for *Jobs* defined as the intersection of the occupation (at the 3-digit level) and establishment of the worker as in FHS. All results are stable across these two definitions.

5.1.2 Defining generalists and specialists

The skills data are measured using an ordinal discrete (integer) scale ranging from 1 to 9. Standard practice in the literature is to treat these data as if continuous and cardinal after standardizing them to mean zero and standard deviation one within each birth cohort. We proceed differently and, whenever we can, instead strive to build our empirical strategies accounting for this discrete ordinal scale. We assume though that the ordinal scales have monotonic relationships to the underlying productive abilities they represent.

[26] An establishment is a physical place of work within one firm. About 10 percent of all workers do not have a fixed physical place of work and these are therefore not included.

[27] The preferences order is to first use observations where the wage can be observed. Wages are sampled in October or November. If there is no (unique) such observation, we select the observation with the highest earnings.

We use as our main empirical tool a classification of workers as *Generalists* or *Specialists* depending on the relationship between their two reported scores (trying to capture the skills ratio, x_1/x_2, defined in the theory Sections in the two skills case). As we are unable to precisely compare the two scales, we allow the data to "wiggle" one step before referring to workers as specialists and therefore count workers with less than a one-step difference between the scores as generalists. We thus heuristically define workers as *Generalists* if $abs(C_i - N_i) < 2$ and consequently define workers as *C-Specialists* if $C_i > N_i + 1$ and *N-Specialists* if $N_i > C_i + 1$. These definitions force us to assume that there is some shared relationship between the two scales (i.e the measures C_i vs. N_i) for each given worker i. On the other hand, the computation does not rely on any cardinal interpretation of differences along each of the scales.

Building on this worker-level classification, we classify establishments as a function of their workers' dominating type (*and not the employing firm's productivity since we examine workers' sorting rather than the workers-to-firms matching*). This classification does, according to the theory, inform us about α, i.e. the type of production function used by the establishment. To ensure that we do not generate any mechanical relationship between the measure of worker skills and this measure of skill-demand, we only use the *co-workers* when classifying establishments.[28] More precisely, an establishment is labelled a Generalist establishment when strictly more than 50% of co-workers are generalists or when it comprises exactly an identical number of C and N specialists.[29] As a consequence, a C-specialists' (resp. N-specialists') establishment has a strictly larger fraction of C-specialists (resp. N-specialists) co-workers. We call "Matched" workers those that are C-Specialists (resp. N-Specialists) in C-Specialists' (resp. N-Specialists) establishments.

For some of our analyses, we classify workers using their overall ability levels or "quality" (parameter λ in the theory). Therefore, we define workers as *low skilled* if the "sum" of (measured) cognitive and non-cognitive ability falls below 9 and *high-skilled* if the same sum is above 11 whereas the *mid skilled* are those where the sum is in-between. This classification is more cardinal in nature as the base is an accumulation of high and low values on to the inherently ordinal scale.[30]

[28]This means that the same establishment, in principle, can be classified differently for different workers within the same establishment (because the excluded worker is different).

[29]The second part of the definition takes account of small establishments. Essentially, the large ones never have an identical number of C and N specialists. In smaller ones, this allows us to have a larger number of specialists establishments. Results are essentially unaffected by small changes in this definition.

[30]This caveat should be kept in mind when interpreting the results but a mitigating factor may be that we only use this classification in contexts where we simultaneously account for the workers' specializations in the C/N dimension.

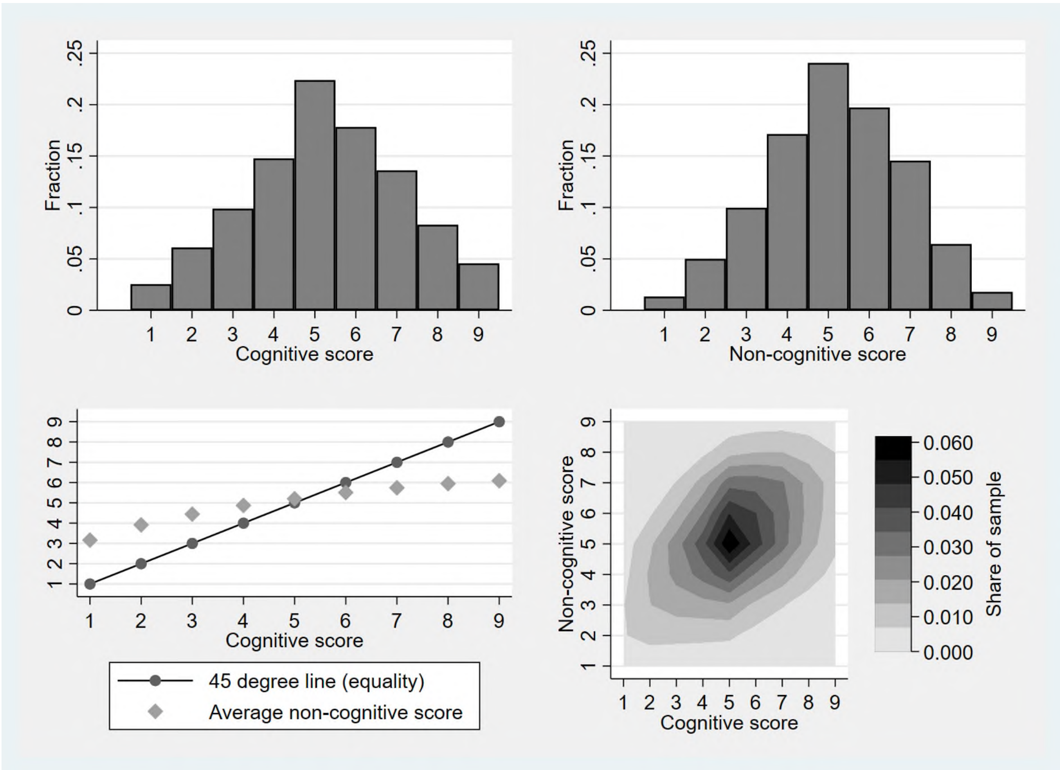


Figure 12: Measured ability scores

Note: The figure shows the test score results in our used data. See restrictions in the text. The bottom panels illustrate the joint distributions.

5.1.3 Descriptive statistics

Figure 12 depicts the joint distributions of the skills as reported on their 1-9 scale. The lower panels show the joint distributions. The two skill-types are correlated (correlation in 0.37 in the used data) but also contain independent information.

Table 1 shows descriptive statistics for the analysis sample. The first column shows the full analysis data. The average score lies marginally above 5 in both dimensions. Around half of the sample is classified as generalists (i.e. being on the diagonal of the joint distribution depicted in Figure 12) and about one quarter each are specialists in either the cognitive or the non-cognitive dimension. The following columns split the data in these three groups (generalists, C-specialists, N-specialists). As expected, the groups are equally distributed across years, ages, and birth cohorts. Cognitive skills are "twice" as large (6.9 vs. 3.6) among cognitive specialists than among non-cognitive specialists but, as discussed above, these scales do not have a natural interpretation in terms of the scores' productive content. The equivalent difference for non-cognitive skills is very similar (6.3 vs. 4.1). Furthermore, C-specialists tend to be over-represented within "highly skilled" workers. Still, all ability levels are present across the three categories. Since most workers are classified as generalists, most establishments are also dominated by generalists. And this also makes it more common for the generalists to be working in an establishment dominated by its own group (in that sense, "sorted"). The final column presents statistics for the part (a half) of the sample for which we can observe wages. As shown, this sample is nearly identical to the sample where we can observe occupations. Most importantly, the data are very similar to the first column (All) in all aspects (such as skill levels and composition), except for establishment size. The latter arises mechanically from an oversampling of large firms. Fortunately, we are able to check the stability of our wage results by estimating the same models for the earnings data that we observe for all.

Table 1: Descriptive statistics

	(1) All	(2) Generalist	(3) C-Specialist	(4) N-Specialist	(5) Wage obs
Year	2004.8	2004.8	2004.9	2004.7	2005.1
Cohort	1965.8	1966.0	1965.4	1965.8	1965.1
Age	39.0	38.8	39.5	39.0	40.0
Worker skills:					
Cognitive (C=1-9)	5.252	5.190	6.914	3.643	5.366
Non-cognitive (N=1-9)	5.179	5.206	4.090	6.267	5.239
$C+N$ low (< 9)	0.252	0.237	0.207	0.339	0.233
$C+N$ mid ($9-11$)	0.376	0.422	0.316	0.325	0.371
$C+N$ high (> 11)	0.371	0.341	0.476	0.336	0.396
Establishment size	82.1	81.9	88.2	76.0	118.4
Generalist establishment	0.767	0.777	0.722	0.787	0.782
Cognitive establishment	0.136	0.125	0.209	0.087	0.141
Non-cognitive est.	0.097	0.098	0.069	0.126	0.077
Matched	0.504	0.777	0.209	0.126	0.507
Observed occupation	0.517	0.514	0.539	0.503	0.978
Observed wage	0.529	0.526	0.551	0.513	1.000
ln(Wage)	10.182	10.182	10.227	10.131	10.182
ln(Earnings)	10.102	10.104	10.138	10.059	10.157
N	12,627,401	6,964,632	2,744,810	2,917,959	6,682,011

Note: Descriptive statistics for the used data covering 1996-2013. Establishments are restricted to be size 6 (i.e. 5 coworkers) to 600. In columns (2) to (4) we split the sample and according to if the worker is a Generalist, defined as $abs(C-N) < 2$ or a Specialist in C or N. Column (5) only uses workers for whom we have information on wages. Generalist establishments have a majority of employees as generalists, or an exactly equal share of specialists of the two types. Non-generalist establishments are classified according to the dominating type of specialists among employees. These classifications only use *co-workers*, i.e. not the subject himself. "Matched" workers are C-Specialists in Cognitive establishments (resp. N). Monthly earnings are recorded for all observations.

5.2 Workers' sorting

We are interested in analyzing how workers skills are related to some common (establishment-level) skill requirement. In the spirit of FHS, we will classify the establishments based on co-workers skill set as explained above (see subsection 5.1.2). We then regress the worker's skill type on the type of her co-workers. As a starting point, we only use one year (2005) and defer the analysis for trends over time to subsection 5.2.3. Thus, we estimate models of the following form:

$$Y_{ij}^{\tau} = \alpha + \lambda^{C,\tau} * C_j^{-i} + \lambda^{N,\tau} * N_j^{-i} + \epsilon_{ij} \tag{37}$$

where Y_{ij}^{τ} represent the type of worker i, employed at workplace j. Types will be captured by indicator functions for being a specialist of type $\tau = C, N$, or a generalist. C_{jt}^{-i} and N_{jt}^{-i} measures the share of co-workers that C-specialists and N-specialists (the residual type is generalists). If workers are (horizontally) sorted into firms where co-workers are of a similar type (because this is what the firm-level technology asks for, following our theory), we expect positive values on $\lambda^{C,C}$, but negative values on $\lambda^{C,N}$ for instance.

5.2.1 Simulating assignment principles

In this subsection, we contrast the sorting patterns observed in the data with patterns that would arise if workers were sorted according to three contrasted assignment principles. The first is random sorting. As noted in the literature on segregation, random assignment does not generate an even distribution of workers across jobs when units are small. This noise will be partly taken care of by using our "leave-out" approach in which we examine the co-workers' types for each individual worker within an establishment. The second assignment principle is sorting on absolute ability, where ability is proxied by $C + N$, consistent with better workers being sorted into similar firms (potentially more productive, something we do not examine here). This principle is related to positive assortative matching even though we prefer to use "vertical sorting" in this text. Third, we study assignment according to the *relative* strength of each ability as proxied by C/N following the above theory.

Two guiding principles are followed. First, and even though the skills are discretely measured in the data, we start by generating simulated raw continuous skills data that exactly aggregate up to the actual data in terms of number of workers with each combination of skills *and* which ensures that the correlations across skill types gets replicated within these types. Second, we keep the exact distribution of establishment sizes unchanged.

Next, we allocate workers into the observed establishment distribution (i.e. number of workers per establishment) using the simulated raw scores. To do so, we rank establishments in a random order. Then, we rank workers according to one of the three criteria (Random, PAM/vertical sorting, CK/horizontal sorting) and assign them to the establishments in this order. Hence, for vertical sorting, we rank the workers according to the sum of the (simulated) cognitive and non-cognitive abilities when for CK/horizontal sorting we divide the two scores and rank workers according to the resulting ratio.

This generates four different allocations (Actual, Random, PAM and CK) all of which have the identical number of workers per ability type, and an identical (real) establishment-size distribution.

Table 2: Leave-out mean regressions on worker types

	(1) Actual sorting	(2) Random sorting	(3) Sorting on $C+N$	(4) Sorting on C/N
Panel A:				
Dependent variable: Being N-specialist				
Coworker share of N-specialists	0.224	0.009	0.283	0.987
	(0.006)	(0.007)	(0.006)	(0.000)
Coworker share of C-specialists	-0.263	0.004	0.124	-0.005
	(0.004)	(0.005)	(0.005)	(0.000)
Constant	0.229	0.215	0.127	0.004
	(0.002)	(0.002)	(0.002)	(0.000)
Panel B:				
Dependent variable: Generalist				
Co-worker share of N-specialists	-0.023	-0.010	-0.417	-0.980
	(0.008)	(0.008)	(0.008)	(0.000)
Co-worker share of C-specialists	-0.155	-0.003	-0.423	-0.974
	(0.007)	(0.008)	(0.008)	(0.000)
Constant	0.593	0.555	0.740	0.990
	(0.003)	(0.003)	(0.003)	(0.000)
Panel C:				
Dependent variable: Being C-specialist				
Co-worker share of N-specialists	-0.201	0.001	0.134	-0.008
	(0.004)	(0.005)	(0.005)	(0.000)
Co-worker share of C-specialists	0.418	-0.001	0.299	0.978
	(0.007)	(0.007)	(0.006)	(0.000)
Constant	0.178	0.230	0.132	0.007
	(0.002)	(0.002)	(0.002)	(0.000)
Observations (all panels)	731,946	731,946	731,946	731,946

Note: Dependent variable is own type, estimates are for the share of co-workers of different types. Reference is the share of generalists. Data are for 2005. At least 6 workers and at most 600 workers with measured skills are employed in each establishment. Three last columns show regression on simulated allocations across the actual establishment size distribution, see text for details. Standard errors are clustered at the establishment level.

The results presented in Table 2 show that workers are systematically sorted across establishments, although not as strongly or one-dimensionally as suggested by the extreme absolute and random sorting scenarios. Each type of worker is more prevalent if there are more co-workers of the same type. Strikingly, there are less C-Specialists in establishments with many N-Specialists (and conversely). In terms of signs (although not magnitudes) this is exactly what is implied by the comparative advantage sorting scenario suggested by in the theory above.

5.2.2 Two-dimensional types

We use now a more detailed set of worker and establishment types by characterizing the workers and co-workers using the ability level combined with the skill type. We define workers as low skilled if the sum of cognitive and non-cognitive abilities falls strictly below 9 and high-skilled if the sum is strictly above 11 whereas the mid-skilled are those in-between. By combining these levels with the types for skill, i.e. generalists, C and N-specialists, we now have 9 types of workers. We run regressions based on equation (37) where we let each of these 9 types be the outcomes and the explanatory variables are the co-worker (leave-out) mean levels of these attributes. We start by estimating the impact of horizontal (specialists) and vertical (high/low) attributes separately (the results from the fully interacted model are presented in Skans, Choné, and Kramarz (2021)).

Table 3 show the resulting estimates. As clearly appears in column (1), panel A, high-level N-Specialists are employed together with high-level N-specialists as well as other high-ability workers, all other estimates are negative. The pattern repeats itself for high-level generalists in Column (2) of the same panel and for high-level C-specialists in Column (3). The following panels exhibit the same patterns for mid- and low-level workers. Although horizontal sorting appears to be stronger higher up in the ability ladder. The one single estimate that deviates somewhat is the positive association between N-specialists and mid-level generalists.

Overall, the results confirm that workers are sorted into establishments where their co-workers are of a similar type. Such results are fully consistent with employers having heterogeneous production functions that differ in their productive values of N and C skills.

Table 3: Leave-out mean regressions on two-dimensional worker types

	(1) N-Specialists	(2) Generalists	(3) C-Specialists
Panel A (High total ability). Dep. var. types:	High N-Specialist	High Generalist	High C-Specialist
Estimates:			
Co-workers N-Specialists	0.075***	-0.055***	-0.105***
	(0.004)	(0.004)	(0.003)
Co-workers C-Specialists	-0.098***	-0.027***	0.223***
	(0.004)	(0.006)	(0.006)
(reference: Generalists)			
Co-workers High ability	0.075***	0.329***	0.184***
	(0.004)	(0.006)	(0.004)
(reference: Mid ability)			
Co-workers Low ability	-0.078***	-0.127***	-0.039***
	(0.003)	(0.004)	(0.003)
Constant	0.072***	0.117***	0.023***
	(0.002)	(0.002)	(0.002)
Observations	731,946	731,946	731,946
Panel B (Mid total ability). Dep. var. types:	Mid N-Specialist	Mid Generalist	Mid C-Specialist
Estimates:			
Co-workers N-Specialists	0.083***	0.029***	-0.049***
	(0.004)	(0.005)	(0.003)
(reference: Generalists)			
Co-workers C-Specialists	-0.063***	-0.079***	0.096***
	(0.003)	(0.005)	(0.004)
Co-workers High ability	-0.078***	-0.211***	-0.027***
	(0.002)	(0.005)	(0.003)
(reference: Mid ability)			
Co-workers Low ability	-0.039***	-0.129***	-0.030***
	(0.003)	(0.005)	(0.003)
Constant	0.106***	0.355***	0.079***
	(0.002)	(0.004)	(0.002)
Observations	731,946	731,946	731,946

Notes: Results from 9 different regressions (table continues on next page) where the worker types are dependent variables. Types are defined from the combination of indicators for C/N-Specialists vs generalist combined with indicators for total ability being low, mid or high. Explanatory variables are co-worker averages of the C/N-specialists (generalists as the reference) and Low/High ability (mid ability as the reference). Data are for 2005. At least 6 workers and at most 600 workers with measured skills are employed in each establishment. Standard errors are clustered at the establishment level.

* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

Table 3: Leave-out mean regressions on two-dimensional worker types (cont'd)

	(1) N-Specialists	(2) Generalists	(3) C-Specialists
Panel C (Low total ability) Dep. var. types:	Low N-Specialist	Low Generalist	Low C-Specialist
Estimates:			
Co-workers N-Specialists	0.042***	-0.005	-0.016***
	(0.003)	(0.004)	(0.002)
Co-workers C-Specialists	-0.051***	-0.034***	0.032***
	(0.003)	(0.003)	(0.003)
(reference: Generalists)			
Co-workers High ability	-0.085***	-0.141***	-0.045***
	(0.002)	(0.003)	(0.002)
Co-workers Low ability	0.126***	0.263***	0.053***
	(0.003)	(0.004)	(0.002)
Constant	0.076***	0.126***	0.047***
	(0.002)	(0.002)	(0.001)
Observations	731,946	731,946	731,946
R-squared	0.030	0.053	0.006

Note: See note in previous table.

5.2.3 Sorting over time

In this part, we document how labor market sorting has changed over time. In doing so, we illustrate how the observed changes are consistent with the *unbundling* process outlined above. Because our data do not cover all cohorts, changes over time will also generate changes in the age-composition of our analysis sample. To eliminate spurious patterns, we follow Håkanson, Lindqvist, and Vlachos (2020) and focus on a specific age group that we can follow consistently over time (age 40 to 45) for the baseline analysis. We then document how the composition of their co-workers has evolved.

We estimate a version of equation (37) where the covariates of interest are interacted with time trends covering our 1996-2013 data period. The model accounts for year indicators and, for robustness tests, various plant-level controls. The model can thus be written as:

$$Y_{ijt}^{\tau} = \alpha + \theta^{C,\tau} * t * C_{jt}^{-i} + \theta^{N,\tau} * t * N_{jt}^{-i} + \lambda^{C,\tau} * C_{jt}^{-i} + \lambda^{N,\tau} * N_{jt}^{-i} + D_t + X_{ijt}\beta^{\tau} + \epsilon_{ijt}^{\tau} \tag{38}$$

where Y_{ijt}^{τ} represent the type of worker i, in year $t = Year - 2005$ employed at workplace j. Types will be indicators for being a specialist of type $\tau = C, N$, or a generalist. C_{jt}^{-i} and N_{jt}^{-i} measures the share of co-workers that are C-specialists and N-specialists (the residual type is generalists). D_t are time indicators and X_{ijt} are additional controls. We discuss now the results we expect to see *if* unbundling indeed took place over the sample period.

Because we constructed a three-type nomenclature of the skills space, with Generalists representing approximately 50% of the space, and each type of Specialist representing about 25%, most types of firms have a fraction of Generalists in them. This fraction is decreasing when the type of the firm specializes more into C-specialists or more into N-specialists (because of the optimal mix implied by its technology). Now, unbundling as seen from Section 3 implies a polarization: a firm's optimal mix moves closer to its axis of choice (more specialized into its "preferred" skill). Hence, when analyzing workers' sorting as we do now, firms mix less generalists (as captured by our definition) with their C or N specialists after a wave of unbundling. This polarization increases when the unbundling cost decreases, as time passes (see Figure 11(b)). Hence, we expect to obtain positive estimates for $\theta^{C,C}$ (i.e. a growing positive presence of co-worker of type C on Y_{ijt}^{C}) and $\theta^{N,N}$, but negative estimates for $\theta^{N,C}$ and $\theta^{C,N}$.

The estimates are displayed in Table 4. Panel A shows the estimates for the outcome Y_{ijt}^{C} and panel B for Y_{ijt}^{N}. Column (1) is the baseline specification without any controls except for time indicators. The estimates suggest that sorting has increased over time as C-specialists increasingly work with C-specialists and less with N-specialists. The converse is true for N-specialists. In column (2), we add controls for occupations. The sample here is a bit smaller as we do not observe occupations for all workers. The picture is, however, very similar. In column (3), we change the concept of co-workers and instead focus on other workers in the same *job* defined as occupation*establishment as in Fredriksson, Hensvik, and Skans (2018). Here the sample is reduced even further as we require that there are at least 5 other employees in the same job, but the estimated time-trends show a pattern similar to that obtained in the main specification. In Column (4), we return to the baseline model, but add controls for establishment size (8 groups) and for the share of low- and high-skilled workers in the establishment. The results remain robust. In Column (5), we remove low-tenured workers as in Fredriksson, Hensvik, and Skans (2018) without much change in results. Finally, in column (6), we widen the age span to also include workers aged 35 to 50 which makes the estimates more modest, although the qualitative results remain. Other evidence are presented in Skans, Choné, and Kramarz (2021), with results fully consistent with those given just above.

Table 4: : Specialist co-workers increasingly predict same-type specialists

Panel A Outcome: C-specialist	(1) Base	(2) Control for Occupation	(3) Coworkers in Job	(4) Additional Controls	(5) Only Tenured	(6) Broader Age Span
Time*C-spec.	0.008***	0.008***	0.006***	0.008***	0.009***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Time*N-spec.	-0.003***	-0.002	-0.003***	-0.003***	-0.003***	-0.001**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
C-specialists	0.415***	0.269***	0.455***	0.345***	0.349***	0.361***
	(0.006)	(0.008)	(0.008)	(0.006)	(0.007)	(0.005)
N-specialists	-0.203***	-0.122***	-0.241***	-0.168***	-0.171***	-0.170***
	(0.004)	(0.007)	(0.006)	(0.004)	(0.005)	(0.003)
Low-skilled cow.				-0.025***	-0.023***	-0.022***
				(0.004)	(0.005)	(0.002)
High-skilled cow.				0.109***	0.121***	0.114***
				(0.005)	(0.005)	(0.003)
N	2,317,898	1,255,003	896,931	2,317,898	1,656,627	8,787,016
Panel B Outcome: N-specialist	(1) Base	(2) Control for Occupation	(3) Coworkers in Job	(4) Additional Controls	(5) Only Tenured	(6) Broader Age Span
Time*N-spec.	0.004***	0.002	0.003**	0.004***	0.004***	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Time*C-spec.	-0.003***	-0.004***	-0.002*	-0.003***	-0.003***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
N-specialists	0.227***	0.144***	0.264***	0.200***	0.203***	0.208***
	(0.005)	(0.008)	(0.008)	(0.006)	(0.007)	(0.004)
C-specialists	-0.251***	-0.147***	-0.261***	-0.198***	-0.198***	-0.207***
	(0.004)	(0.006)	(0.005)	(0.004)	(0.005)	(0.003)
Low-skilled cow.				0.019***	0.019***	0.014***
				(0.004)	(0.005)	(0.003)
High-skilled cow.				-0.080***	-0.085***	-0.085***
				(0.004)	(0.004)	(0.003)
N	2,317,898	1,255,003	896,931	2,317,898	1,656,627	8,787,016

Notes: Dependent variable is a an indicator for being a C-specialist in panel A (N-specialist in Panel B). Subjects are 40 to 45 years old. Explanatory variables are share of co-workers that are C/N-specialists interacted with time. Normalised so that main effects of co-workers reflect 2005. All specifications include year dummies. Col (2) also controls for occupation dummies at the 3-digit level (sample requires that occupations are observed). Column (3) measures co-workers in job (occupation*establishment) instead (sample requires at least 5 co-workers in job). Columns (4) to (6) controls for eight plant size dummies and the share high/low skilled among co-workers. Column (5) only include workers with at least 3 years of tenure. Column (6) widens the age span to 35 to 50. Standard errors clustered at the establishment level. Data cover 1996-2013.
* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

5.3 Skills and Wages

In this subsection, we use our data to document how sorting relates to wages. In particular, we are interested in assessing the extent to which market returns to each skill are higher in settings where the technology is likely to use more intensively this exact skill.[31]

We again define the type of employer based on the share of each type of specialists that are employed by the establishment (see the definition in 5.1.2). As we are particularly interested in the sorting of specialists, we only include establishments where the majority of workers are specialists, and separate them into C and N establishments based on the dominating kind of specialists it employs. Thus, our data are drawn from the set of firms where the α-parameter in the production function is likely to correspond to a firm that employs a large fraction of either type of specialist. We then interact the type of the establishment with the specialization of the worker and estimate if the returns to being a C-intensive worker are higher if the employer uses a C-intensive technology (and conversely for N). To properly identify the interaction term net of the general returns to skill levels, the model controls non-parametrically for the level of skills in each dimension. Hence, the estimated model is:

$$\ln W_{ijt} = \alpha^C_{C(i)} + \alpha^N_{N(i)} + D^{N-plant}_{jt} + \lambda^N_j * D^{N-in-N}_{ijt} + \lambda^C_j * D^{C-in-C}_{ijt} + X_{ijt}\beta \qquad (39)$$

where $\ln W_{it}$ represents the (log-)wage of worker i in establishment j in year t and where the α's are indicators for each value of C and N skills. The two key variables of interest are the interaction terms D^{N-in-N} (for N-specialists in N-establishments) and D^{C-in-C} which captures the additional returns to N-skills in N-intensive employers, and C-skills in C-intensive employers, respectively. The vector of control variables will always include time and plant size indicators together with an age polynomial.

The results are presented in Table 5. Throughout, the results suggest that the wages in segments where employers rely intensively on C-skills also pay higher returns to these exact skills. Similarly, the results suggest a premium for N-skills in market segments dominated by N-intensive firms. These patterns are robust to controls for occupations, analyzing data at the job-level, controlling for very detailed skills, focusing on tenured workers, or restricting attention to the center year of 2005. In panel B, we show that the results are identical if we instead use monthly earnings, allowing us to expand the data set to include all observations rather than just the half for whom we observe wages.

[31] Some evidence in this direction at the *job*-level is presented in Fredriksson, Hensvik, and Skans (2018), with a focus on new hires, but here we revisit the issue at the *establishment* level for the *stock* of employees.

Table 5: Returns to specific skills are higher when co-workers are specialist in those skills

	(1) Base	(2) Control for Occupation	(3) Coworkers in Job	(4) Interacted Skills	(5) Only Tenured	(6) Only 2005
Panel A: Wages						
C-sp. in C-est.	0.027***	0.009***	0.040***	0.027***	0.028***	0.020***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.007)
N-sp. in N-est.	0.016***	0.005*	0.023***	0.021***	0.017***	0.009
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.007)
C-establishment	0.087***	0.020***	0.126***	0.089***	0.092***	0.075***
	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.006)
N	1,458,790	1,432,159	1,259,521	1,458,790	961,640	85,291
Panel B: Earnings						
C-sp. in C-est.	0.036***	0.009***	0.044***	0.036***	0.032***	0.033***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.007)
N-sp. in N-est.	0.023***	0.005*	0.026***	0.029***	0.025***	0.019***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.007)
C-establishment	0.081***	-0.002	0.108***	0.083***	0.082***	0.067***
	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)	(0.006)
N	2,945,409	1,432,159	1,259,521	2,945,409	1,899,162	168,815

Notes: Dependent variable is log wages. Control variables are the indicators for each C-skill (1 to 9) and N-skill (1 to 9), dummies for being a C- or an N-specialist, as well as year dummies, an age polynomial and eight plant size dummies. Displayed estimates are for C-specialists in C-establishments (and conversely for N-specialists). Sample excludes establishments where the majority of workers are generalists. Specialization of establishment is based on the specialization among co-workers. Column (2) adds controls for occupations. Column (3) performs the analysis at the job (occupation times establishment) level instead. Column (4) interacts the skills controls (C, N) into 81 groups Column (5) only include workers with at least 3 years of tenure. Column (6) zooms in on data for 2005. Panel A uses wages that only exist for a 50 percent sample. Panel B uses monthly earnings instead. Sample overlap when conditioning on observed occupations (col 2 and 3). Standard errors clustered at the establishment level. Data cover 1996-2013.
* $(p < 0.10)$, ** $(p < 0.05)$, *** $(p < 0.01)$

5.4 The growing wage of generalists

According to our theory, a process of "unbundling" should lead to an increase in generalists' wages when compared to those of specialists'. Indeed, the bundling constraint results in lower market wages for generalists when compared with the equivalent skills supplied by specialists. In order to test this prediction, we estimate wage regressions where our variable of interest is the interaction between time and an indicator for being a generalist (defined as above). The model controls for overall wage growth using year indicators. It also includes a fixed effect for each "detailed type" of worker, the type being defined as the interaction of the raw cognitive and non-cognitive scores (thus, 81 types). Our identification thus comes from the relative wage changes among workers on the generalists skill-diagonal relative to other types of workers. The model can be written as:

$$\ln W_{it} = \alpha_{CN(i)} + \theta^G * G_i * t + D_t + X_{ijt}\beta$$

where $\ln W_{it}$ represents the (log-) wage of worker i in year t, and where $\alpha_{CN(i)}$ is the fixed effect for the worker type. We estimate the model for 40 to 45 year old workers as above, and allow for a set of control variables X_{ijt} that will vary across specifications. We provide separate estimates for the sample of workers who are "well matched" (or, not bunched) in the sense that they work at an establishment where the own type is in majority among the work force.

The estimates are displayed in Table 6. Panel A shows the estimates for the overall population and Panel B zooms in on the "matched" sample (see again the definition in subsection 5.1.2). Column (1) is the baseline specification without any controls except for time indicators and the type-specific fixed effects. The estimates suggest that wages of generalists have grown more than wages for workers in general. The magnitudes suggest a modest 1.2 percent additional wage increase across one decade. In column (2), we add controls for occupations interacted with the worker type. In Column (3), we introduce a set of controls for competing time trends that interact each possible value of N and C with time (thus, 18 trends) as well as controls for establishment size (8 groups). In Column (4), we remove low-tenured workers and in column (5), we widen the age span to include all workers aged 35 to 50. Panel B uses the same set of specifications but only includes those workers who are employed in establishments where the majority of other workers are of the same broad type (Generalist, C-specialist, N-specialist). Estimates are unchanged in qualitative terms, but the magnitudes are at least twice as large, suggesting that wages of "matched" generalists have grown by 2-3 percent more across a decade than wages of matched specialists. This amounts to one-tenth of the average real wage growth during the period.

Table 6: Generalists' relative wages grows over time

Panel A All workers	(1) Base	(2) Control for Occupation	(3) Additional Controls	(4) Only Tenured	(5) Broader Age Span
Generalist interacted with time	0.0012***	0.0007***	0.0007***	0.0011***	0.0009***
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
N	1,281,151	1,255,003	1,281,151	928,127	4,723,064
Panel B Matched sample only					
Generalist interacted with time	0.0031***	0.0020***	0.0018***	0.0024***	0.0019***
	(0.0006)	(0.0004)	(0.0006)	(0.0007)	(0.0006)
N	654,687	641,005	654,687	476,688	2,415,481

Notes: Dependent variable is log wages. Subjects are 40 to 45 years old. Estimates are for interaction between year and a generalist indicator. All specifications include year indicators and control for 81 fixed effects for interactions between measured C (1 to 9) and N (1 to 9). Column (2) has more detailed fixed effects that also interact with occupation indicators at the 3-digit level (sample requires that occupations are observed). Column (3) controls for eight plant size indicators and 18 additional time trends, each interacted with one of the possible 9 values of C and N. Column (4) only includes workers with at least 3 years of tenure. Column (5) widens the age span to 35 to 50. Standard errors clustered at the establishment level. Data cover 1996-2013.
* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$).

6 Connecting Literatures

Up to now, we presented our model and the associated theoretical results as well as the exploration of some empirical counterparts to this theory. However, we also view our work as offering new connections between literatures that are rarely envisaged simultaneously. These "innovations" are presented in the following paragraphs, insisting particularly on the *Labor* and *IO* connection. To do so, we present the essential contributions that we believe to be related to what we have studied above.

To summarize, our theoretical contribution incorporates three ingredients – 1) a continuum of heterogeneous workers with multi-dimensional skill-types; these skills being either bundled or unbundled; 2) a continuum of firms with heterogeneous and multi-dimensional production functions in which the (intermediary) inputs are tasks; 3) tasks are obtained by (type by type) aggregation of workers' skills employed at the firm *rather than by the aggregation of workers' individual production.* A (potentially) non-linear wage schedule will allow the matching (sorting) of these multi-dimensional workers to their multi-dimensional firms within a general equilibrium framework (GE, hereafter).

We now examine in turn the various articles that incorporate some (but we believe not all) of these three ingredients. We insist most particularly on the (lack of) role played by firms in these contributions and try to contrast them with our approach which places firms center-stage.

The paper that prompted our investigation, Heckman and Scheinkman (1987), comprises two of our three ingredients; missing are the firms since their framework comprises n sectors (and identical firms within each sector, with the firms playing essentially no role). Heckman and Scheinkman (1987) were trying to understand whether bundling of skills (first ingredient) together with production obtained from an aggregation of workers' skills (third ingredient) could generate differential returns to the same skill in two different sectors. Indeed, explicit in our approach (through GE) as well as in theirs, workers choose the firm (for us) or sector (for them) that compensates them best, in the spirit of a Roy model. The answer to the above question is positive – returns to a given skill may differ across sectors – but general conditions for when this happens appear to be missing in their contribution. And the nature of the matching between workers and firms (sectors) is not their object of interest.

Bundling Multi-Dimensional Skills: Lindenlaub (2017) also contains two of our ingredients; missing is the firm-level production component since the model only comprises jobs. Conditions for sorting (through positive assortative matching, PAM, or its negative counterpart, NAM) are fully characterized in this multi-dimensional framework. The article also points to the connections of her problem to optimal transport theory (the so-called Monge-Kantorovich problem, see Villani (2009) for the mathematical theory, Galichon (2018) for applications to the economics of matching, and Peyré and Cuturi (2019) for computational optimal transport). Indeed, Chiappori, McCann, and Nesheim (2010) present the detailed mathematical connections between the type of questions studied in Lindenlaub (2017), optimal transport, and hedonic pricing. Lindenlaub and Postel-Vinay (2020) builds on Lindenlaub (2017) by adding random search to the initial sorting problem. This yields an extremely rich contribution in dimensions that we do not examine in the present article. Clearly, the search dimension brings important insights into skill-specific job ladders and the induced sorting of workers' skills bundles to jobs. However, and as in Lindenlaub (2017), the model is about jobs, not firms. Because Lindenlaub (2017) is an important step in the study of the matching of workers to jobs in this multi-dimensions context, we have related her results to ours directly within the body of our theory Sections (in particular on purity of the matching scheme, but not only).

Edmond and Mongey (2020) also examine a question directly related to ours, albeit within a macroeconomic framework. Their article has multiple connections to ours; they study a model with two tasks and two skills, with or without bundling of skills (using this word as we do). As in our approach, their workers are heterogeneous in their skill endowments. As in Murphy (1986) and Heckman and Scheinkman (1987), they have two firms in their economy (or, rather, two occupations). As we do here, each task (occupation, in their model) is produced from skills (using a CES function, in their model). Again, as we did above, output is produced using the supply of both tasks as inputs. Because they have two occupations producing output, the question of sorting of workers to the two occupations is the one they ask rather than sorting of workers across firms (or sectors as in Costinot and Vogel (2010)). And the general equilibrium they have to solve is very similar to that of Heckman and Scheinkman (1987). But, they very clearly and convincingly examine how unbundling operates, something that none of their predecessors had looked at. *Unbundling is indeed a central element of our paper.* But, even though Edmond and Mongey (2020) and our paper share some questions, their analytical framework has a macroeconomic perspective whereas our approach is mostly microeconomic in its focus on the multi-dimensional assignment/sorting of workers who are heterogeneous in their skill endowments to firms which are heterogeneous in their tasks needs, in *both* a bundling and an unbundling context. Hernnäs (2021) studies the consequences of bundling in a world where tasks can be automated, using a framework close to that of Edmond and Mongey (2020). The paper shows that skill returns in the automated task decline if tasks are gross complements. More generally, Hernnäs (2021) allows to examine automation in a richer setting than what was provided in the robotization literature.

Two very interesting contributions must be mentioned here.[32] In Rosen (1983), firms aggregate skills as we do here (or as Heckman and Scheinkman (1987) do), but bundling is not explicitly considered. However, the central question in Rosen's note is sorting/assignment of multi-skilled workers to firms (see his concluding paragraph). In his dissertation Murphy (1986), *Specialization and Human Capital*, Kevin Murphy examines very similar topics. His framework and some of his initial questions are very close to those of Heckman and Scheinkman (1987). However, the bulk of Murphy's thesis is concerned with how workers' investment decisions in skills prior to entering the labor force and workers' on-the-job investments affect their choice of specialization, two topics that Heckman and Scheinkman (1987) or ourselves do not study.

In Lazear (2009), skills are multi-dimensional as in our approach; some of these skills are general and some firm-specific. On the demand side, firms have different valuations of these skills. On the supply side, workers may choose their skill-set knowing that firms may need a skill-set comprising multiple skills. This setup generates a skill-weights view which is then used to understand the nature of wage losses when workers are displaced or forced to move. Lazear's perspective is unique in the labor literature, extremely original, but very far from our perspective in which firms aggregate (multi-dimensional) skills of their workers and where the (non-linear) equilibrium wage schedule is a central object of interest (see his footnote 16 where he differentiates the questions he asks with those of Heckman and Scheinkman (1987)).

Workers' Skills and Products' Characteristics: At this stage, it is useful to go back to Lancaster (1966) and Rosen (1974). Lancaster and Rosen were among the first to study questions related to those examined here. More precisely, the first of these two papers introduces the idea that goods are valued by consumers for their characteristics, and consumers do not derive utility from the good, per se. Furthermore, goods always possess multiple characteristics (some being shared by multiple goods), and finally combination of goods may possess characteristics that differ from those of the goods taken separately. The price function Lancaster (1966) investigates is linear in the characteristics. Rosen (1974) discusses this restriction very thoroughly (pages 37 and 38) in order to reject the constraints it imposes. His article examines the location of consumers in the product space as well as that of the producer, in a market equilibrium of pure competition. What Rosen proposes constitutes the foundation of hedonic pricing. It is a landmark contribution to both the analysis of consumption decisions and that of production decisions within industrial organization. To understand how Rosen's work is related to ours, essential to note that his consumers are our firms

[32] We would like to thank Simon Mongey for attracting our attention to them.

when his goods and products are our workers and their skills. In this respect, his problem is a mirror image of ours. However, from a modeling perspective, his approach is different from that of Heckman and Scheinkman (1987) or ours by not accounting for such "productive" interactions. In particular, he assumes "divisibility" in production in contrast to Lancaster (1966) (Rosen (1974), page 38). It is important to be more precise on divisibility.[33] In our approach, our assumption of bundling implies that any given worker cannot divide (or unpack) her skills between two firms (within-worker indivisibility). However, as clearly appears from our above example with two skills and supply on each of the axes and explicit in our production function, a firm can aggregate the skills of two workers or more, skill by skill. By contrast, Rosen (1974) focuses on products and their attributes, rather than skills. And, for such products, he clearly rules out such an aggregation, something he calls buyer's arbitrage (i.e. generating a new good by taking a linear combination of two goods' attributes) that would force the price of the product to be linear (page 37, last paragraph).[34] Hence, we allow for such arbitrage **across workers** *in portions of the skills space for which such arbitrage behavior is an equilibrium outcome* and where the equilibrium wage is linear. Because many components of Lancaster (1966)' theory are taken to be linear, such arbitrage possibilities are assumed rather than outcomes of an equilibrium behavior.

Giving Firms Substance: Our research is also inspired by a recent and important contribution, Eeckhout and Kircher (2018), in which assortative matching in so-called large firms is analyzed. In contrast to Lindenlaub (2017), workers in their approach have one dimension of skills (hence, one type). However, to obtain firms that are more than a collection of jobs, in line with firms being our main object of interest, they separate workers' quality from workers' quantity and assume constant returns to scale in those quantity variables. Then, their model gives a role to management that decides the firm's span of control by setting the firm's "resources". This allows them to study rich patterns of sorting in which quality and quantity dimensions both play a role. The resulting sorting condition combines four different dimensions: 1) complementarity between workers' and firms' qualities; 2) complementarity in workers' quantities and firms' resources; 3) span of control complementarity between manager's (firm's) quality and number of workers; and 4) complementarity between workers' quality and firms' resources. The production function adopted is $F(x, y, l, r)$ where x and y are quality variables (the types), respectively for the workers and the firms whereas l and r are quantity variables, respectively the number of workers and the firms' resources. In

[33] We restrict our use of divisibility to the skills bundling context.

[34] Two cars with 50 horsepower each are not equivalent to one with 100 horsepower is an obvious example.

addition, the production function has constant returns in both quantity variables, l and r. Hence, $F(x, y, l, r) = r * F(x, y, l/r, 1)$. In line with most of the literature, Eeckhout and Kircher (2018) define market equilibrium using $\int F(x, y, l, r)$, rather than $F(\int(.))$ as in Heckman and Scheinkman (1987) or as in the present article.

A point of interest to us is how the notion of efficiency units of labor à la Stigler (1961)) connects to the recent literature. As Eeckhout and Kircher (2018) explain, this notion reflects the assumption that "workers of a given skill are exactly replaceable by a number of workers of a different skill proportional to their skill difference: workers with half the skill level are perfect substitutes as long as there are twice as many of them." The Appendix A.11 of the latter paper shows that their one-dimensional setting with constant returns to scale in (worker and firm) quantity variables incorporates efficiency units of labor as a special case. In this case, their "sorting condition is satisfied with equality, (capturing) the well-known fact that sorting is arbitrary" (Eeckhout and Kircher (2018), p.102). However, we show that this is no longer true in our multi-dimensional framework with bundling; sorting between firm's technology and workers' bundles of skills emerges as a feature of the competitive equilibrium. Indeed, in our model, these efficiency units of labor are captured by our measure of one-dimensional worker quality – what we called the "vertical dimension" of the worker type. Our framework, however, incorporates multiple "horizontal" dimensions, namely the mix of worker skills. We find non-trivial sorting in these horizontal dimensions.

Comparative Advantage: A more macroeconomic literature studying trade, comparative advantage, and technical change also has connections with our approach. In Costinot and Vogel (2010), and as we do here, firms use workers to produce intermediate goods ("tasks" or "sectors"). The tasks are then combined into a final product. Firms that produce the final good use no labor and purchase their inputs on upstream markets. All firms operate under constant returns to scale and hence make zero profit. There is no heterogeneity across firms within sectors: all the firms that produce a given (intermediate or final) good share the same technology. Workers are heterogeneous in a single dimension. This allows them to study a Roy-like assignment model where high-skill workers have a comparative advantage in tasks with high-skill intensity. In equilibrium, this results in sorting between skills and tasks, in which each worker performs a single task.[35]

At this stage, it is also useful to spell out the differences between the bundling setting adopted here and the unbundling setting adopted in Costinot and Vogel (2010).

[35] Related to Costinot and Vogel (2010)'s super-modularity condition, Eeckhout and Kircher (2018) shows that sorting occurs provided that the complementarity between the quality (or vertical) dimensions of the firms and the workers is strong enough.

In both approaches, production of a final good requires a number of intermediate tasks to be completed. In our *bundling* regime though, there are no markets for tasks. The tasks are performed by the firms' employees as part of the firm's production process. Each employee contributes to all the intermediate tasks that are produced by the firm (or, more plausibly from an empirical standpoint a specific occupation). The workers are heterogeneous in their multi-dimensional skills (one skill-type per task) and the firms are heterogeneous in the technological intensities of the various tasks.[36] For the considered occupation and the associated set of firms which use this occupation for their output, sorting occurs between workers and firms, rather than between workers and tasks/sectors as is the case in Costinot and Vogel (2010). Firms with different technologies employ workers with different skill profiles, paying different implicit prices for the different tasks. The situation considered in Costinot and Vogel (2010) corresponds to our *full unbundling* regime, where firms and workers can trade all the types of skills on intermediate markets. We indeed show that the full opening of such markets makes the wage linear in skills, and skill-types and the resulting tasks can thus be thought of as intermediate inputs. Despite their differences, both frameworks deliver a role for sorting of workers to firms through a comparative advantage mechanism, one-dimensional for Costinot and Vogel (2010), multi-dimensional here (or in Lindenlaub and Postel-Vinay (2020)).

Old Tools, New Tools ...: The Industrial Organization (IO) literature is also deeply connected to our contribution, in particular its technical tools. We take stock of the screening literature when we study sorting. Non-linear pricing provides us with elements of the apparatus necessary to solve our "labor economics" problem. Stochastic dominance is also useful in what follows. However, whereas IO mostly relies on partial equilibrium concepts, we work within a General Equilibrium framework. Even though bundling is a word often used in IO, the way it is used here vastly differs from its meaning in the non-linear pricing literature. In the latter, sellers endowed with market power often find it optimal to grant rebates in return for the purchase of multiple units, which Wilson (1993) interprets as a form of (endogenous) bundling: the price charged for a "bundle" made of many units is lower than the sum of the prices of its components.

[36] Evidence abound that tasks for workers with similar occupations vary from firm to firm or even within firm for the same occupation. The nature of skills is also specified by contrasting cognitive and non-cognitive skills. A large literature uses such measures as O-NET to characterize occupations in terms of their (fixed) cognitive and non-cognitive content. Newly studied data sources help measure these two types of skills at the worker level. The distribution of such skills has been shown to vary across occupations, across firms, but also within firms (see Fredriksson, Hensvik, and Skans (2018), Håkanson, Lindqvist, and Vlachos (2020)).

Here, on the contrary, bundling appears as an exogenous friction (skills cannot be sold separately) in an otherwise competitive environment.

Our analysis also contributes to a totally different strand of the literature: the so-called many-to-one matching with transferable utility. A fraction of this literature has examined the problem in its discrete (game-theoretic) version whereas we work with a continuum of workers and a continuum of firms. Crawford (1991), Kelso and Crawford (1982), Hatfield and Milgrom (2005), and, more recently, Pycia (2012) and Pycia and Yenmez (2019) have contributed to this strand. Unfortunately, very few contributions – but Lindenlaub (2017) – examine many-to-one matching in the continuous types setting adopted here. However, in the spirit of many-to-one matching with a continuum and close to what we do here, Chiappori, McCann, and Pass (2017) examine multi-dimensional matching in which these multiple dimensions refer to the attributes on one or both sides of the market. They pay particular attention to problems in which the two dimensions of the market are not identical. In particular, they focus on the case where one side of the market is one-dimensional. In their contribution (see also Chiappori, McCann, and Pass (forthcoming) which only examines the one-dimensional case), they distinguish stable and pure matchings from others where purity imposes that all agents of type x on one side of the market match to the same $y = F(x)$ on the other side. We will see that, even though our matching solution is stable, it may not be pure. More important for us, they uniquely focus on problems with no "aggregate" production (in contrast to Heckman and Scheinkman (1987) and the present paper). Hence, some of their theorems do not apply to our setup. They also mention in passing the potential case when the multi-dimensional matching generates many-to-one patterns rather than one-to-one (again see Chiappori, McCann, and Pass (forthcoming)). Interestingly, these last authors connect their work to the multi-dimensional screening literature. Rochet and Choné (1998) investigate the multi-product monopolist problem in a setting where the dimensions of consumer heterogeneity and product attributes coincide. They find that consumers with different tastes may choose the same quality mix. This "bunching" phenomenon is due to a strong tension between participation constraints and second-order incentive compatibility conditions. Chiappori, McCann, and Pass (2017) show that the tension is inherent to the monopolist's market power and disappears together with bunching when perfect competition prevails. In the present paper, we find something akin to bunching in a competitive environment with multi-dimensional types where firms and workers have the same dimension of heterogeneity. Indeed as explained above, in any bundling equilibrium, each firm has a preferred mix of skill-types that depends on its productive characteristics. And firms with different characteristics have different optimal mix (full sorting between firm-types and optimal mix of workers' types). However, in conditions of workers' supply of skill-types that we characterize, this optimal mix can be achieved **only** by combining workers endowed with different skill-types. In this precise situation, firms of different types optimally hire workers endowed with the exact same type to achieve their (different) optimal mix; a phenomenon we call "bunching". The economic forces at work here – the bundling of skills and their aggregation in the production functions of firms – are entirely different from the screening mechanism of Rochet and Choné (1998), even though the mathematical characterization of bunching is similar in the two environments.

... and Very New Tools (WOT): The class of weak optimal transport (WOT) problems has been recently introduced by Gozlan, Roberto, Samson, and Tetali (2017). Given two probability measures μ and ν, and a cost function $c(\phi, m)$ that is convex in m, they consider the problem of minimizing

$$\int c(\phi, p^\phi)\,\mathrm{d}\mu(\phi) \tag{40}$$

over all couplings π of μ and ν, where p^ϕ is the (μ-almost surely unique) probability kernel such that

$$\mathrm{d}\pi(x, \phi) = \mathrm{d}p^\phi(x)\ \mathrm{d}\mu(\phi). \tag{41}$$

Gozlan, Roberto, Samson, and Tetali (2017) prove existence and duality results for Problem (40) under the main requirement that $c(\phi, m)$ is convex in m. The problem of maximizing total output in the economy, which is given by (3), has the same form as (40), with $\mu = H^f$, $\nu = H^w$, and the transport cost defined (for any given $x_0 \in \mathcal{X}$) by

$$c(\phi, m) = -F\left(\int x\,\mathrm{d}m(x); \phi\right) + F(x_0; \phi) + \nabla_x F(x_0; \phi).\left(\int x\,\mathrm{d}m(x) - x_0\right).$$

The above cost function is nonnegative by concavity of F in X. Under the equilibrium condition (2), minimizing (40) is equivalent to maximizing (3) because $\iint x\,\mathrm{d}p^\phi(x)\,\mathrm{d}\mu(\phi)$ equals $\int x\,\mathrm{d}\nu(x)$, which is a fixed and exogenous quantity.

Yet, as mentioned at the beginning of Section 2, the framework developed in the present article has an important difference with that of Gozlan, Roberto, Samson, and Tetali (2017). We do not impose here that the firms' demands for skill, $\mathrm{d}N^d(x;\phi)$, are *probability* measures as is required in the kernel decomposition (41). We only impose the weaker condition (2). Integrating the latter condition with respect to x yields

$$\int \bar{N}(\phi)\,\mathrm{d}H^f(\phi) = 1,$$

where $\bar{N}(\phi) = N^d(\mathcal{X};\phi)$ is the mass of the positive measure $N^d(x;\phi)$. In other words, the measure defined by $\tilde{H}^f(\phi) = \bar{N}(\phi)H^f(\phi)$ is a probability measure on Φ. Furthermore, by definition of $\bar{N}(\phi)$, the measure defined by $q^\phi(x) = N^d(x;\phi)/\bar{N}(\phi)$ is a probability measure on $\mathcal{X}$. We can thus rewrite the probability measure π that characterizes the matching between workers and firms as

$$\pi(x;\phi) = N^d(x;\phi)\, H^f(\phi) = q^\phi(x)\, \tilde{H}^f(\phi).$$

It follows that the probability measure $q^\phi(x)$ transports $H^w(x)$ onto $\tilde{H}^f(\phi)$. We are therefore back to the framework of Gozlan, Roberto, Samson, and Tetali (2017), with the major difference that the probability measure $\tilde{H}^f$ is now endogenous. In Choné, Gozlan, and Kramarz (2021), we extend the results of the latter paper to this context. In particular, we show that relaxing the kernel normalization $p^\phi(\mathcal{X}) = 1$ considerably enlarges the space over which the maximization takes place. We are nevertheless able to prove the existence of solutions for the primal and dual problems, as well as the duality equality:

$$\sup_\pi \int F\left(\int x\,\mathrm{d}N^d(x;\phi);\phi\right)\,\mathrm{d}H^f(\phi) = \inf_w \int \Pi(\phi;w)\,\mathrm{d}H^f(\Phi) + \int w(x)\,\mathrm{d}H^w(x),$$

where $\pi = N^d\,H^f$ and the firm's profit functions $\Pi(\phi;w)$ are given by (6).[37]

Getting back to the WOT framework, Gozlan and Juillet (2020) show that for quadratic costs $c(\phi,p^\phi) = \left(\phi - \int x\,\mathrm{d}p^\phi(x)\right)^2$ optimal plans are composition of a deterministic transport given by the gradient of a continuously differentiable convex function followed by a martingale coupling. Although our cost function is *not* quadratic, we have a similar composition pattern, with the deterministic part being the firm's aggregate skill function:

$$X^d(\phi) = \mathbb{E}(x|\phi) = \int x\,\mathrm{d}N^d(x;\phi).$$

The deterministic part, which satisfies the envelope condition (7), reflects the sorting of aggregate skill profiles, which itself comes from our single-crossing assumption (or twist condition). This suggests that the composition result established by Gozlan and Juillet (2020) for the quadratic transport cost can be generalized to a larger class of cost functions.

[37] The inequality is obvious in one direction as

$$\sup_{\pi=N^dH^f} \int F\left(\int x\,\mathrm{d}N^d(x;\phi);\phi\right)\,\mathrm{d}H^f(\phi) \le \inf_w \int \Pi(\phi;w)\,\mathrm{d}H^f(\Phi) + \int w(x)\,\mathrm{d}H^w(x)$$

follows from the very definition of the profit function $\Pi(\phi;w)$. (Use that $\Pi(\phi;w)$ is higher than the value obtained for the measure N^d that achieves the maximum at the left-hand side.)

References

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

Acemoglu, D., and D. Autor (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, vol. 4, pp. 1043–1171. Elsevier.

Acemoglu, D., and P. Restrepo (2018): "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 108(6), 1488–1542.

Autor, D. H. (2015): "Why are there still so many jobs? The history and future of workplace automation," *Journal of economic perspectives*, 29(3), 3–30.

Becker, G. S. (1973): "A theory of marriage: Part I," *Journal of Political economy*, 81(4), 813–846.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and labor market inequality: evidence and some theory," *Journal of Labor Economics*, 36(S1), S13–S70.

Chiappori, P.-A., R. McCann, and B. Pass (2017): "Multidimensional matching," *Working Paper*.

Chiappori, P.-A., R. J. McCann, and L. P. Nesheim (2010): "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 42(2), 317–354.

Chiappori, P.-A., R. J. McCann, and B. Pass (forthcoming): "Multi-to one-dimensional transportation," *Communications in Pure and Applied Mathematics*.

Choné, P., N. Gozlan, and F. Kramarz (2021): "Weak optimal transport with unnormalized kernels," On-going work.

Costinot, A., and J. Vogel (2010): "Matching and inequality in the world economy," *Journal of Political Economy*, 118(4), 747–786.

Crawford, V. P. (1991): "Comparative statics in matching markets," *Journal of Economic Theory*, 54(2), 389–400.

Edmond, C., and S. Mongey (2020): "Unbundling labor," Discussion paper.

Eeckhout, J., and P. Kircher (2018): "Assortative Matching With Large Firms," *Econometrica*, 86(1), 85–132.

Fredriksson, P., L. Hensvik, and O. N. Skans (2018): "Mismatch of talent: Evidence on match quality, entry wages, and job mobility," *American Economic Review*, 108(11), 3303–38.

Galichon, A. (2018): *Optimal transport methods in economics*. Princeton University Press.

Gozlan, N., and N. Juillet (2020): "On a mixture of Brenier and Strassen theorems," *Proceedings of the London Mathematical Society*, 120(3), 434–463.

Gozlan, N., C. Roberto, P.-M. Samson, and P. Tetali (2017): "Kantorovich duality for general transport costs and applications," *Journal of Functional Analysis*, 273(11), 3327–3405.

Håkanson, C., E. Lindqvist, and J. Vlachos (2020): "Firms and skills: the evolution of worker sorting," *Journal of Human Resources*.

Hatfield, J. W., and P. R. Milgrom (2005): "Matching with contracts," *American Economic Review*, 95(4), 913–935.

Heckman, J., and J. Scheinkman (1987): "The importance of bundling in a Gorman-Lancaster model of earnings," *The Review of Economic Studies*, 54(2), 243–255.

Hernnäs, S. (2021): "Automation, Task Content, and Occupations when Skills and Tasks are Bundled," Discussion paper.

Kelso, A. S., and V. P. Crawford (1982): "Job matching, coalition formation, and gross substitutes," *Econometrica: Journal of the Econometric Society*, pp. 1483–1504.

Laffont, J.-J., and D. Martimort (2009): *The theory of incentives: the principal-agent model*. Princeton university press.

Lancaster, K. J. (1966): "A new approach to consumer theory," *Journal of Political Economy*, 74(2), 132–157.

Lazear, E. (2009): "Firm Specific Human Capital: A Skill Weights Approach," *Journal of Political Economy*, 117(5), 914–940.

Lindenlaub, I. (2017): "Sorting multidimensional types: Theory and application," *The Review of Economic Studies*, 84(2), 718–789.

Lindenlaub, I., and F. Postel-Vinay (2020): "Multidimensional Sorting under Random Search," *Working Paper*.

Mandelbrot, B. (1962): "Paretian distributions and income maximization," *The Quarterly Journal of Economics*, 76(1), 57–85.

Murphy, K. M. (1986): "Specialization and Human Capital," Ph.D. thesis, University of Chicago.

Ostroy, J. M. (1984): "On the existence of Walrasian equilibrium in large-square economies," *Journal of Mathematical Economics*, 13(2), 143–163.

Peyré, G., and M. Cuturi (2019): "Computational optimal transport," *Foundations and Trends® in Machine Learning*, 11(5-6), 355–607.

Pycia, M. (2012): "Stability and preference alignment in matching and coalition formation," *Econometrica*, 80(1), 323–362.

Pycia, M., and M. B. Yenmez (2019): "Matching with externalities," .

Rochet, J.-C., and P. Choné (1998): "Ironing, sweeping, and multidimensional screening," *Econometrica*, pp. 783–826.

Rockafellar, R. T. (1970): *Convex analysis*, no. 28. Princeton university press.

Rosen, S. (1974): "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of Political Economy*, 82(1), 34–55.

——— (1983): "A Note on Aggregation of Skills and Labor Quality," *Journal of Human Resources*, 18(3), 425–431.

Skans, O. N., P. Choné, and F. Kramarz (2021): "The Empirics of Labor Market Sorting When Workers Have Bundles of Skills and Jobs Are Heterogeneous," Discussion paper.

Stigler, G. J. (1961): "The economics of information," *Journal of political economy*, 69(3), 213–225.

Strassen, V., et al. (1965): "The existence of probability measures with given marginals," *Annals of Mathematical Statistics*, 36(2), 423–439.

Villani, C. (2009): *Optimal transport: old and new*, vol. 338 of *Grundlehren der mathematischen Wissenschaften*. Springer Science, New York.

Wilson, R. B. (1993): *Nonlinear pricing*. Oxford University Press.

A Appendix

A.1 Existence of the equilibrium under bundling

Lemma A.1. *There exist Walrasian equilibria.*

Proof. As Eeckhout and Kircher (2018), we interpret our economy in terms of a classical exchange economy: "Consumers" in the classical model are firms $f = (\alpha, z)$. They consume bundles $(x_1, \ldots, x_k)$. We denote by n^d is the amount of numeraire it consumes. To make this an endowment economy, we assume that each firm is initially endowed with some of the workers and a sufficiently high level of the numeraire. The exact endowment of workers to firms does not matter because of the presence of the numeraire, so endowing each firm with the average distribution of workers would suffice. Firm preferences are represented by utility function $u(\,\mathrm{d}N^d, n^d; \phi)$:

$$u(\,\mathrm{d}N^d, n^d; \phi) = n^d + F(X^d; \phi),$$

with the aggregate skills X^d being given by (1). Because X^d is a linear function of $\mathrm{d}N^d$, the utility function is concave in $(\,\mathrm{d}N^d, n^d)$ and the firm's preference are convex. We can apply Ostroy (1984) for the existence of Walrasian equilibria. $\square$

A.2 CES technology and twist conditions

For the CES production function (4), we have

$$\nabla_\phi F(X; \phi) = (Y_0, Y_1, \ldots, Y_{k-1})',$$

with

$$Y_0 = (1/\eta)\left[\sum_{j=1}^{k} \alpha_j X_j^\sigma\right]^{\eta/\sigma} \quad \text{and} \quad Y_j = (z/\sigma) X_j^\sigma \left[\sum_{j=1}^{k} \alpha_j X_j^\sigma\right]^{\eta/\sigma - 1}$$

for $j = 1, \ldots, k-1$. It follows that

$$X_j^\sigma = (\sigma/z)\,(\eta Y_0)^{\sigma/\eta - 1}\, Y_j,$$

for $j = 1, \ldots, k$. The map $X \to \nabla_\phi F(X; \phi)$ is therefore invertible, hence injective.

A.3 Properties of the wage schedule

Proof of Lemma 1 To prove homogeneity, consider two workers with proportional skills x and λx for some $\lambda > 0$. These workers have the same relative skill endowments but differ in their overall quality, embodied by the multiplicative factor λ. Assume, by contradiction, that $w(\lambda x) < \lambda x$. Then no firm would hire worker type x as diminishing $N(x;\phi)$ by ε and increasing $N(\lambda x;\phi)$ by ε/λ leaves the firm aggregate skill unchanged while reducing the wage bill. It follows that the demand for worker x is zero, a contradiction. The reverse inequality, $w(\lambda x) > \lambda x$ is ruled out by the same argument.

Next, we show the wage schedule is quasi-convex. Suppose, by contradiction, that there exist worker types x, x', and x'' such that $w(x) = w(x') = 1$, $x'' = \nu x + (1-\nu)x'$, and $w(x'') > 1$. Then no firm would demand x'' as diminishing demand $N(x'')$ by ε and increasing $N(x)$ by $\nu\varepsilon$ and $N(x')$ by $(1-\nu)\varepsilon$ leaves the firm aggregate skill unchanged while reducing the wage bill.

Finally, since w is quasi-convex and homogenous of degree one, it is convex.

□

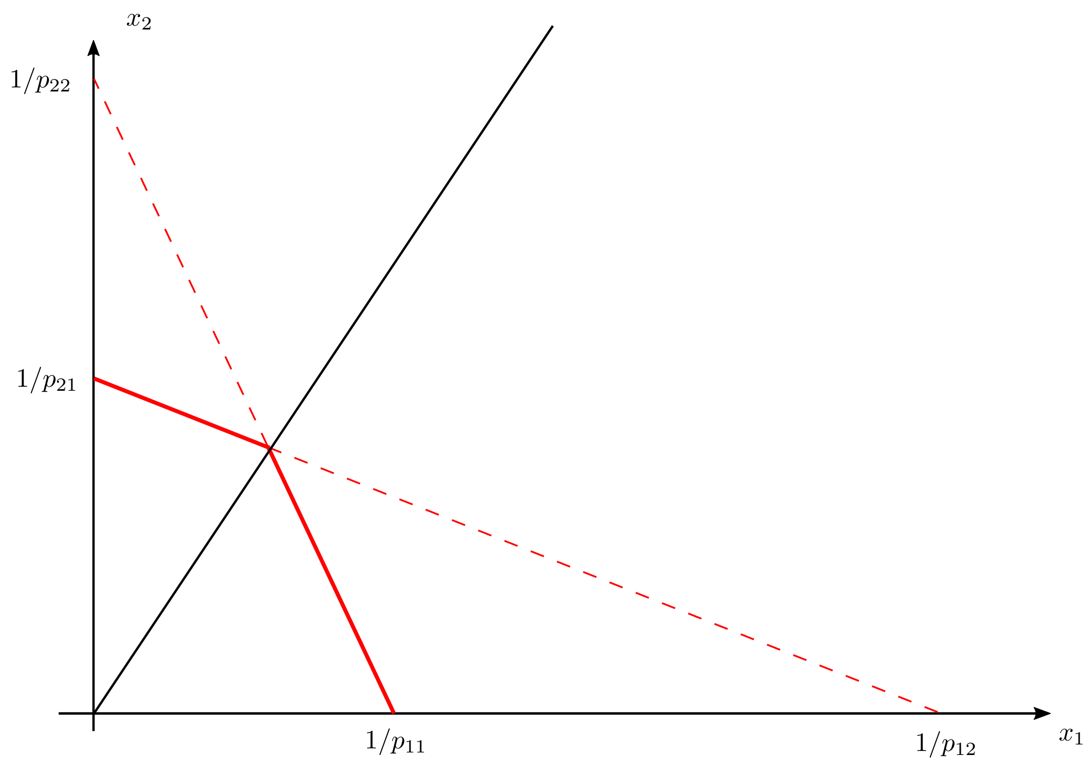


Figure A.1: Two-part wage schedule

Proof of (9) The first equality follows from Euler's homogenous function theorem. The second equality is a consequence of convexity:

$$\sum_{i=1}^{k} w_i(y)x_i = w(y) + \sum_{i=1}^{k} w_i(y)(x_i - y_i) \leq w(x).$$

Lemma A.2. *Let x_0 and x_1 be two distinct points in $\mathbb{R}_+^k$. The wage schedule is linear on $[x_0; x_1]$ if and only if the segment $[x_0/w(x_0); x_1/w(x_1)]$ is included in the iso-wage curve $\partial\mathcal{C}$.*

A.4 Proof of Lemmas 4 and 5

When there are two skills ($k = 2$), the average profile of the workers, θ, and their total quality, Λ, satisfy the first-order conditions

$$\begin{aligned} K_1(\theta, \Lambda) &\stackrel{\mathrm{d}}{=} zF_1(\Lambda\cos\theta, \Lambda\sin\theta; \alpha_2, 1) - w_1(\theta) &= 0 \qquad &\text{(A.1)} \\ K_2(\theta, \Lambda) &\stackrel{\mathrm{d}}{=} zF_2(\Lambda\cos\theta, \Lambda\sin\theta; \alpha_2, 1) - w_2(\theta) &= 0. \qquad &\text{(A.2)} \end{aligned}$$

where K_1 and K_2 are the first derivatives of the firm's objective $F(X;\phi) - w(X)$. Differentiating the first-order conditions (A.1) and (A.2) and inverting the Jacobian of K yields

$$\begin{pmatrix} \dfrac{\partial\theta}{\partial\alpha_2} & \dfrac{\partial\theta}{\partial z} \\ \dfrac{\partial\Lambda}{\partial\alpha_2} & \dfrac{\partial\Lambda}{\partial z} \end{pmatrix} = -\frac{1}{d} \begin{pmatrix} z\dfrac{\partial F_2}{\partial\Lambda} & -z\dfrac{\partial F_1}{\partial\Lambda} \\ -\left(z\dfrac{\partial F_2}{\partial\theta} - w_2'\right) & z\dfrac{\partial F_1}{\partial\theta} - w_1' \end{pmatrix} \begin{pmatrix} z\dfrac{\partial F_1}{\partial\alpha_2} & F_1 \\ z\dfrac{\partial F_2}{\partial\alpha_2} & F_2 \end{pmatrix}, \tag{A.3}$$

where d is the determinant of the Jacobian of $K = (K_1, K_2)$ in polar coordinates, i.e., the determinant of

$$\begin{pmatrix} \dfrac{\partial K_1}{\partial\theta} & \dfrac{\partial K_1}{\partial\Lambda} \\ \dfrac{\partial K_2}{\partial\theta} & \dfrac{\partial K_2}{\partial\Lambda} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial K_1}{\partial x_1} & \dfrac{\partial K_1}{\partial x_2} \\ \dfrac{\partial K_2}{\partial x_1} & \dfrac{\partial K_2}{\partial x_2} \end{pmatrix} \begin{pmatrix} -\Lambda\sin\theta & \cos\theta \\ \Lambda\cos\theta & \sin\theta \end{pmatrix}.$$

By concavity of the firm's problem, the determinant of the first matrix at the right-hand side is positive, hence $d < 0$.

First, we prove Lemma 5. The derivative of total quality with respect to total factor productivity is

$$\frac{\partial\Lambda}{\partial z} = -\frac{1}{d}\left[F_2\left(z\frac{\partial F_1}{\partial\theta} - w_1'\right) - F_1\left(z\frac{\partial F_2}{\partial\theta} - w_2'\right)\right]. \tag{A.4}$$

Consider the bracketed terms in (A.4). The first term $F_1 w_2' - F_2 w_1' = w_1 w_2' - w_2 w_1'$ is positive because the w_2/w_1 increases with θ by concavity if the iso-wage curve. The second term $F_2 \partial F_1/\partial\theta - F_1 \partial F_2/\partial\theta$ is positive by convexity of the production isoquants. It follows that the bracketed terms is positive and hence that Λ increases with z.

Second, the derivative of skill profile with respect to technological intensity is

$$\frac{\partial\theta}{\partial\alpha_2} = -\frac{z^2}{d}\left[\frac{\partial F_1}{\partial\alpha_2}\frac{\partial F_2}{\partial\Lambda} - \frac{\partial F_2}{\partial\alpha_2}\frac{\partial F_1}{\partial\Lambda}\right].$$

If production isoquants are homothetic, we have $F_1 \partial F_2/\partial\Lambda = F_2 \partial F_1/\partial\Lambda$. Because F_1 and F_2 decrease with Λ, we get from Assumption 2 that

$$\frac{\partial F_1}{\partial\alpha_2}\frac{\partial F_2}{\partial\Lambda} - \frac{\partial F_2}{\partial\alpha_2}\frac{\partial F_1}{\partial\Lambda} \geq 0. \tag{A.5}$$

which together with $d < 0$ yields $\partial\theta/\partial\alpha_2 > 0$.

Third, the determinant of the sorting matrix at left-hand side of (A.3) is positive because by concavity of the firm problem and Assumption 2 the two matrices at the right-hand side have a negative determinant.

A.5 Proof of Proposition 2

For any test function h, we have

$$\begin{aligned}
< W_\# X^d_\# H^f, h > &= \int_\phi h\left(\frac{X^d(\phi)}{w(X^d(\phi))}\right) w(X^d(\phi))\,\mathrm{d}H^f(\phi) \\
&= \int_\phi h\left(\frac{X^d(\phi)}{w(X^d(\phi))}\right) \int_x w(x)\,\mathrm{d}N^d(x;\phi)\,\mathrm{d}H^f(\phi) && \text{(A.6)} \\
&= \int_x \int_\phi h\left(\frac{x}{w(x)}\right) w(x)\,\mathrm{d}N^d(x;\phi)\,\mathrm{d}H^f(\phi) && \text{(A.7)} \\
&= \int_x h\left(\frac{x}{w(x)}\right) w(x)\,\mathrm{d}H^w(x) && \text{(A.8)} \\
&= < W_\# H^w, h > .
\end{aligned}$$

Equation (A.6) follows from Lemma 2. Equation (A.7) uses that $X^d(\phi)/w(X^d(\phi)) = x/w(x)$ for all x in the support of $\mathrm{d}N^d$, i.e., for all x proportional to $\tilde{X}^d(\alpha)$. Equation (A.8) uses the equilibrium condition (2).

A.6 Proof of Proposition 3

For any convex test function h, we have, by Jensen inequality

$$h\left(\frac{X^d(\phi)}{w(X^d(\phi))}\right) = h\left(\frac{\int (x/w(x)w(x)\,\mathrm{d}N^d(x;\phi)}{w(X^d(\phi))}\right) \leq \int h\left(\frac{x}{w(x)}\right) w(x)\,\mathrm{d}N^d(x;\phi),$$

which yields inequality rather than equality in (A.7) .

A.7 Demand for CES with two tasks

Consider the CES production function (4).

$$\begin{cases} X_1^d(p_1,p_2;\alpha,z) = \left[\dfrac{(1-\alpha)z}{p_1}\right]^{\frac{1}{1-\eta}} (1-\alpha+\alpha t^{\sigma})^{\frac{\eta-\sigma}{\sigma(1-\eta)}} & \text{(A.9a)} \\ X_2^d(p_1,p_2;\alpha,z) = \left[\dfrac{\alpha z}{p_2}\right]^{\frac{1}{1-\eta}} \left(\alpha+(1-\alpha)t^{-\sigma}\right)^{\frac{\eta-\sigma}{\sigma(1-\eta)}}, & \text{(A.9b)} \end{cases}$$

where $t = X_2/X_1$. From the sorting condition (16), we can replace $X_2/X_1 = \tan\theta(\alpha)$ with its value expressed in terms of the implicit prices faced by the firm. Under unbundling with no wedge, those prices do not depend on α. The aggregate skills at firm α can be expressed in terms of the implicit prices

$$X_1^d(\alpha,z) = (1-\alpha)^{\frac{1}{1-\sigma}}\, z^{\frac{1}{1-\eta}}\, p_1^{-\frac{1}{1-\eta}}\, p_2^{\frac{\sigma-\eta}{(1-\eta)(1-\sigma)}} \left[\alpha^{\frac{1}{1-\sigma}} p_1^{\frac{\sigma}{1-\sigma}} + (1-\alpha)^{\frac{1}{1-\sigma}} p_2^{\frac{\sigma}{1-\sigma}}\right]^{\frac{\eta-\sigma}{\sigma(1-\eta)}} \tag{A.10}$$

Similarly for skill 2

$$X_2^d(\alpha,z) = \alpha^{\frac{1}{1-\sigma}}\, z^{\frac{1}{1-\eta}}\, p_2^{-\frac{1}{1-\eta}}\, p_1^{\frac{\sigma-\eta}{(1-\eta)(1-\sigma)}} \left[\alpha^{\frac{1}{1-\sigma}} p_1^{\frac{\sigma}{1-\sigma}} + (1-\alpha)^{\frac{1}{1-\sigma}} p_2^{\frac{\sigma}{1-\sigma}}\right]^{\frac{\eta-\sigma}{\sigma(1-\eta)}}. \tag{A.11}$$

It follows that the demand for skills exhibits complementarity, i.e., the demand of skill 1, X_1^d, *decreases* with the price of skill 2, p_2, if and only if $\sigma < \eta$. If $\sigma = \eta$, the two skills are independent. If $\sigma > \eta$, the skills are substitutes (X_1^d increases with p_2).

Under bundling, the uniform prices p_1 and p_2 in (A.10) and (A.11) must be replaced with $w_1^b(\alpha)$ and $w_2^b(\alpha)$.

If we assume that the equilibrium wage is *a circle in polar coordinates*: $w(\theta) = w(\cos\theta, \sin\theta) = 1$. In this case, the wage decomposition in logarithms only comprises a person effect (λ), with no firm effect (since $w(\theta)$ is a constant). In this case, the sorting equation yields:

$$\frac{1-\alpha}{\alpha}\,(\tan\theta)^{1-\sigma} = \frac{1}{\tan\theta} \tag{A.12}$$

Hence

$$\theta(\alpha) = \arctan[(\frac{1-\alpha}{\alpha})^{\frac{1}{\sigma-2}}] \tag{A.13}$$

A.8 Proof of Proposition 4

We prove here that a weighted average of the ratio $r(\alpha)$ is larger than one. To do this, we compute the quantity $p_1\bar{X}_1^w + p_2\bar{X}_2^w$ in two different ways. We first consider the bundling environment. Under bundling, the aggregate demand for skill 1 of firms with factor intensity α is

$$\begin{aligned} X_1^b(w_1^b(\alpha), w_2^b(\alpha); \alpha) &= D(\alpha) \frac{(1-\alpha)^{\frac{\eta}{\sigma(1-\eta)}}}{(w_1^b(\alpha))^{\frac{1}{1-\eta}}} \left[1 + \frac{\alpha}{1-\alpha}(t^b(\alpha))^\sigma\right]^{\frac{\eta-\sigma}{\sigma(1-\eta)}} \\ &= D(\alpha) \frac{(1-\alpha)^{\frac{\eta}{\sigma(1-\eta)}}}{(w_1^b(\alpha))^{\frac{1}{1-\eta}}} \left[1 + \frac{w_2^b(\alpha)}{w_1^b(\alpha)} t^b(\alpha)\right]^{\frac{\eta-\sigma}{\sigma(1-\eta)}} \\ &= D(\alpha) \left[\frac{1-\alpha}{w_1^b(\alpha)}\right]^{\frac{\eta}{\sigma(1-\eta)}} \left[w_1^b(\alpha) + w_2^b(\alpha)t^b(\alpha)\right]^{\frac{\eta-\sigma}{\sigma(1-\eta)}}, \end{aligned} \tag{A.14}$$

where $D(\alpha) = \int_z z^{1/(1-\eta)} h^f(z|\alpha) h^f(\alpha)$ and (A.14) uses the sorting condition (16). Shortening the notation $X_i^b(w_1^b(\alpha), w_2^b(\alpha); \alpha)$ as $X_i^b(\alpha)$, it follows that

$$\begin{aligned} p_1 X_1^b(\alpha) + p_2 X_2^b(\alpha) &= D(\alpha) \left[\frac{1-\alpha}{w_1^b(\alpha)}\right]^{\frac{\eta}{\sigma(1-\eta)}} \left[w_1^b(\alpha) + w_2^b(\alpha)t^b(\alpha)\right]^{\frac{\eta-\sigma}{\sigma(1-\eta)}} \left[p_1 + p_2 t^b(\alpha)\right] \\ &= D(\alpha) \left[\frac{1-\alpha}{w_1^b(\alpha)}\right]^{\frac{\eta}{\sigma(1-\eta)}} \left[w_1^b(\alpha) + w_2^b(\alpha)t^b(\alpha)\right]^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}} r(\alpha), \end{aligned} \tag{A.15}$$

where $r(\alpha)$ is the ratio (29). Next, we compute the same quantities in the unbundling environment. The same computation yields

$$\begin{aligned} p_1 X_1^u(\alpha) + p_2 X_2^u(\alpha) &= D(\alpha) \left[\frac{1-\alpha}{p_1}\right]^{\frac{\eta}{\sigma(1-\eta)}} \left[p_1 + p_2 t^u(\alpha)\right]^{\frac{\eta-\sigma}{\sigma(1-\eta)}} \left[p_1 + p_2 t^u(\alpha)\right] \\ &= D(\alpha) \left[\frac{1-\alpha}{p_1}\right]^{\frac{\eta}{\sigma(1-\eta)}} \left[p_1 + p_2 t^u(\alpha)\right]^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}, \end{aligned} \tag{A.16}$$

where $X_i^u(\alpha) = X_i^u(p_1, p_2; \alpha)$ denotes the demand for skill i of firms with factor intensity α. Comparing the sorting conditions (16) under bundling and unbundling yields

$$t^u(\alpha) = t^b(\alpha) \left(\frac{p_1 w_2^b(\alpha)}{p_2 w_1^b(\alpha)}\right)^{\frac{1}{1-\sigma}}.$$

Replacing $t^u(\alpha)$ with the above value, we rewrite $p_1X_1^u(\alpha)+p_2X_2^u(\alpha)$ as

$$p_1X_1^u(\alpha)+p_2X_2^u(\alpha)=D(\alpha)\left[\frac{1-\alpha}{w_1^b(\alpha)}\right]^{\frac{\eta}{\sigma(1-\eta)}}\left[p_1\left(\frac{w_1^b(\alpha)}{p_1}\right)^{\frac{1}{1-\sigma}}+p_2t^b(\alpha)\left(\frac{w_2^b(\alpha)}{p_2}\right)^{\frac{1}{1-\sigma}}\right]^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}. \tag{A.17}$$

Let $K=p_1\bar{X}_1^w+p_2\bar{X}_2^w$. From (A.15), the market equilibrium for the two skills under bundling yields

$$K=\int a(\alpha)\,b(\alpha)^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}r(\alpha)\,\mathrm{d}\alpha, \tag{A.18}$$

where

$$a(\alpha)=D(\alpha)\left[\frac{1-\alpha}{w_1^b(\alpha)}\right]^{\frac{\eta}{\sigma(1-\eta)}}h^f(\alpha)\quad\text{and}\quad b(\alpha)=w_1^b(\alpha)+w_2^b(\alpha)t^b(\alpha).$$

From (A.17), the market equilibrium for the two skills under unbundling yields

$$K=\int a(\alpha)\,c(\alpha)^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}\,\mathrm{d}\alpha, \tag{A.19}$$

where

$$c(\alpha)=p_1\left(\frac{w_1^b(\alpha)}{p_1}\right)^{\frac{1}{1-\sigma}}+p_2t^b(\alpha)\left(\frac{w_2^b(\alpha)}{p_2}\right)^{\frac{1}{1-\sigma}}=w_1^b\left(\frac{p_1}{w_1^b(\alpha)}\right)^{\frac{\sigma}{\sigma-1}}+w_2^bt^b(\alpha)\left(\frac{p_2}{w_2^b(\alpha)}\right)^{\frac{\sigma}{\sigma-1}}.$$

Letting $d(\alpha)=b(\alpha)/c(\alpha)$, we now prove the inequality

$$d(\alpha)^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}\leq r(\alpha)^{\frac{\eta}{1-\eta}}. \tag{A.20}$$

Suppose first that $\sigma<0$. Because $x^{\sigma/(\sigma-1)}$ is concave in x, we have

$$\frac{c(\alpha)}{b(\alpha)}\leq\left[\frac{p_1+p_2t^b(\alpha)}{w_1^b(\alpha)+w_2^b(\alpha)t^b(\alpha)}\right]^{\frac{\sigma}{\sigma-1}}=r(\alpha)^{\frac{\sigma}{\sigma-1}}. \tag{A.21}$$

This yields

$$\frac{b(\alpha)}{c(\alpha)}\geq r(\alpha)^{\frac{\sigma}{1-\sigma}} \tag{A.22}$$

and hence (A.20). Next, suppose that $\sigma>0$. Then $x^{\sigma/(\sigma-1)}$ is convex in x, the inequalities (A.21) and (A.22) are reversed, and (A.20) remains true. Combining

(A.20) with (A.18), we get

$$K = \int a(\alpha)\, c(\alpha)^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}\, d(\alpha)^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}\, r(\alpha)\, \mathrm{d}\alpha \leq \int a(\alpha)\, c(\alpha)^{\frac{\eta(1-\sigma)}{\sigma(1-\eta)}}\, r(\alpha)^{\frac{1}{1-\eta}}\, \mathrm{d}\alpha,$$

which, combined with (A.19), shows that a weighted average of $r(\alpha)^{1/(1-\eta)}$ is larger than one, and hence, as explained above, that generalist workers with mix $\hat{\theta}$ benefit from skill unbundling.

A.9 Other examples of wage schedules

We assume that the supply of workers is uniform in the same polar coordinates. Now, we can introduce the above equation in equation (23) (the equality between supply of workers and labor demand from firms) to get:

$$\eta^{1/(1-\eta)} \int_z z^{1/(1-\eta)}\, \mathrm{d}H^f(z|\alpha)\, [(1-\alpha)(\cos\theta(\alpha))^\sigma + \alpha(\sin\theta(\alpha))^\sigma]^{1/(1-\eta)}\, h^f(\alpha) = \theta'(\alpha). \tag{A.23}$$

Noticing that

$$\cos\theta(\alpha) = \frac{1}{\sqrt{1+(\frac{1-\alpha}{\alpha})^{\frac{2}{\sigma-2}}}},$$

$$\sin\theta(\alpha) = \frac{(\frac{1-\alpha}{\alpha})^{\frac{1}{\sigma-2}}}{\sqrt{1+(\frac{1-\alpha}{\alpha})^{\frac{2}{\sigma-2}}}},$$

and

$$\theta'(\alpha) = -\frac{1}{\alpha^2}[\frac{(\frac{1-\alpha}{\alpha})^{\frac{1}{\sigma-2}-1}}{1+(\frac{1-\alpha}{\alpha})^{\frac{2}{\sigma-2}}}],$$

we can rewrite the equilibrium equation as:

$$\eta^{1/(1-\eta)} \int_z z^{1/(1-\eta)}\, \mathrm{d}H^f(z|\alpha) h^f(\alpha) = \frac{\theta'(\alpha)\int_\lambda \lambda\, \mathrm{d}H^w(\lambda|\theta(\alpha)) h^w(\theta(\alpha))}{[(1-\alpha)(\cos\theta(\alpha))^\sigma + \alpha(\sin\theta(\alpha))^\sigma]^{1/(1-\eta)}}.$$

The sorting condition for the ellipse is given now.

$$\theta(\alpha) = \arctan[(\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{1}{\sigma-2}}] \tag{A.24}$$

becomes our new sorting equation. In contrast to the case of the circle, $w(\theta)$ is not constant any more. The log-wage decomposition now includes a firm-specific effect that captures the firm-specific technology at equilibrium. Indeed,

$$\cos^2\theta(\alpha) = \frac{1}{1 + (\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{2}{\sigma-2}}},$$

$$\sin^2\theta(\alpha) = \frac{(\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{2}{\sigma-2}}}{1 + (\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{2}{\sigma-2}}},$$

and

$$\theta'(\alpha) = -\frac{a^2}{b^2}\frac{1}{\alpha^2}[\frac{(\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{1}{\sigma-2}-1}}{1 + (\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{2}{\sigma-2}}}],$$

All these elements yield in particular the sorting equilibrium and the wage using the above expression for $w(\theta)$:

$$w(\theta(\alpha)) = \frac{\frac{1}{a^2} + \frac{1}{b^2}(\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{2}{\sigma-2}}}{1 + (\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{2}{\sigma-2}}}$$

Therefore,

$$w'(\alpha) = \frac{2(b^2-a^2)}{(\sigma-2)}\frac{a^2}{b^2\alpha^2}\frac{(\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{4-\sigma}{\sigma-2}}}{[1 + (\frac{1-\alpha}{\alpha}\frac{a^2}{b^2})^{\frac{2}{\sigma-2}}]^2}$$

and because $\sigma < 1$, the firm-specific component of the wage is increasing (resp. decreasing) with α when $a > b$, a flat ellipse, (resp. $a < b$).

We can contrast the above case with the ellipse turned counter-clockwise by $\frac{\pi}{4}$. In this case,

$$w(\theta) = \sqrt{\frac{cos^2(\theta - \frac{\pi}{4})}{a^2} + \frac{sin^2(\theta - \frac{\pi}{4})}{b^2}}$$

which can be rewritten as

$$w(\theta) = \frac{\sqrt{2}}{2}\sqrt{\frac{1}{a^2} + \frac{1}{b^2} + 2\cos\theta\sin\theta(\frac{1}{a^2} - \frac{1}{b^2})}$$

The sorting equation becomes

$$(\frac{1-\alpha}{\alpha})(\tan\theta)^{1-\sigma} = \frac{\Delta^+ + \Delta^-\tan\theta}{\Delta^- + \Delta^+\tan\theta}.$$

with $\Delta^+ = \frac{1}{a^2} + \frac{1}{b^2}$ and $\Delta^- = \frac{1}{a^2} - \frac{1}{b^2}$ Unfortunately, the sorting equation is less straightforward. However, when $\sigma = 0$ hence assuming a Cobb-Douglas production function, it is possible to go a step further:

$$(1-\alpha)(\Delta^-\tan\theta + \Delta^+\tan^2\theta) = \alpha(\Delta^+ + \Delta^-\tan\theta)$$

or

$$(1-\alpha)\Delta^+ \tan^2\theta + (1-2\alpha)\Delta^- \tan\theta - \alpha\Delta^+ = 0$$

Hence,

$$\theta(\alpha) = \arctan[(\frac{1-2\alpha}{2(1-\alpha)}\frac{\Delta^-}{\Delta^+} \pm \sqrt{\frac{1-\alpha}{\alpha} + (\frac{1-2\alpha}{2(1-\alpha)}\frac{\Delta^-}{\Delta^+})^2}] \tag{A.25}$$

Finally, note that $w(0) = w(\pi/2) = \frac{\sqrt{2}}{2}\sqrt{\Delta^+}$ and $w(\pi/4) = \frac{\sqrt{2}}{2}\sqrt{\Delta^+ + \Delta^-}$. Therefore, in the case of $a > b$ considered here, $\Delta^- < 0$ and the price of generalists ($\theta = \pi/4$) is lower than the price of specialists ($\theta = 0$ and $\theta = \pi/2$). Demand for the former is larger than demand for the latter. Hence sorting is first decreasing and then increasing.

CRESTER FOR RESEARCH IN ECONOMICS AND STATISTICS

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

The Center for Research in Economics and Statistics (CREST) is a leading French scientific institution for advanced research on quantitative methods applied to the social sciences.

CREST is a joint interdisciplinary unit of research and faculty members of CNRS, ENSAE Paris, ENSAI and the Economics Department of Ecole Polytechnique. Its activities are located physically in the ENSAE Paris building on the Palaiseau campus of Institut Polytechnique de Paris and secondarily on the Ker-Lann campus of ENSAI Rennes.