DISCUSSION PAPER SERIES

DP17645

MATCHING WORKERS' SKILLS AND FIRMS' TECHNOLOGIES: FROM BUNDLING TO UNBUNDLING

Philippe Chone and Francis Kramarz

INDUSTRIAL ORGANIZATION AND LABOUR ECONOMICS

ISSN 0265-8003

MATCHING WORKERS' SKILLS AND FIRMS' TECHNOLOGIES: FROM BUNDLING TO UNBUNDLING

Philippe Chone and Francis Kramarz

Discussion Paper DP17645
Published 04 November 2022
Submitted 25 October 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

MATCHING WORKERS' SKILLS AND FIRMS' TECHNOLOGIES: FROM BUNDLING TO UNBUNDLING

Abstract

How are workers matched to firms when their skills are multidimensional and firms differ in how they value each skill dimension for their production? When workers' skills cannot be unpacked and sold separately on skill-specific markets, the implicit price of each skill can vary across firms. The wage function is shown to be log-additive in worker quality and a firm-specific effect that reflects the firm's aggregate skill-mix and equilibrium matching. When individual skills can be unpacked or purchased on markets thanks to new technologies and increasing access to outsourcing, 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. We discuss the empirical content of our theory and supportive evidence which uses Swedish data on workers' skills and their employing firms.

JEL Classification: D20, D40, D51, J20, J24, J30

Keywords: Bundling, Multidimensional skills, Matching, Sorting, Heterogenous firms, Polarization

Philippe Chone - Philippe.Chone@ensae.fr
CREST-ENSAE-Ecole Polytechnique and CEPR

Francis Kramarz - kramarz@ensae.fr
CREST-ENSAE-Ecole Polytechnique and CEPR

Acknowledgements
We are grateful to Marco Cuturi, Xavier d'Haultfoeuille, Alfred Galichon, Nathael Gozlan, Philipp Kircher, Samuel Kortum, Clémence Lenoir, Ilse Lindenlaub, Simon Mongey, Giuseppe Moscarini, Bérengère Patault, François-Pierre Paty, Torsten Persson, Roland Rathelot, Jean-Marc Robin, Bernard Salanié, David Schönholzer, Oskar Nordström Skans, and Bob Topel for extremely helpful conversations, remarks, and suggestions. We also thank participants at multiple seminars and conferences. We thank Gaspard Choné-Ducasse and Pierre Krakovitch for excellent research assistance. 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).

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*

November 3, 2022

Abstract

How are workers matched to firms when their skills are multidimensional and firms differ in how they value each skill dimension for their production? When workers' skills cannot be unpacked and sold separately on skill-specific markets, the implicit price of each skill can vary across firms. The wage function is shown to be log-additive in worker quality and a firm-specific effect that reflects the firm's aggregate skill-mix and equilibrium matching. When individual skills can be unpacked or purchased on markets thanks to new technologies and increasing access to outsourcing, 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. We discuss the empirical content of our theory and supportive evidence which uses Swedish data on workers' skills and their employing firms.[1]

JEL Codes: D20, D40, D51, J20, J24, J30

Keywords: bundling; multidimensional skills; matching; sorting; heterogeneous firms; polarization

---

*We are grateful to Marco Cuturi, Xavier d'Haultfoeuille, Alfred Galichon, Nathael Gozlan, Philipp Kircher, Samuel Kortum, Clémence Lenoir, Ilse Lindenlaub, Simon Mongey, Giuseppe Moscarini, Bérengère Patault, François-Pierre Paty, Torsten Persson, Roland Rathelot, Jean-Marc Robin, Bernard Salanié, David Schönholzer, Oskar Nordström Skans, and Bob Topel for extremely helpful conversations, remarks, and suggestions. We also thank participants at multiple seminars and conferences. We thank Gaspard Choné-Ducasse and Pierre Krakovitch for excellent research assistance. 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] This empirical analysis is fully based on Skans, Choné, and Kramarz (2022), jointly written with Oskar Nordström Skans.

1 Introduction

Uberization, the Gig Economy ... Words often used in the press to identify the new forms of labor. 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 these new jobs appear to be missing.

To understand how labor markets operate now, we start by modelling older forms of labor. We characterize such forms by building on Mandelbrot (1962), the first to note "the impossibility of renting the different factors to the different employers", as cited in Heckman and Scheinkman (1987). Hence, firms are forced to hire workers endowed with their entire skill-set. Heckman and Scheinkman (1987) (HS, hereafter) use the word *Bundling* to name this constraint: the impossibility to unpack a worker's package of skills (hence, the impossibility for workers to sell each skill separately on a market).

By contrast, to characterize the new forms of labor, we examine how the labor markets are transformed when markets for individual skills open, potentially at a cost; a process we call *Unbundling*.

More precisely, we first study how workers are matched to firms in a bundled world and the resulting wage structure. Then, we look at how labor markets change in an unbundled world. In doing so, we try to capture the role of new technologies, increasing access to outsourcing, to temp agencies, or platforms in shaping the allocation of workers to firms, as well as the ensuing wage structure. This contrast between the old world and the new, based on theoretical modelling inspired from (what we hope to be) a deep knowledge of labor markets and their economic environment, will be shown to be informative about this very human and therefore very different from other markets (product markets, in particular).

In this article, we build on Heckman and Scheinkman's theoretical insight.[2] 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 or purchased separately, an employing firm has access to all skill components a person is endowed with. We follow Heckman and Scheinkman in assuming that *each firm's production function depends on its workers' (bundled) skills aggregated by skill-types*, $X = (X_1, \ldots, X_j, \ldots, X_k)$ with $X_j = \int x_j$ (the integral being taken over the measure of

[2] We discuss how our approach is connected to those found in the literature, later in this Introduction.

workers employed in the firm), to produce a bundle of k tasks *rather than each worker's (job) production aggregated over workers (jobs) employed at the firm.*[3]

Importantly, both firms and workers display rich multidimensional heterogeneity, allowing us to examine the *matching* of workers to firms and the induced *sorting*. More precisely, we study how a continuum of workers, endowed with multidimensional (exogenously given) skills, match with a continuum of firms, also endowed with multidimensional (also exogenously given) heterogeneity (rather than a 2-sector setup with a continuum of identical firms within each sector, as in HS). Firms are allowed to choose their size, a well-defined concept in our approach. We derive the wage schedule that prevails at the general competitive equilibrium of this economy and show that it is a) a homogenous function of degree one in the "quality" of the worker; b) a non-linear convex function in the bundle.[4] Hence, in equilibrium, the implicit price of each skill-type varies across firms and the law of one price does not apply: *there is more than one price per type of skill*, potentially an infinite number of such prices.[5] This result is a direct consequence of the inefficiency – *constrained efficiency* – induced by bundling: the impossibility of unpacking a worker's multidimensional skills.

Crucially, we exhibit the allocation of workers to firms and the *sorting* patterns displayed at this equilibrium. Under usual single-crossing conditions of the firm's technology, aggregate sorting obtains and firms hire their unique preferred mix of skill-types, say the ratio X_2/X_1 in a two-skills world, a phenomenon that we label "sorting in the horizontal dimension".[6] Depending on the skills supply prevailing in the economy, this preferred mix is obtained by hiring either workers with exactly that preferred mix or a combination of workers delivering the same exact preferred mix (a pattern we call "Bunching" in this paper). To give an intuition of this last result, consider a world with two skills, 1 and 2.[7] In this world, let us assume 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 will hire a mixture of workers of type 1 and type 2 so as to obtain its optimal mix X_2/X_1. In this example, no worker in the firm is endowed with the optimal mix, and the wage is linear in the two skills, with one unique price per skill. By contrast, 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 are endowed with their employing firm's optimal mix. In this case, the wage is nonlinear, with the implicit price of each skill depending on the worker's employing firm.

[3]In most of our analysis, we equate skills and tasks. We discuss this assumption in Section 5.

[4]Rather than a linear function in skills with returns allowed to differ in each sector, again as in HS.

[5]Even though there is a unique price per bundle.

[6]This two-skills world seems to resemble the so-called "Roy model" but we discuss below why our model vastly differs from it.

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

The model also delivers predictions about sorting patterns in the vertical dimension. First, a given firm does not necessarily employ workers of the same quality. For instance, in the absence of bunching, when supply is located away from the axes, the employees of a given firm have skill sets of the form $x = (\lambda X_1, \lambda X_2)$: while they are all endowed with the firm's optimal mix X_2/X_1, they may be heterogeneous in their quality, i.e., λ may vary within a subset of $\mathbb{R}_+$. Yet we demonstrate the *uniqueness* of the firm-aggregated vector of skills at any competitive equilibrium and show that high-productivity firms will employ a high-quality labor force (endowed with a high total amount of the different skills). Hence, a high-quality labor force, a well-defined firm-level concept, may stem from hiring many mediocre workers, i.e., by increasing the size of the firm, or from hiring a smaller number of excellent workers. It follows that *conditional on employment* high-productivity firms employ high-quality individual workers.

Another consequence of our results in this bundled world is the log-additivity of the wage function in worker's quality and in a firm-specific effect. The latter 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 holds exactly *in the convex portions of the wage schedule.* As mentioned above, however, supply together with demand conditions may yield an equilibrium in which firms must mix workers with skills that differ from the optimal mix. Bunching is shown to prevail in regions where the wage schedule in skills is linear. When those regions are "small" enough, the wage function is close to such log-additivity. Hence, in our bundled world – with multidimensional skills and firms with heterogeneous production functions – 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 (coming from technology rather than profit-sharing or monopsony) is *pervasive.* Because high-productivity firms also employ a high-quality (total) labor force, these two effects may well be positively correlated. 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 routinely does.

In a world of opening markets, through better technology, globalization, temp agencies, or, more recently, platforms, the unbundling of skills is facilitated, potentially at a cost.[8] To analyze the effect of increased market access, we examine how the matching of workers to firms is altered when opening all markets for skill-types simultaneously. *Full unbundling* (i.e. with no unbundling cost for workers or firms) restores unconstrained efficiency. In a bundled world, workers must supply all their labor to their employing

[8]Before analyzing full opening, we look at an intermediate setting in which workers – with their skills still bundled – are allowed to alter their skills supply. This setting offers an important contrast with the bundling environment but interesting similarities with unbundling which we discuss now.

firm. In the unbundled world, the one studied by most of the previous literature, a market exists for each skill. There, 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.

The first consequence of the existence of such markets is that wages become linear combinations of workers' skills endowments. Hence, going from a bundled to an unbundled world, the "flattening" of the wage schedule implies an increase in within-firm workers' skills heterogeneity and a progressive elimination of firm effects.[9]

A second characterization of these changes (going from a world with bundled skills to one where they are unbundled) is obtained by identifying those workers benefiting 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. This "markdown" affecting generalists in a bundled world is reminiscent of monopsonistic models of the labor market. However, in our bundling framework, there is no labor supply per se; all the effects come from firms' labor demand. Endogenous labor supply only kicks in when markets for skills open. And, as stated just above, generalists benefit from this opening. The contrast with monopsony becomes even more interesting: markets opening in a model of bundling, potentially resulting from public policies (as in the Hartz laws) eliminates the wage "markdown" when introduction of a minimum wage, another public policy, has a similar effect in monopsony models of the labor market. This parallel holds despite the opposite origin of such markdowns, coming from the demand side in one model and from an upward-sloping labor supply in the other.

Third, again after unbundling, comparative advantage in sorting continues to hold, 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

[9] See, among many others, Goldschmidt and Schmieder (2017) who study the impact of the Hartz reforms – through increased access to skills by using temp agencies in particular – on wages in the aftermath of domestic outsourcing.

widening of polarization and the flattening of the equilibrium wage schedule – and the associated elimination of firms effects – in detail.

In a companion paper, co-written with Oskar Nordström Skans, Skans, Choné, and Kramarz (2022), summarized in this article just before our conclusion, we use Swedish data on workers' skills, employers, and occupations, and provide descriptive evidence of some of the consequences of our model. Indeed, our empirical results appear to confirm the role of comparative advantage in sorting (on top of absolute advantage). They also demonstrate that generalists have seen their position getting better over time wrt generalists, in accordance with our model's predictions.

Connecting Literatures We believe that our theoretical contribution incorporates four ingredients – 1) a continuum of heterogeneous workers with multidimensional skill-types; these skills being either bundled or unbundled; 2) a continuum of firms with heterogeneous and multidimensional 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*; 4) an endogenous firm size. A (potentially) non-linear wage schedule will allow the matching (sorting) of these multidimensional workers to their multidimensional 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 ingredients.

Bundling Multidimensional Skills: HS is the first paper, which we are aware of, examining the consequences of bundling of skills. These authors were trying to understand whether bundling of skills (first ingredient above) together with production obtained from an aggregation of workers' skills (third ingredient) could generate different returns to each skill in two different sectors, in an economy with n sectors (and identical firms within each sector, the firms playing essentially no role). Their answer was positive: returns for skills could differ across sectors, in this Roy-style model. Unfortunately, they did not provide general conditions for their result. Nor did they examine the structure of the matching between workers and firms (sectors). By contrast, Lindenlaub (2017) focuses on sorting and provides a full characterization of positive assortative matching, PAM, or its negative counterpart, NAM, in a multidimensional framework with jobs but no aggregation of skills used in a firm-level production function. 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. [10] Because Lindenlaub (2017) is an important step in the study of the matching of workers to jobs in this multi-dimensional (with bundling) context, we will relate her results to ours directly within the body of our theory Sections.

Edmond and Mongey (2020) also examine bundling using a model with two tasks and two skills, with bundling or after an unbundling of skills (using this word as we do), adopting a purely macroeconomic perspective. As in our approach, their workers are heterogeneous in their skill endowments. As in Murphy (1986) and HS, 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 do, 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. Importantly, and very much as we will do here, they examine how unbundling operates, something that none of the previous papers had looked at. In a recent contribution, 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.

Consequences of skills-bundling were also studied in International Trade (Ohnsorge and Trefler (2007)). There, workers have bundled skills and the production side of the economy is much simpler, with jobs rather than firms. These authors' interests lie in sources of comparative advantage generated by such bundling constraints in a country. We come back to this point just below.

Comparative Advantage in the Vertical Dimension: A more macroeconomic literature studying trade, comparative advantage, and technical change has also connections with our approach. In Costinot and Vogel (2010), and as we do here, firms use workers to produce intermediate goods ("tasks" or "sectors" for them, firm-aggregated skills for us). The tasks are then combined into a final product. In contrast with our assumptions however, firms that produce the final good use no labor and purchase their inputs on upstream markets. Furthermore, the upstream firms operate under constant returns to scale and hence make zero profit (in contrast to the final

[10] Lise and Postel-Vinay (2020) also have a search-theoretic component and allow for multidimensional skills, without bundling though, and on-the-job learning.

good producers). There is no heterogeneity across firms within sectors: all firms that produce a given (intermediate or final) good share the same technology. Workers are heterogeneous in a single dimension, hence there is no bundling (and because there is a market for each task, full unbundling of tasks/skills prevails). This allows Costinot and Vogel (2010) to study a Roy-like assignment model where high-skill workers have a comparative advantage in tasks with high-skill intensity, what we call the vertical dimension. In equilibrium, this results in sorting between skills and tasks, in which each worker performs a single task. Indeed, in our approach, we show that individuals with a comparative advantage in one skill will work in firms that value this exact skill more. In Appendix B, we study a Dixit-Stiglitz variant of our model.[11] It allows to clearly see how versatile our bundling model can be and also how different environments (pure competition versus monopolistic competition, in this case) deliver similar effects based on different formulas. Both approaches, Costinot and Vogel (2010)'s and ours, deliver a role for sorting of workers to firms through a comparative advantage mechanism, one-dimensional in the vertical dimension for the former, multidimensional in the horizontal dimension for us.

Ohnsorge and Trefler (2007) also have multidimensional skills (but no firms) and show that international differences in the distribution of workers' skill bundles, such as Japan's abundance of workers with a modest mix of both quantitative and teamwork skills, have important implications for international trade, industrial structure, and domestic income distribution.

Connected to this trade literature, with a clear focus on labor markets, two contributions must be mentioned. First, Teulings (2005) (cited in Costinot and Vogel (2010)) presents a theory of factor substitutability in a model with a continuum of worker and job (both uni-dimensional) types, with highly skilled workers having a comparative advantage in complex jobs. This model allows to generate patterns of substitutability between types that decline with their skill distance. Second, in a recent and very interesting article, Haanwinckel (2020) contributes to this labor literature. His task-based production function requires combining tasks of different complexity levels, with task requirements depending on the good the firm decides to sell. As in Teulings (2005), the comparative advantage structure is uni-dimensional, corresponding to what we label the vertical dimension of skills. Interestingly, the firm assigns (optimally) each worker to tasks, resulting in within-firm heterogeneity in workers' types. Labor market imperfections (e.g. a minimum wage or monopsony) are then added to the model. Some of its predictions are also examined empirically using Brazilian data.

[11] We thank Sam Kortum for this suggestion.

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, Eeckhout and Kircher (2018) separate workers' quality from workers' quantity and assume constant returns to scale in those quantity variables. In addition, management 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. As a result of the constant returns assumptions in particular, at the equilibrium, a firm of quality y hires only one quality of worker x, with the mapping between x and y being one-to-one, hence the model generates no within-firm worker's heterogeneity. Unfortunately, very few contributions address this firm's substance challenge. We mentioned above Haanwinckel (2020). A recent and interesting contribution is Boerma, Tsyvinski, and Zimin (2021) with firms of exogenous size (equal to two). Their model includes a team production function with bundling and heterogeneous firms (in productivity only, though). Their interest lies in the matching between such firms and workers. We briefly mention some of the mathematical techniques they use in the paragraphs just below.

Firms also play a role in recent GE models of monopsonistic labor markets, such as Berger, Herkenhoff, and Mongey (2022) (see also references, therein). A finite number of firms in a market, each firm having an upward sloping labor supply curve, face workers endowed with different tastes for firms. The resulting equilibrium yields a markdown of wages. Workers have an active supply behavior when, in our bundled world, workers make essentially no choice and just respond to firms' labor demand. And, as mentioned earlier, generalists – most constrained by bundling – face a "markdown". When markets open, with the associated unbundling of skills, generalists are better off and the bundling markdown vanishes.

Connecting Optimal Transport and Matching Problems: Our analysis contributes to the vast literature that studies many-to-one matching with transferable utility. A fraction of this literature has examined the problem in its discrete (gametheoretic) version[12] whereas we work with a continuum of workers and a continuum of firms.

A growing strand of the literature leverages the insights of optimal transport theory to study the matching of agents in competitive markets.[13] Important papers in this strand actually consider one-to-one matching, e.g. in the labor market (Lindenlaub (2017)) or in the market marriage market, Galichon and Salanié (forthcoming)). Boerma, Tsyvinski, and Zimin (2021), briefly presented just above, use the multi-marginal version of optimal transport, with the marginal distributions of the transport plan being prescribed on three sets that represent firms', workers', and co-workers' types.

Using the optimal transport perspective, hedonic models share many features with matching problems (see Chiappori, McCann, and Nesheim (2010)). In hedonic models, the focus is on matching firms and products on the one hand, and consumers and products on the other, with the equilibrium imposing equality of the (products') marginals between the two transport plans. In matching on the labor market such as here, the focus is or *should be* on matching workers' skills and tasks on the one hand, and tasks and firms on the other. Importantly though, *tasks are not observed* by the researcher in this type of problem.[14] Hence, hedonic models share multiple – but not all – features of what we are studying here. In particular, consumers in hedonic models correspond to firms for us when goods and products in hedonic models correspond to workers and their skills in our approach.

Whereas we insisted above on similarities between hedonic pricing and our matching problem, there is at least one important difference: the firm's ability to aggregate workers' skills for production. Most of the literature initiated by Rosen (1974) 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).[15]

To deal with this aggregation of skills within firm, we use new methods and results from OT theory, namely the so-called weak optimal transport (WOT) introduced by Gozlan, Roberto, Samson, and Tetali (2017). To allow for endogenous firm sizes, we rely on the extension of WOT introduced by Choné, Gozlan, and Kramarz (2022).

[12] 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.

[13] 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.

[14] Indeed, we are not aware of any data source that would offer a comprehensive picture: workers' exact skills, the exact tasks each worker performs, together with the worker's employing firm. Often occupations are used as a proxy even though the tasks performed by the worker in her employing firm are never measured.

[15] Two cars with 50 horsepower each are not equivalent to one with 100 horsepower is an obvious example. See also Lancaster (1966).

In Appendix A.11, we describe the latter set-up. Paty, Choné, and Kramarz (2022) develop efficient algorithms to numerically approximate the equilibrium solutions.

Bunching and bundling Using the literature on multidimensional optimal transport, Chiappori, McCann, and Pass (2016) derive conditions under which stable matches are unique and pure. They connect their work to the multidimensional screening literature and argue that the bunching phenomena, observed by Rochet and Choné (1998) in the monopoly context, do not occur in the competitive context. In the present paper, we find something akin to bunching in a competitive environment with multidimensional 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 **only** be achieved by combining workers endowed with different skill-types. In this precise situation, firms of different types optimally hire workers endowed with the exact same skill-type to achieve their (different) optimal mix; a phenomenon we call "bunching".

In the next Section, we present our model setup, when bundling prevails. Then, Section 3 examines how firms and workers are matched, again under bundling. In Section 4, we first look at what happens when skills are bundled but workers are allowed to choose their skills supply, and then look at consequences of skills unbundling. Next, we discuss the empirical consequences of our model (Section 5) before confronting these consequences with data evidence (Section 6).[16] Section 7 briefly concludes. All proofs are relegated to the Appendix.

2 Model Setup Under Bundling

The production process involves k intermediary inputs produced by workers, which we call tasks. Firms aggregate the tasks performed by their employees and transform them into final output. They are heterogeneous in their production technologies. Denoting by $T = (T_1, \dots, T_k)$ the aggregate vector of tasks produced by its employees, a firm of type ϕ produces final output $F(T; \phi)$, with F being concave in T. Firms' types are distributed according to a probability measure $H^f(\mathrm{d}\phi)$ on $\Phi \subset \mathbb{R}^k_+$.

[16] Section 6 contains chosen elements of a paper, co-written with Oskar Nordström Skans, Skans, Choné, and Kramarz (2022), in which we study some aspects of the empirics of bundling and unbundling using Swedish data.

Performing tasks requires skills. Workers are heterogeneous in their skill endowments, our primitive on the supply side. Each worker's endowment is given by a skill vector $x = (x_1, \dots, x_k)$, $k \geq 2$. Skills are distributed according to a probability measure $H^w(dx)$ on $\mathcal{X} \subset \mathbb{R}^k_+$. We define the overall quality of a worker as the Euclidian norm $|x|$ of her skill vector $\tilde{x} = x \in \mathbb{R}^k_+$ and her skill profile as $x/|x|$. We refer to the former and latter respectively as to the vertical and horizontal dimensions of workers' heterogeneity.

In this Section, as well as in the next two sections, *we simply equate skills with tasks*. We briefly discuss the relationship between skills and tasks when analyzing the empirical content of our model in (Section 5).

As in Acemoglu and Autor (2011), the total amount of task j in a firm is obtained by linear aggregation:

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

where $N^d(\mathrm{d}x; \phi)$ is a positive measure on $\mathcal{X}$ that represents the number of workers of each type x hired by a firm of type ϕ.

An assignment of workers to firms is a family of a positive measures $N^d(\mathrm{d}x; \phi)$ on $\mathcal{X}$. Important to stress that we use "unnormalized" positive measures. Hence, the size of firms, which we denote by $N(\phi) = N^d(\mathcal{X}; \phi)$, need not be one and $N^d(\mathrm{d}x; \phi)$ need not be a probability measure. In fact, the firms' sizes are endogenously determined in equilibrium.

An assignment N^d "clears" the labor market if

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

for H^w-almost all worker types $x \in \mathcal{X}$. In other words, market clearing assignments "disintegrate" the skill distribution $H^w(\mathrm{d}x)$ and quantify the number of workers of any type x hired by firms of any type ϕ. Below, we often write the market clearing equation (2) in the shorter form $N^d H^f = H^w$. Integrating this equation with respect to x shows that, for any market clearing assignment N^d, the expected firm size is one:

$$\int N(\phi) H^f(\mathrm{d}\phi) = 1. \tag{3}$$

In other words, the distribution of firms type $\tilde{H}^f(\mathrm{d}\phi) = N(\phi)H^f(\mathrm{d}\phi)$ is a probability measure. Introducing $q(\mathrm{d}x; \phi) = N^d(\mathrm{d}x; \phi)/N(\phi)$, a probability measure for any ϕ, shows that the matching between workers' and firms's types

$$\pi(\mathrm{d}x, \mathrm{d}\phi) = N^d(\mathrm{d}x; \phi)H^f(\mathrm{d}\phi) = q(\mathrm{d}x; \phi)\tilde{H}^f(\mathrm{d}\phi) \tag{4}$$

is a transport plan between the original skill distribution $H^w(x)$ and the modified firm distribution $\tilde{H}^f(\mathrm{d}\phi)$.[17]

We say that *a market clearing assignment N^d is optimal* if it maximizes total output in the economy, i.e., if it solves

$$Y^* \overset{\mathrm{d}}{\equiv} \sup_{N^d \,|\, N^d H^f = H^w} \int F\left(\int x N^d(\mathrm{d}x;\phi);\phi\right) H^f(\mathrm{d}\phi). \tag{5}$$

Whenever the production function F is nonlinear in the firm-aggregate vectors of tasks T, the total output in the economy is a nonlinear function of the assignment N^d. By contrast, if firms' production were just the sum of each of their employees' production – which is not what we do here –, total output $\iint F(x;\phi)N^d(\mathrm{d}x;\phi)H^f(\mathrm{d}\phi)$ would be linear in N^d.

Finally, we introduce the notion of competitive equilibrium. Under bundling, a worker's set of skills cannot be untied, hence firms must purchase her entire skill package $x = (x_1, \ldots, x_k)$. The workers' skills are observed by firms and are contractible. The wage of a worker of type x is denoted by $w(x)$. The wage schedule $w(.)$ is therefore a map: $\mathcal{X} \to \mathbb{R}_+$. We rule out agency problems: a firm that hires a worker of type x pays $w(x)$ and obtains the vector of intermediary inputs x. Given a wage schedule $w(.)$, the demand for skill is the assignment $N^d(\mathrm{d}x;\phi)$ on $\mathcal{X}$ that maximizes the firms' profit:

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

A **competitive equilibrium** is a pair (w, N^d) composed of a wage schedule and a market-clearing assignment of workers to firms such that the assignment N^d reflects the demand for skills under the wage w, i.e., N^d solves the firms' problem (6).

Of particular interest to us are the production functions of the form $F(T;\phi) = zF(T;\alpha)$ with the firms' types $\phi = (\alpha, z)$ having two components: z reflects total factor productivity and α reflects the relative importance of each task in the production process. We assume that the worker and firm heterogeneities have the same dimension, hence α lies in a space of dimension $k-1$. Our leading example exhibits constant elasticity of substitution and decreasing returns to scale:

$$zF(T;\alpha) = (z/\eta)\;\left[\sum_{j=1}^{k} \alpha_j T_j^{\rho}\right]^{\eta/\rho}, \tag{7}$$

with $\sum_{j=1}^{k} \alpha_j = 1$, $\eta < 1$, and $\rho < 1$.

[17] See Section 1 for more details about the connection of our framework to optimal transport theory.

When $\rho < \eta$, the function displays increasing marginal productivities of aggregate skill types, $\partial^2 F/\partial T_j \partial T_k > 0$ for all $j \neq k$.[18] In other words, the marginal productivity of a worker in one skill increases with her co-workers' other skills. Under this specification, complementarities across workers result from complementarities across skill types.

Important, the bundling constraint is a) versatile enough to be embedded in a perfectly competitive world (but for bundling) as we study below but also in a Dixit-Stiglitz framework (see Appendix B) where firms operate under constant returns to scale and quantities are set by monopolistic competition; b) simple enough to deliver explicit results with testable consequences in these different environments.

3 Matching Workers and Firms Under Bundling

We continue to assume that there are no markets for individual skills/tasks. Firms can acquire intermediary inputs only from their employees. Once hired, a firm can use the entirety of a worker's skills. In addition, we assume that a worker cannot be employed by more than one firm.

In Subsection 3.1, we prove the existence of competitive equilibria using new insights from optimal transport theory. In Subsection 3.2, we examine how the firm-*aggregated* vectors of tasks depend on the firms' technologies. Then, assuming homothetic production functions, we study the sorting of *individual* workers into firms. In Subsection 3.3, we focus on cases where pure sorting in the horizontal dimension obtains. In Subsection 3.4, we describe situations where, by contrast, skill profiles are heterogeneous within firms.

3.1 Competitive Equilibria and the Structure of Wages

Competitive equilibria will be shown to exist under the following assumptions:

Assumption 1. *(i) For all $x \in \mathcal{X}$ and $\phi \in \Phi$, $F(\lambda x; \phi)/\lambda$ tends to 0 as $\lambda \to +\infty$; (ii) $\inf_{x \in \mathcal{X}, \phi \in \Phi} F(\lambda x; \phi)$ tends to $+\infty$ as $\lambda \to +\infty$; (iii) The convex hull of $\mathcal{X}$ does not contain 0.*

Part (i) is true in particular for homogenous production functions with diminishing returns to scale, as is the case in our leading example (7). Part (ii) implies that increasing the number of workers even for the poorest match between the workers' and firms' types allows to produce an arbitrary large quantity of final output. Finally, part (iii) captures the impossibility to find convex combinations of workers' skills in $\mathcal{X}$ that

[18] In the case of two skills, the condition $\partial^2 F/\partial T_1 \partial T_2 > 0$ ensures that the aggregate skills are complements, i.e., that the demand for one skill decreases with the price of the other skill.

are arbitrarily close to 0 in $\mathbb{R}^k_+$. Hence, all workers have a positive amount of skills in at least one skill dimension $j = 1, \ldots, k$.

As already explained, the bundling environment is characterized by missing markets. Firms cannot purchase some amount of skills, separately for each skill type $j = 1, \ldots, k$. Proposition 1 below states versions of the two fundamental theorems of welfare economics that are adapted to this constrained environment. In particular the notion of optimality refers to the "primal" Problem (5), which includes the constraints that only workers can be hired and that only skill-*vectors* can be traded.

Proposition 1 (The Fundamental Theorems Under Bundling)**.** *Suppose Assumption 1 holds. Then there exist optimal market clearing assignments of workers to firms. Any such assignment can be decentralized by a wage schedule w. Conversely, any equilibrium assignment is optimal.*

The next proposition, which describes in more detail the structure of wages, continues to assume the identity between skills and tasks. We briefly discuss in Section 5 how wages are affected when skills and tasks are allowed to differ.

Proposition 2 (Structure of wages)**.** *Suppose Assumption 1 and equation* (1) *both hold. Then any optimal market clearing assignment can be decentralized by a wage schedule w that is convex and homogenous of degree one.*

The convexity and homogeneity of the wage schedule 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 skills.

Suppose for instance that there exist worker types x, x', and x'' such that $x'' = \nu x + (1-\nu)x'$ with $0 < \nu < 1$, $w(x) = w(x') = 1$, and $w(x'') > 1$. Then, no firm would want to hire type-x'' workers because a combination of type-x and type-x' workers would deliver the same amount of intermediary inputs in return for a lower wage bill. Specifically, diminishing demand $N^d(x'', \phi)$ by ε and increasing $N^d(x, \phi)$ by $\nu\varepsilon$ and $N^d(x', \phi)$ by $(1-\nu)\varepsilon$ leaves the firm-aggregated vector of tasks unchanged and reduces the wage bill.

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 $w(\lambda x) < \lambda w(x)$. 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, $w(\lambda x) > \lambda w(x)$, is ruled out by the same argument.

Proposition 2 implies that the wage is sub-additive, which has important economic implications. Let (e_i) be the canonical basis of $\mathbb{R}^k$, i.e., $e_i = (0, \ldots, 1, \ldots, 0)$, with 1 in the ith coordinate. Because w is convex and homogenous of degree one, it is sub-additive, hence

$$w(x) = w\left(\sum_{i=1}^{k} x_i e_i\right) \leq \sum_{i=1}^{k} w(e_i x_i) = \sum_{i=1}^{k} w(e_i) x_i. \tag{8}$$

Hereafter, we call a worker *specialist* if she is endowed with an unbalanced set of skills, with one dominating skill, and *generalist* if she is endowed with a balanced set of skills. The subadditivity property (8) expresses that it is less costly for firms to hire a generalist worker with skill set $x = (x_1, \ldots, x_k)$ than k specialist workers endowed with the corresponding amount x_i of skill in each dimension.

We now describe in more detail the structure of the convex and homogenous wage schedules and connect our model to Roy (1951). To do this, we define the *implicit price* of skill i for workers of type x as $w_i(x) = \partial w / \partial x_i$. These implicit prices are homogenous of degree zero, and as such depend on skill profiles $\tilde{x} = x/|x|$ but not on workers' qualities $|x|$. Using Euler's homogenous function theorem and the convexity of wages, we get

$$w(x) = \sum_{i=1}^{k} w_i(x) x_i \geq w(y) + \sum_{i=1}^{k} w_i(y)(x_i - y_i) = \sum_{i=1}^{k} w_i(y) x_i. \tag{9}$$

In a Roy-like assignment model, workers would decide to self-select into their preferred option among the menu of linear wage schedules $\sum_{i=1}^{k} w_i(y) x_i$ indexed by y. In a Roy-model context, Equation (9) would be thought of as an incentive constraint expressing that a worker with skills $x = (x_i)_{i=1,\ldots,k}$ prefers the linear schedule "designed for her", i.e., chooses $y = x$.[19] By contrast, our paper's modeling framework (under skill bundling) involves no supply-side decisions on workers' side. Hence, Equation (9) is purely demand-driven: it results from the structure of our production function, in particular from the aggregation of skills within firms.

Geometrically, convex and homogenous wage schedules are entirely determined by the associated iso-wage surfaces $w(x) = 1$, i.e., the sets of skill types that firms can

[19] The empirical results of Section 6.3 illustrate that a worker is paid less if she deviates from $y = x$ and in this sense is not well "matched".

obtain in return for one dollar. Figure 1 shows that the iso-wage surfaces are the envelopes of their tangents.[20]

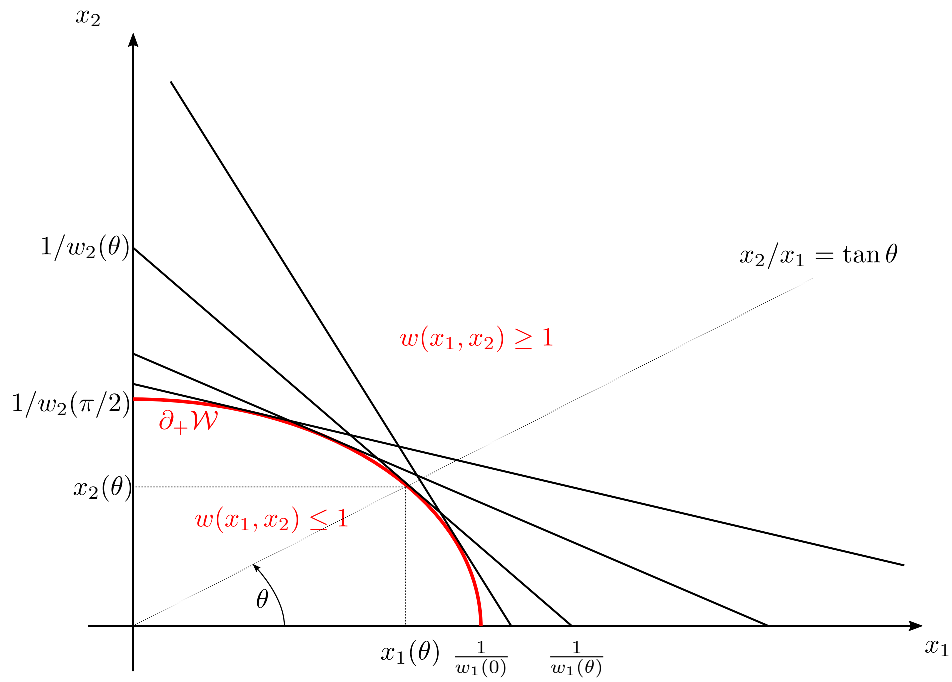


Figure 1: The set of workers paid less than one dollar is convex. The implicit prices of skills 1 and 2 for workers with skill profile θ are $w_1(\theta)$ and $w_2(\theta)$

In the case of two skills, $k = 2$, the worker's skill profiles $\tilde{x} = (x_1/|x|, x_2/|x|)$ can be parameterized as $\tilde{x} = (\cos\theta, \sin\theta)$, where θ belongs to $[0, \pi/2]$. For brevity, we often refer to θ as the worker's skill profile. The worker's comparative advantage in skill 2 over skill 1 is simply $x_2/x_1 = \tan\theta$. The implicit prices of the two skills, $w_1(\theta)$ and $w_2(\theta)$, depend only on the profile θ. 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 $\theta' = \theta$. As shown on Figure 1, 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'.[21] The literature that deals with multidimensional 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, both of these papers assume two sectors with homogenous firms within each sector and a sector-specific wage schedule, in other words they restrict attention to two-part wage schedules.

[20]See the nonlinear pricing literature, e.g., Wilson (1993) and Laffont and Martimort (2009)).

[21]The iso-wage curve $w = 1$ can transparently be parameterized as $(x_1(\theta), x_2(\theta))$, with $x_1(\theta) = \cos\theta/\tilde{w}(\theta)$ and $x_2(\theta) = \sin\theta/\tilde{w}(\theta)$.

3.2 Aggregate Sorting

The firms' problem (6) can be broken down into two subproblems that consist respectively in finding the firm-aggregated skill vector $T(\phi) = \int xN^d(\mathrm{d}x;\phi)$ and in achieving that aggregate vector in the most economical way. In this subsection, we study the properties of the aggregated skill $T(\phi)$ and examine how it varies with the firms' technological characteristics ϕ.

Proposition 3 (Uniqueness of the firm-aggregated vector of skills)**.** *Suppose Assumption 1 holds and assume furthermore that $F(T;\phi)$ is* strictly *concave in T. Then the firm-aggregated skill vector $T(\phi) = \int xN^d(\mathrm{d}x;\phi)$ is unique among all optimal market clearing assignments N^d. It solves*

$$\Pi(\phi; w) = \max_T F(T;\phi) - w(T), \tag{11}$$

where w is any equilibrium wage schedule that is convex and homogenous of degree one.

Since F is concave and w is convex, the above problem is well-posed, with a unique solution characterized by

$$F_j(T(\phi);\phi) = w_j(T(\phi)). \tag{12}$$

At any competitive equilibrium, the productivity of each skill equals its marginal price. When the wage schedule is locally linear, i.e., is of the form $< \bar{p}, x >$, we are back to $F_j(T(\phi);\phi) = \bar{p}$, i.e., price equals marginal productivity. Otherwise, the *implicit* price of skill i in the neighborhood of the aggregate skill T is the partial derivative $w_i = \partial w/\partial x_i$ evaluated at that point. Figures 2 and 4 show the tangency of the firm's production isoquant and the iso-wage surface.

From the envelope theorem, the firm-aggregated skill vector $T(\phi)$ can be expressed in terms of the firm' profit (11) provided that the function $T \to \nabla_\phi F(T;\phi)$ is invertible.[22] We check in the Appendix that the invertibility condition holds for the CES production function (7).

Corollary 1 (Envelope theorem)**.** *Assume that the function $T \to \nabla_\phi F(T;\phi)$ is invertible, and denote its inverse by $(\nabla_\phi F)^{-1}$. The firm-aggregated vector of skill $T(\phi)$ can be written as*

$$T(\phi) = (\nabla_\phi F)^{-1} \nabla_\phi \Pi(\phi; w). \tag{13}$$

In the rest of this subsection, we study how the aggregate vector $T(\phi)$ varies with the firm's type ϕ. We distinguish the (quality-adjusted) size of a firm and the aggregate

[22] The latter condition is stronger than the twist condition defined in Chiappori, McCann, and Pass (2016), which requires only injectivity: For any ϕ, $T \neq T'$ implies $\nabla_\phi F(T;\phi) \neq \nabla_\phi F(T';\phi)$.

profile of its employees. Specifically, we write the firm-aggregated skill vector of firm ϕ as $T(\phi) = \Lambda(\phi)\tilde{X}^d(\phi)$, where $\Lambda(\phi) = |T(\phi)|$ is the total quality of the firm's employees and $\tilde{X}^d(\phi)$ is their average skill profile.

Corollary 2 (Matching of aggregate skill profiles)**.** *Assume that production functions have homothetic isoquants. Then, when a firm's technology is more intensive in skill j, it uses relatively more of that skill.*

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

The aggregate profile of the workers employed by a firm therefore depends on the marginal rates of technical substitution. When ϕ takes the form $\phi = (\alpha, z)$, where z reflects total factor productivity, i.e., $F(T, \phi) = zF(T, \alpha)$, these rates do not depend on TFP, z. As a consequence, the same is true for aggregate skill profile: $\tilde{X}^d(\phi)$ depends only on the technological intensity parameters α that reflect the importance of each task. This is the case for instance in our leading example (7), for which $F_j/F_j = (\alpha_j/\alpha_k)(X_k/X_j)^{1-\rho}$.

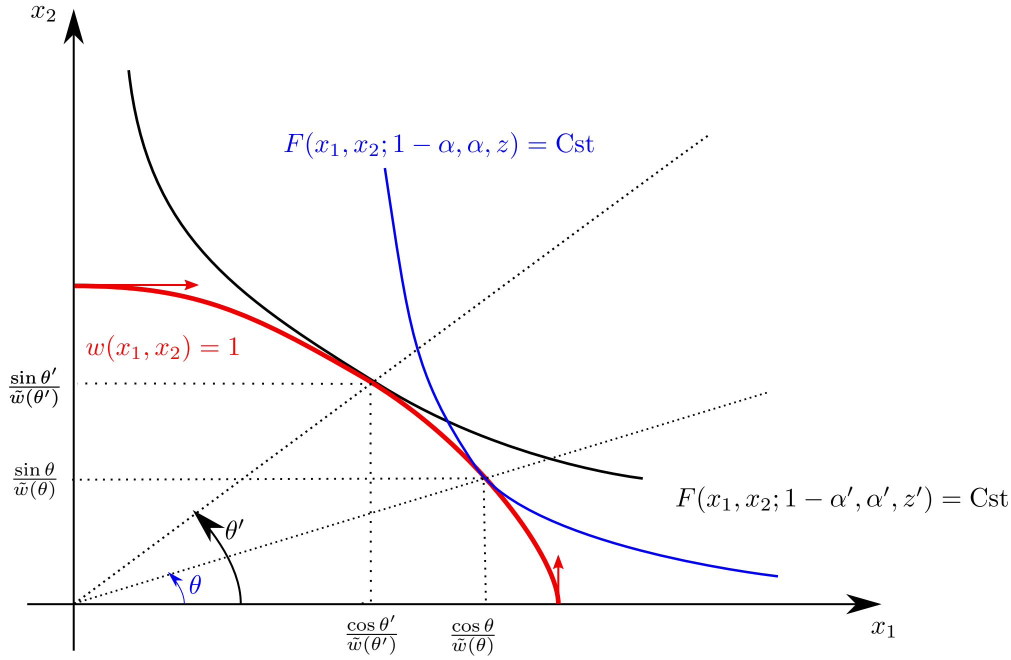


Figure 2: Matching in the skill dimension: Firm $(1-\alpha, \alpha, z)$ is more intensive in skill 1 than firm $1-\alpha', \alpha', z$.

Corollary 3 (Homogenous production functions and TFP)**.** *Assume furthermore that the production functions are homogenous of degree $\eta < 1$. Then the firm-aggregated* *intermediary input T, the firm's wage bill, and the firm's profits are proportional to $z^{1/(1-\eta)}$, where z denotes firm's total factor productivity.*

Two tasks: As explained above, when $k = 2$, we may represent the firm-aggregated skill vector $T = (\Lambda^d \cos\theta^d, \Lambda^d \sin\theta^d)$ in polar coordinates, where Λ^d is the total quality of workers employed at firm ϕ.

Proposition 4. *Assume that there are two skills/tasks and that the production $zF(T;\alpha)$ is concave in T. Then the total quality of the workers employed by a firm, $\Lambda^d(\alpha, z)$, increases with the firm's total factor productivity z.*

Assume furthermore that the production functions have homothetic isoquants and that F_2/F_1 increases with α. Then the firm-aggregated matching $(\theta^d(\alpha, z), \Lambda^d(\alpha, z))$ exhibits positive assortative matching in the sense of Lindenlaub (2017).

Hence, total quality Λ^d increases with TFP z. In addition, with homothetic isoquants, 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.[23] In contrast to Lindenlaub (2017), however, 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 5.1 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, as we discuss in Section 3.4.

CES with two tasks example We consider the production function (7):

$$zF(T_1, T_2; \alpha) = \frac{z}{\eta}\left[(1-\alpha)T_1^\rho + \alpha T_2^\rho\right]^{\eta/\rho}.$$

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

$$\left[\tan\theta^d(\alpha)\right]^{1-\rho} = \frac{\alpha}{1-\alpha}\,\frac{w_1\left(\theta^d(\alpha)\right)}{w_2\left(\theta^d(\alpha)\right)}. \tag{15}$$

The matching between workers and firms is represented by the increasing function $\theta^d(\alpha)$ implicitly defined by (15). The relative skill endowment in skill 2 of the workers, $\theta^d(\alpha)$, increases with the demand intensity in skill 2, α, as illustrated on Figures 2 and 4. Equation (A.6) in the appendix gives the aggregate quality of the firms' employees.

[23] In Appendix A.4, we provide a sufficient condition for PAM that does not require homothetic production isoquants, see inequality (A.12).

3.3 Pure Sorting in the Horizontal Dimension

We now examine the matching of worker types to firm types represented by the transport plan π given by (4). In this subsection as well as in the next one, we focus on the horizontal dimension, i.e., on the skill profiles $x/|x|$ of workers employed by any given firm. To do this, we examine the second part of a firm-ϕ's problem, namely achieving the aggregated skill vector $T(\phi)$ in the most economical way:

$$w(T(\phi)) = \inf \left\{ \int w(x) N^d(\mathrm{d}x) : N^d \in \mathcal{M}(\mathcal{X}), \int x N^d(\mathrm{d}x) = T(\phi) \right\}, \tag{16}$$

where w is convex and homogenous of degree one.

We start with the case where the iso-wage surface $\partial_+\mathcal{W}$ is strictly concave. Under this circumstance, the minimization of the wage bill at a given aggregate skill in (16) imposes that firm ϕ hires only workers with skill profile $\tilde{X}^d(\phi) = T(\phi)/\Lambda(\phi)$. It follows that the support of the matching transport π is included in the graph of $\tilde{X}^d(\phi)$.

To characterize the equilibria under strict concavity of the iso-wage surface $\partial_+\mathcal{W}$, we first notice that, for any skill vector x, the wage earned by a worker of type $\tilde{x} = x/w(x)$ is equal to one, or equivalently $\tilde{x}$ belongs to $\partial_+\mathcal{W}$. It follows that the integral $\int \lambda H^f(\mathrm{d}\lambda|\tilde{x})$ represents the sum of the wage earned by workers with the same skill profile as $\tilde{x}$. More generally, 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 $\partial_+\mathcal{W}$:[24]

$$W_\# H = \left(\frac{x}{w(x)} \right)_\# w(x) H. \tag{17}$$

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

Proposition 5. *When the iso-wage schedule surface is strictly concave, all employees within the same firm share the same skill profile, i.e., the matching is pure in the horizontal dimension*

$$\textit{Support}\ \pi \subset \{(\tilde{X}^d(\phi) \times \mathbb{R}_+\,,\ \phi) \mid \phi \in \varPhi\}. \tag{18}$$

In equilibrium, the total value of efficiency units of labor offered by workers and demanded by firms coincide for each skill profile separately. Formally, we have

$$W_\# H^w = W_\# T_\# H^f, \tag{19}$$

[24] The push-forward operator is defined in Appendix A.5.

where W is given by (17).

When the iso-wage is strictly concave, any 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 $T_\# H^f$ represents the demand for skill vectors expressed by all firms in the economy.[25] The measure $W_\# T_\# H^f$ is based on a weighted sum of skill vectors with the same profile (using wages as weights) and can be thought of as the demand for skill *profiles*, while similarly $W_\# H^w$ is the (weighted) supply of skill profiles in the economy. The equilibrium conditions (19) says that the demand and supply of skill profiles coincide. It translates into an ordinary differential equation for the matching map as we now illustrate in the case of two tasks.

Back to the two skills-tasks example: Assume that the production function is homogenous of degree $\eta < 1$ and F_2/F_1 increases with α as in Proposition 4. As above, the firm-aggregated skill vector is represented as $T = (\Lambda^d \cos\theta^d, \Lambda^d \sin\theta^d)$, where Λ^d is the total quality of workers employed at firm ϕ. The workers-to-firms matching condition (14) can be written in this context

$$\frac{F_1\left(\cos\theta^d(\alpha), \sin\theta^d(\alpha); \alpha\right)}{F_2\left(\cos\theta^d(\alpha), \sin\theta^d(\alpha); \alpha\right)} = \frac{w_1\left(\theta^d(\alpha)\right)}{w_2\left(\theta^d(\alpha)\right)}, \tag{20}$$

which implicitly defines an increasing matching map $\theta^d(\alpha)$. Setting $\tilde{w}(\theta) = w(\cos\theta, \sin\theta)$ as in (10), and using expression (A.7) for the wage bill of firm $\phi = (\alpha, z)$, we can write the equilibrium condition (19) for any α as

$$\int_0^{\theta^d(\alpha)} \Lambda^w(\theta)\tilde{w}(\theta)H^w(\mathrm{d}\theta) = \int_0^{\alpha} Z^f(\alpha)\left[F\left(\frac{\cos\theta^d(\alpha)}{\tilde{w}(\theta^d(\alpha))}, \frac{\sin\theta^d(\alpha)}{\tilde{w}(\theta^d(\alpha))}; \alpha\right)\right]^{1/(1-\eta)} H^f(\mathrm{d}\alpha), \tag{21}$$

where $\Lambda^w(\theta) = \int_z \lambda H^w(\mathrm{d}\lambda|\theta)$ and $Z^f(\alpha) = \int_z (\eta z)^{1/(1-\eta)} H^f(\mathrm{d}z|\alpha)$ are exogenous quantities that depend on the primitive distributions H^f and H^w. The left-hand side of (21) represents the total wages earned by workers with skill profile below $\theta^d(\alpha)$. According to (A.7), the right-hand side represents the total wage bill paid by the employing firms of those workers, namely all the firms with technological parameter below α.

[25] $T_\# H^f$ is the push-forward of the distribution of the firms' technological parameters H^f by their skill aggregate skill demand T, see Appendix A.5.

Differentiating with respect to α yields the ordinary differential equation for the matching map $\theta^d(\alpha)$

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

where h^f and h^w are the densities of the distributions of θ and α. Equation (22) relates the matching map $\theta^d(\alpha)$ implicitly given by (20) and its derivative $\mathrm{d}\theta^d/\mathrm{d}\alpha$ to the distributions of workers' skills and firms' technologies. It follows that for any strictly wage schedule $w(x)$ such that $\partial_+\mathcal{W}$ is strictly concave, any homogenous production functions $zF(.;\alpha)$ such that F_2/F_1 increases with α, and any skill 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 (22) only determines (for any α) the quantity $Z^f(\alpha)h^f(\alpha)$ that drives the demand for workers with skill profile $\theta^d(\alpha)$ by firms with intensity α in skill 2.

3.4 The Impact of Bunching

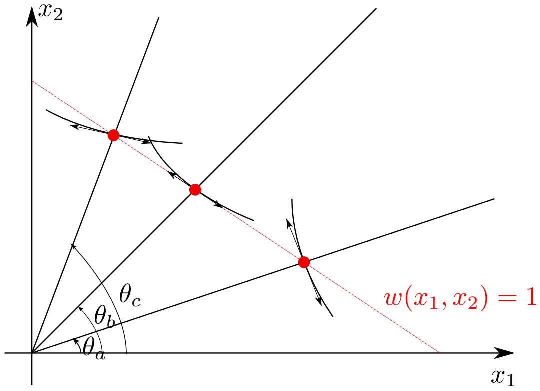
We now turn to situations in which different firm-types hire workers with similar skill-types (albeit never using the same combination because of the aggregate workers-to-firms sorting condition). We refer to this phenomenon as bunching. First, we explain intuitively how bunching can arise in equilibrium, 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, the skill profiles of workers *within firms* become heterogeneous, in short, bunching emerges.

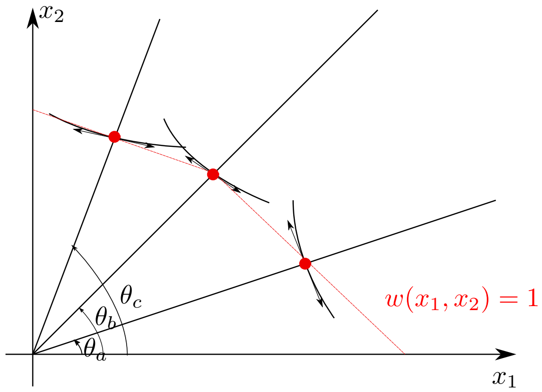
We illustrate the mechanism in a setting with two tasks and three skill profiles $\theta_a < \theta_b < \theta_c$, see Figure 3. Recall $\tan\theta_i = x_{i2}/x_{i1}$ is the endowment of workers $i \in \{a, b, c\}$ in skill 2 relative to skill 1. We pick any $w_1 > 0$ and $w_2 > 0$ and construct distributions H^w and H^f for which the linear wage schedule $w(x_1, x_2) = w_1x_1 + w_2x_2$ prevails in equilibrium. We choose three values for the technological intensities in skill 2,

α_k, $k \in \{a, b, c\}$, such that

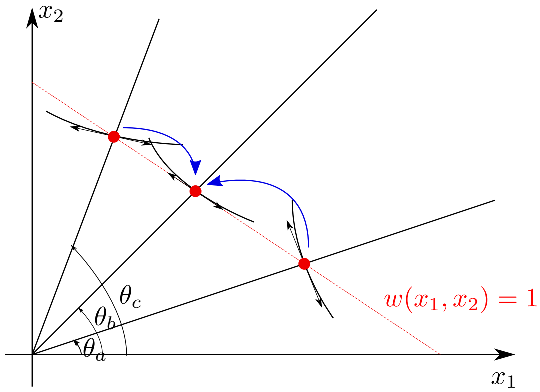
$$\frac{1-\alpha_c}{\alpha_c}\,(\tan\theta_c)^{1-\rho} < \frac{w_1}{w_2} = \frac{1-\alpha_b}{\alpha_b}\,(\tan\theta_b)^{1-\rho} < \frac{1-\alpha_a}{\alpha_a}\,(\tan\theta_a)^{1-\rho}\,.$$



(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 endowed 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 (22) takes the form

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

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

We now slightly increase the (quality-adjusted) number of generalist workers in the economy, $\Lambda^w(\theta_b)\, h^w(\theta_b)$. 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, as well as the behavior of firms with type a and c. The wage schedule, however, has become nonlinear.

To generate bunching, we on the contrary decrease the number of generalist workers 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 $T(\phi)$ 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.

Characterization of equilibrium under bunching: When the wage schedule is strictly concave as was assumed in Subsection 3.3, all the points of the iso-wage surface $\partial_+\mathcal{W}$ are extremal points of $\mathcal{W}$. Extremal points are degenerated faces of $\mathcal{W}$.[26] By contrast, when the schedule is locally linear, the set $\mathcal{W}$ has proper faces, i.e., faces that are neither a singleton nor the whole set $\mathcal{W}$ itself. For instance, on Figure 4, the segment $[AB]$ is a proper face of $\mathcal{W}$, while A is an extremal point. We now use the faces of $\mathcal{W}$ to characterize the equilibria under bunching.

Rockafellar (1970), Theorem 18.2., states that any convex set is the *disjoint* union of the relative interiors of all its faces. For any T, let $\mathcal{F}(T)$ be the (unique) face of $\mathcal{W}$

[26]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.

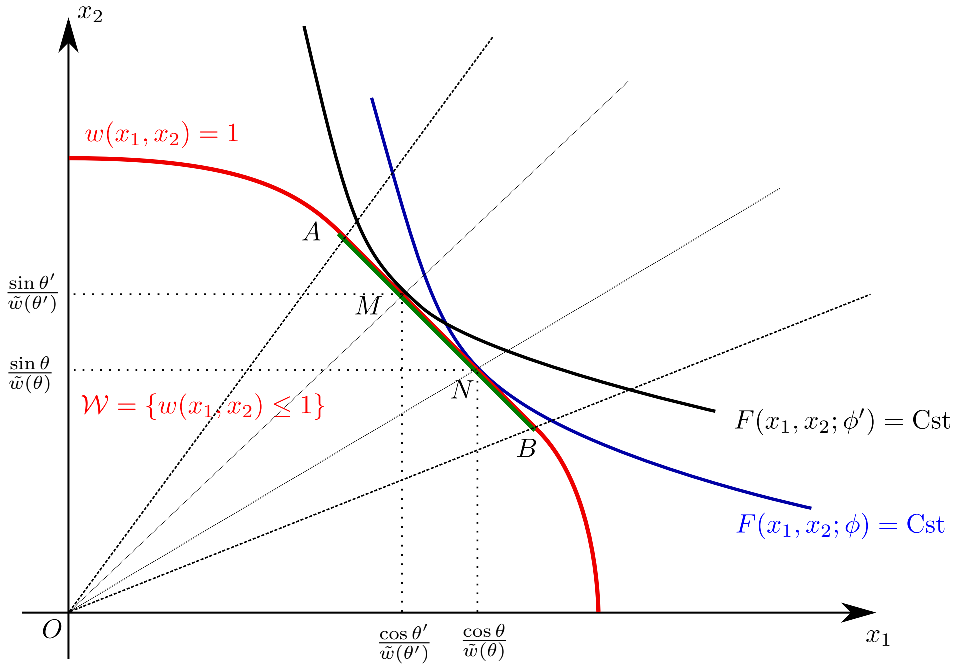


Figure 4: Matching is not pure. Firms $\phi = (1-\alpha, \alpha, z)$ and $\phi' = (1-\alpha', \alpha', z')$, pick their employees in the cone generated by the face $[AB]$ of $\mathcal{W}$ in $\mathbb{R}^2_+$. Firm ϕ' is more intensive in skill 2: $\alpha' > \alpha$ and $\theta^d(\alpha') > \theta^d(\alpha)$.

such that $T/w(T)$ belongs to the relative interior of $\mathcal{F}(T)$. The cone

$$\mathcal{C}\left(T(\phi)\right) = \mathcal{F}\left(T(\phi)\right) \times \mathbb{R}_+ \tag{23}$$

is the largest set $\mathcal{C}$ in $\mathcal{X}$ such that (i) w is linear on $\mathcal{C}$; and (ii) the relative interior of $\mathcal{C}$ contains $\tilde{X}^d(\phi)$, the average skill profile of workers employed by firm with type ϕ, see Lemma A.1 in the Appendix.

Figure 4 illustrates a case where w is linear on the non-degenerated cone lying between the rays (OA) and (OB). 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 $\{T/w(T)\}$ and the cone is reduced to a ray (the ray containing A in the example). For the firms ϕ and ϕ', $\mathcal{F}\left(T(\phi)\right)$ and $\mathcal{F}\left(T(\phi')\right)$ are equal to the segment $[AB]$, which generates the cone (AOB).

When the wage schedule w is locally linear, the minimization of the wage bill, problem (16), is compatible with a firm hiring employees with different skill profiles. To minimize the firm's wage bill, the support of the assignment measure $N^d(\mathrm{d}x; \phi)$ must be included in $\mathcal{C}\left(T(\phi)\right)$. Because the wage schedule w is linear on that cone, we have

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

For instance, firms with type 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 6. *When the equilibrium wage schedule is locally linear, the matching is not pure in the horizontal dimension*

$$\textit{Support}\ \pi \subset \{\mathcal{C}\left(T(\phi)\right)\,,\,\phi) \mid \phi \in \Phi\}, \tag{24}$$

where $\mathcal{C}\left(T(\phi)\right)$ *is the cone given by* (23)*. In equilibrium condition, the measure* $W_{\#}T_{\#}H^f$ *is dominated by* $W_{\#}H^w$ *in the convex order:*

$$W_{\#}H^w \succeq_C W_{\#}T_{\#}H^f \tag{25}$$

where the operator W *is given by* (17)*.*

When bunching prevails, it is no longer true that the total value of efficiency units of labor supplied by workers and demanded by firms coincide for each skill profile, i.e., that the distributions $W_{\#}T_{\#}H^f$ and $W_{\#}H^w$ are equal. Recall that a measure μ_1 is dominated by a measure μ_2 in the convex order if and only if $\mu_2 h \geq \mu_1 h$ for all convex functions h.[27] The condition (25), which is weaker than (19), expresses that there is a local excess supply of specialist workers and an excess demand for generalist ones. In terms of efficiency units of labor (valued by wage), the distribution of workers' skills H^w lies closer to the boundary of the cone than the demand distribution $T_{\#}H^f$. For instance, on Figures 4 and 5, the supply of skills is more concentrated along the rays OA and OB, while the demand is more concentrated in the interior of the cone.

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 workers with the same type may be employed by firms with different technologies. For instance, firms F and F' on the figure, which have different technological intensities in skill α, 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.[28]

To conclude this section, we connect our primal problem (5) to the classic optimal transport (OT) framework, used for instance in Lindenlaub (2017)'s study of worker-to-job matching. To fully understand how OT is connected to our contribution, a small

[27] It means that μ_2 is "riskier" than μ_1.

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

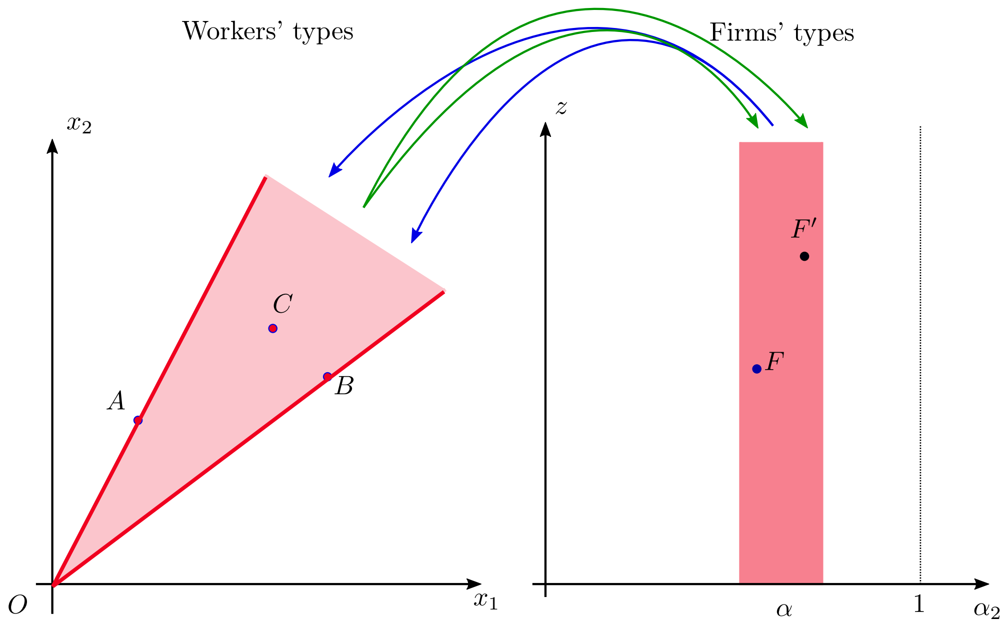


Figure 5: Sorting with bunching: Within-firm heterogeneity in skill profiles

detour is required. Our approach requires to account for workers' skill aggregation within firms, and endogenous firm size. These requirements demand a new mathematical framework, developed in Choné, Gozlan, and Kramarz (2022). In particular, it allows us to define precisely when one distribution is more "generalist" than another. Intuitively, in a two-skills world, it means that there are more generalists than specialists. Indeed, and back to our problem, we show in Appendix A.7 that the distribution of firm-aggregated skill vectors, $T_\# H^f$, is more "generalist" than the original distribution of workers' skills in the economy, H^w, in the sense that $\int h(x) T_\# H^f(\mathrm{d}x) \leq \int h(x) H^w(\mathrm{d}x)$ for all positively 1-homogenous convex functions h. When this property holds, Choné, Gozlan, and Kramarz (2022) say that $T_\# H^f$ is dominated by H^w in the positively 1-homogenous convex order, something we denote by $T_\# H^f \preceq_{phc} H^w$.

Proposition 7. *For any given map $T : \Phi \to \mathbb{R}^n_+$, the two properties are equivalent:*

1. *There exists a market clearing assignment N^d such that $T(\phi)$ is the firm-aggregated skill vector $T(\phi) = \int x N^d(\mathrm{d}x; \phi)$;*

2. *The probability distributions $T_\# H^f$ and H^w satisfy: $T_\# H^f \preceq_{phc} H^w$.*

Furthermore, if N^d is an optimal market clearing assignment, $T_\# H^f$ is solution to

$$Y^* = \max_{\gamma \preceq_{phc} H^w} \max_{\pi \in \Pi(\gamma, H^f)} \int F(x; \phi) \pi(\mathrm{d}x \mathrm{d}\phi). \tag{26}$$

The first part of Proposition 7 states that the ordering $T_{\#}H^f \leq_{phc} H^w$ is not only necessary but also sufficient for T being generated by a skill-aggregation process. The second part, namely equation (26), expresses that the optimal output under bundling, see (5), is the maximal output that can obtained *without* skill-aggregation *i.e., with classic OT* among all skill distributions that are "more generalist" than the original distribution H^w.

Hence, when there are enough generalist workers in the economy, there is no bunching: $T_{\#}H^f =_{phc} H^w$ as in Proposition 5.[29] If, on the contrary, $T_{\#}H^f$ is strictly dominated by H^w in the convex positively homogenous order – for instance if there are mostly specialist workers in the economy – then $T(\phi)$ is obtained by using workers with different skill profiles as in Proposition 6. In the latter case, there is within-firm heterogeneity in *skill profiles*. As a consequence, in equilibrium, complementarities across workers within the firm materialize. Hence, the productivity of workers endowed with (mostly) one skill and deprived of the other skills is enhanced by the presence of co-workers endowed with the other, complementary, skills.

4 From Bundling to Unbundling: Endogenizing the Supply of Skills

In this Section, we continue to assume that skills and tasks are evaluated in the same metric, i.e., one unit of a skill supplied by a worker corresponds to one unit of the corresponding task (intermediary input) used by firms.

Another assumption, adopted until now – a firm uses the exact endowment of the workers it hires – constitutes a clear limit of our basic setup. In the following, we relax this assumption in two directions. These two directions are not opposed but rather complementary and *it is simple to combine them in practice*, even though exposition is easier when each one is presented separately.

A first manner to relax the fixed endowment assumption, while *continuing to assume* that a worker's skills are bundled, is to allow each worker (resp. each firm) to decide how much of each skill she will supply to her employing firm (resp. to decide how much of each skill will be supplied by each worker) at a "cost". Indeed, we will explicitly allow each worker to choose the exact quantity of inputs supplied to the firm *within a set.* This quantity will maximize her compensation given the equilibrium prices of skills. The shape of the allowed set will reflect the trade-offs implied in converting one skill into another, hence the associated "cost". It is intended to reflect each worker's production function when she manages her time (even though time is never explicitly introduced).

[29]The projections of the distributions $T_{\#}H^f$ and H^w onto the iso-wage surface coincide.

Indeed, a worker may potentially exhaust oneself in a cognitive task if using all her endowment in that cognitive skill. It may therefore be optimal to convert some of her time for a less cognitive one. Or, by contrast, a worker may decrease her time in a non-cognitive task (manual, for instance) to increase her time in a more cognitive one. However, the conversion rate will be below one, to reflect potential exhaustion, boredom, or skills transformation costs. We also present the "mirror" problem with firms making this decision for their workers.

A second manner to relax the *fixed-labor-supply-to-a-single-firm assumption*, while *not assuming anymore* that a worker's skills are bundled, is "Unbundling". Unbundling may take place when a market for each skill is opened. In this setup, we allow each worker to supply skills to the market in addition to those supplied to a firm.[30] Therefore, a worker's labor supply to, respectively, the main employing firm and the platforms becomes endogenous. In what follows, we discuss both full unbundling (i.e. with no associated cost) and costly unbundling (e.g. a platform aggregates workers' skills and supplies them at a cost paid, say, by the firm). In practice, one may ask what do platforms or temp agencies trade ? Skills or tasks. In this Section, we examine the role of markets equating skills and tasks. However, we have another look at this question in Subsection 5.4, when trying to understand how skills and tasks are connected.

To simplify the exposition, we assume in the remainder of this Section that workers are endowed with two different skills, i.e., $k = 2$.

4.1 Endogenous Supply of Skills

In this Subsection, we maintain the bundling restriction, i.e., only bundles of skills can be traded. But we examine two closely related variants: in the first, the workers are allowed choose to specialize into a particular skill when, in the second, the firm decides how workers specialize into their skills. We study them in turn.

When workers decide the specialization in skills: In this setting, a worker endowed with skills $x = (x_1, x_2)$ may use part or all of her endowment x_1 to produce more of skill 2. Specifically, the worker can produce any bundle of skills $s = (s_1, s_2)$ in the set

$$S(x_1, x_2) = \{ (s_1, s_2) \mid \tau s_1 + s_2 \leq \tau x_1 + x_2 \},$$

where $\tau \geq 0$ is an economy-wide parameter that reflects the conversion rate of skill 1 into skill 2. A worker of type x can thus choose to offer any skill in the set $S(x)$.

[30] When the markets for skills operate through platforms, a worker will be allowed to supply her skills to, at least, two firms.

An example of such set is represented on Figure 6. A choice of skill $s(.)$ can thus be described as a selection of the set-valued function S.

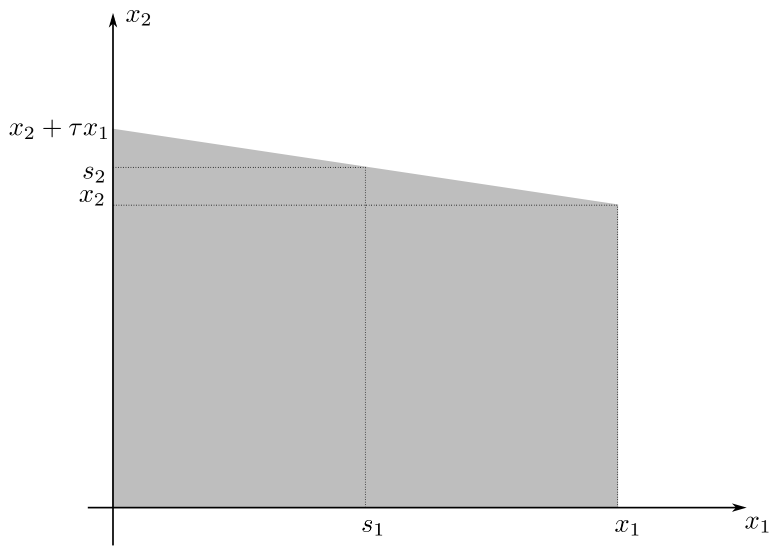


Figure 6: Technology set $S(x_1, x_2)$ of worker with type (x_1, x_2). The worker can produce any couple of skills (s_1, s_2) in the shaded area

In this context, the maximization of output in the economy, i.e., the equivalent of the primal problem (5), is given by

$$Y^* \stackrel{\mathrm{d}}{\equiv} \sup_{s(x) \in S(x)} \sup_{N^d \,|\, N^d H^f = H^w} \int F\left(T(\phi); \phi\right) H^f(\mathrm{d}\phi), \tag{27}$$

where the firm-aggregated vector of tasks $T(\phi)$ is given by

$$T(\phi) = \int s(x) N^d(\mathrm{d}x; \phi). \tag{28}$$

Equation (28), which replaces equation (1), reflects that the link between individual skills and firm-aggregated tasks is now endogenous.

In a competitive environment, let w denote the wage schedule. A worker with initial skills x who sells her transformed bundle of skills $s(x)$ earns $w(s(x))$. An equilibrium is defined as a triplet (N^d, t, w) made of a market-clearing assignment of workers to firms $N^d(\mathrm{d}x; \phi)$, a skills transformation function t, and a wage schedule w such that:

(i) the assignment N^d reflects the demand for skills under the wage schedule w, i.e., is solution to the firms' problem

$$\Pi(\phi; w) = \max_{N^d} \; F\left(\int s(x) N^d(\mathrm{d}x; \phi); \phi\right) - \int w(s(x)) N^d(\mathrm{d}x; \phi); \tag{29}$$

(ii) the transformation function reflects the supply of skills under the wage schedule w, i.e., t is solution to the workers' problem

$$U(x; w) = \max_{s \in S(x)} w(s). \tag{30}$$

When firms decide the specialization in skills: The employer assigns the worker with skill x to the new skills set, $s(x)$, (something equivalent to a "task" set, as in Haanwinckel (2020)). In this case, the selection s is decided by the firms, not by the workers.

(i) Equation (29) must be changed as

$$\Pi(\phi; w) = \max_{N^d, s} F\left(\int s(x) N^d(\mathrm{d}x; \phi); \phi\right) - \int w(x) N^d(\mathrm{d}x; \phi). \tag{31}$$

Notice that compared to (29), $w(x)$ now replaces $w(s(x))$ in the wage bill;

(ii) No-arbitrage for firms and workers $w(s(x)) \geq w(x)$ (otherwise it would be cheaper for the firm to hire $s(x)$ than to hire x and require her to do $s(x)$) and $w(s(x)) \leq w(x)$ (otherwise the worker would sell $s(x)$ instead of x), so in fact we must have $w(s(x)) = w(x)$.

In both situations (workers or firms deciding), by the same argument as in Proposition 2, there is no loss of generality in restricting to wage schedules that are convex and homogenous of degree one. (Recall that these properties derive from the cost minimization carried out by firms.) In the present context, the endogeneity of the worker's labor supply yields the following additional restrictions on the shape of equilibrium wage schedules.

Proposition 8. *When workers' skill 1 can be transformed into skill 2 at rate τ, any equilibrium wage schedule satisfies*

$$\frac{w_1(x)}{w_2(x)} \geq \tau \tag{32}$$

for all skill vector $x \in \mathcal{X}$.

Recall that $w_1 = \partial w / \partial x_1$ and $w_2 = \partial w / \partial x_2$ denote the implicit prices of the two skills. The intuition is that any transformed skill-vector t for which (32) is not satisfied is irrelevant. If the ratio $w_1(t)/w_2(t)$ is lower than τ, then at the margin skill 2 is generously paid relatively to skill 1, and as a result the skill vector t is dominated for all workers by skill vectors of the form $(s_1 - \varepsilon, s_2 + \tau\varepsilon)$ for small $\varepsilon > 0$.

We now compare the shape of the wage schedule to the situation that prevailed in Section 3 where the supply of skills was exogenous, i.e. when the conversion rate was zero. Starting from that bundling situation, we assume that it becomes possible for workers to convert skill 1 into skill 2 at rate τ.[31] To make things interesting, we consider values of τ large enough for the constraint $w_1(x)/w_2(x) \geq \tau$ to be violated in some region under bundling. This means that, when conversion becomes possible at rate τ, workers with skill vectors in this region will indeed convert their skills, breaking the bundling equilibrium.

Using the algorithms developed in Paty, Choné, and Kramarz (2022), we numerically compute the equilibrium configurations with endogenous skill supply for various values of the conversion rate τ. Figure 7 assumes that the technical intensity in skill 2 – the parameter α – is uniformly distributed on $[0, 1]$. There is no heterogeneity in workers' quality or in firms' total factor productivity. The two skills are complements ($\sigma = -1$) and the returns to scale are decreasing ($\eta = .5$). It is also assumed that the workers' skill profiles are distributed as a Beta(9,9) random variable, so that specialist profiles are rare in the economy, and therefore very expensive under bundling. As a result, there is no bunching under exogenous skill supply: the implicit prices of skills 1 and 2, represented by the black lines on the Figure, are respectively decreasing and increasing with α over the whole interval $[0, 1]$.

When the workers can convert skill 1 into skill 2 at rate $\tau = .75$, the constraint $w_1/w_2 \geq \tau$ is binding for large values of α. The implicit prices are represented by the red lines on Figure 7. When the two lines are horizontal, the wage schedule is linear, the ratio of implicit prices w_1/w_2 equals τ. In this region, the implicit prices of skill 1 and skill 2 are respectively greater and lower than under bundling because there is respectively less (more) supply of that skill due to skill conversion. When the workers can convert skill 1 into skill 2 at rate $\tau = 1$, the implicit prices of the two skills are equal for large values of α. The skill conversion region – the horizontal part of marginal prices – widens as τ rises from zero to one.

In Figures 8(a) and 8(b) respectively, we show the change in sorting (so θ as a function of α) and the change in wages when τ moves from 0 (no specialization possible) to $\tau = 0.75$. Sorting barely changes for small values of α. For large values of α, where $w_1/w_2 = \tau$, sorting is driven by labor supply. The kink in the implicit implicit prices shown Figure 7 translates into a kink in the sorting map on Figure 8(a). Firms with large α benefiting from the conversion of skill 1 into skill 2 focus on their comparative advantage in production. As a consequence, the same firms increase their demand for skill 2, leading to increased polarization. But, this increase in demand is more than

[31] We discuss below, at the end of this Subsection, the symmetric case when both skills can be converted (into one another).

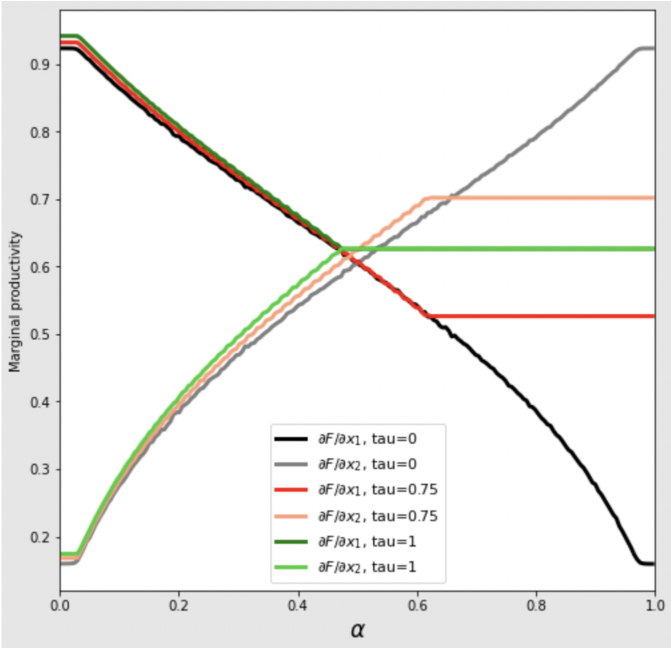
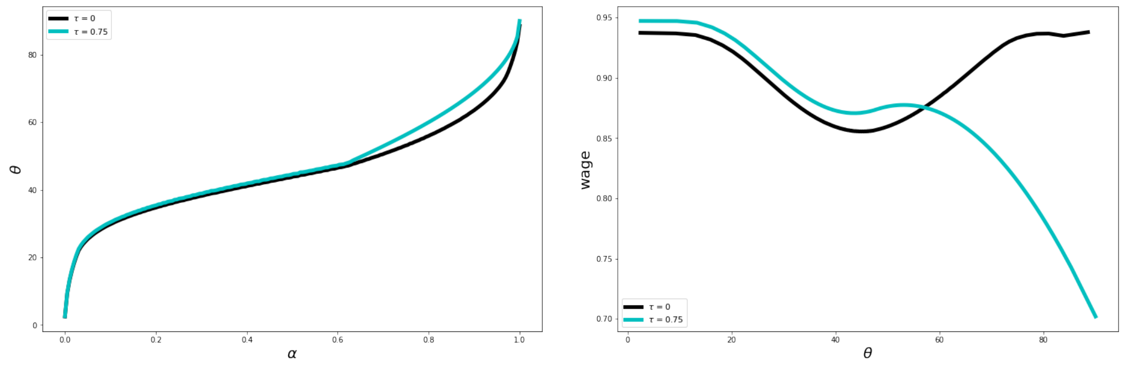


Figure 7: Implicit prices under different conversion rates. Workers can convert skill 1 into skill 2 at rate $\tau = .75$ (in Red) and at rate $\tau = 1$ (in Green). Skill conversion is not possible (bundling case, $\tau = 0$, in Black).

compensated by the supply of skill 2 from workers transforming their skill 1 into skill 2. The resulting effect is seen on Figure 8(b). Workers with a comparative advantage in skill 2 over skill 1 are negatively affected and their total wage falls. This is the reverse for those without such a comparative advantage who benefit from the decrease in the supply of skill 1 used to produce some skill 2 (not in their employing firm but in those with a large α). In other words, "specialist workers" (those being endowed mostly with skill 2) are harmed by increased competition from generalist workers (those endowed with a more balanced set of skills) taking advantage of the new possibility to convert. This redistributive effect will be a recurring theme in what follows.

Skill-biased technical change: Given a worker's conversion technology (i.e., the conversion rate of skill 2 into skill 1, τ, is fixed), the model can be used to understand the effect on our bundled economy of skill-biased technical change in firms' production functions. More precisely, as an example among many, in the following we focus on an increase in the TFP parameter z of those firms with technological intensity in skill 2 (α) greater than .75.[32] Figures 9, 10, and 11 show the effect of the change on respectively

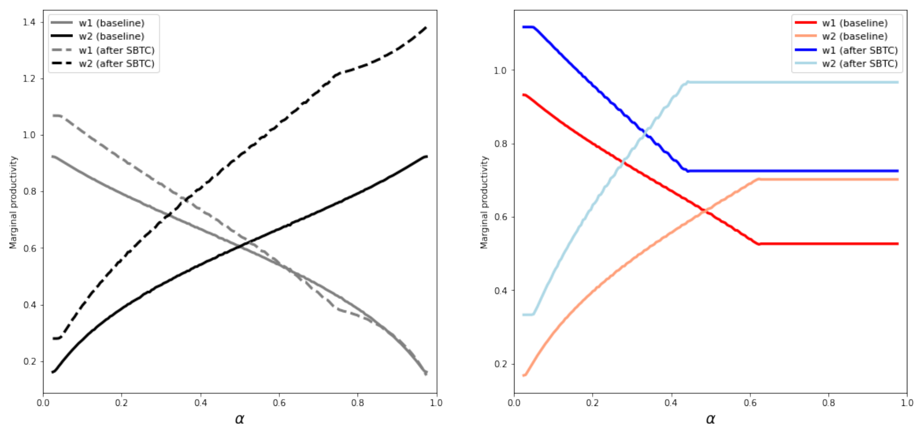
[32]Keeping our cognitive and non-cognitive example in mind, the shock benefits firms with production techniques requiring a lot of cognitive skills.



(a) Sorting (b) Wage

Figure 8: Sorting and Wage allowing for Worker Specialization : From $\tau = 0$ to $\tau = .75$

the implicit prices, sorting, and wages under exogenous labor supply (left panel) and when workers can convert skill 1 into skill 2 at rate $\tau = .75$.

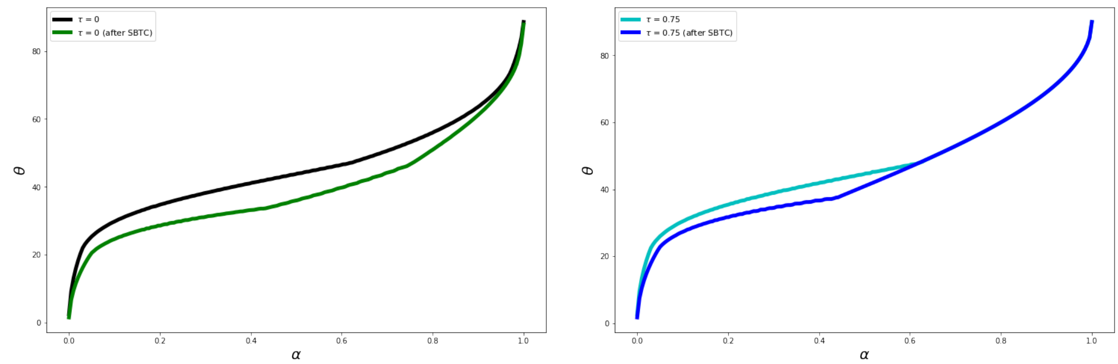


(a) No conversion possible ($\tau = 0$) (b) Conversion rate $\tau = .75$

Figure 9: Effect of skill-biased technical change (skill 2 intensive firms experience a positive productivity shock) on the implicit prices of the two skills

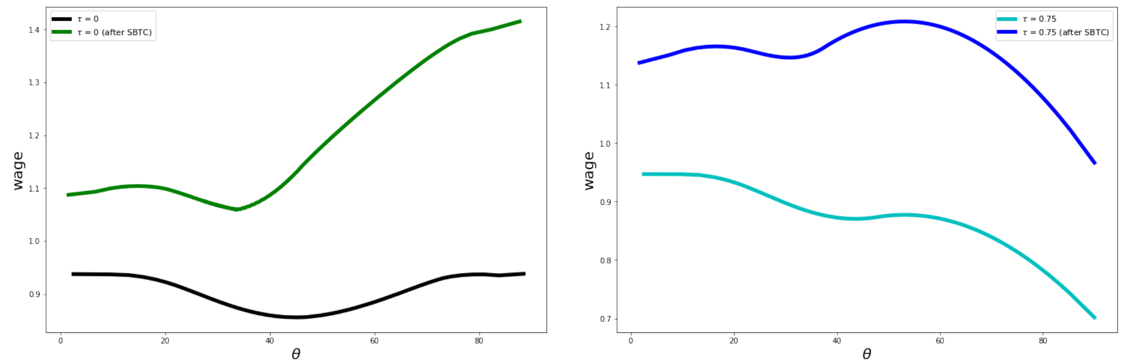
The direct effect of the shock is to increase demand for skill 2 in the firms hit by the SBTC shock (demand for workers with a strong comparative advantage in skill 2). As a result, the implicit price of skill 2 rises, see Figure 9. Because the price of skill 2 relative to skill 1 also rises, firms change their skill mix in favor of skill 1 (the sorting map $\theta(\alpha)$ is weakly lower after the shock, see Figure 10), which increases demand for skill 1 and hence the implicit price of that skill (see, again, Figure 9). As a consequence, all workers benefit from SBTC, see Figure 11.

The above effects are at work whether or not workers can convert skill 1 into skill 2. Yet the possibility to convert strengthens connections between the two skills at the equilibrium. Indeed, for large values of the skill profile x_2/x_1, the two implicit prices are linked through the constraint $w_2/w_1 = \tau$. Accordingly, for large values of the technical intensity α in skill 2, sorting is driven by the conversion rate τ. SBTC enlarges the region where the constraint is active, i.e., the region where workers specialize (i.e. convert skill 1 into skill 2), see Figure 9(b). As a result, the sorting maps before and after the shock coincide in the high end of the interval (see Figure 10(b)): labor supply adjusts itself so as to maintain sorting constant.[33] In this example, SBTC causes the fraction of skill 1 in the economy that is converted into skill 2 to increase from 6.3% to 22.0%. This increased transformation has a countervailing effect on the price of skill 2 through an expansion of its supply. As a result, after SBTC, specialist workers (those endowed mostly with skill 2) are worse off than generalist workers. This effect, already discussed above, is exacerbated by SBTC, see Figure 11(b).



(a) No conversion possible ($\tau = 0$) (b) Conversion rate $\tau = .75$

Figure 10: Effect of skill-biased technical change (skill 2 intensive firms experience a positive productivity shock) on the sorting map



(a) No conversion possible ($\tau = 0$) (b) Conversion rate $\tau = .75$

Figure 11: Effect of skill-biased technical change (skill 2 intensive firms experience a positive productivity shock) on wages

[33] As already observed, the kinks in the sorting apparent on Figure 10(b) correspond to the kinks in the implicit prices on Figure 9(b).

The above analysis assumes that skill conversion is possible only in one direction: skill 1 can be transformed into skill 2; the opposite being impossible. In an Appendix available upon request, we study a symmetric environment in which workers can convert each of their two skills into the other, at the same rate τ. Here again, the special case $\tau = 0$ corresponds to skills being exogenously supplied (bundling). At the other extreme, if $\tau = 1$, the two skills become one unique commodity, and must therefore have the same price, implying that the wage schedule is fully linear ($w_1/w_2 = 1$). The qualitative insights found in the case of symmetric skill conversion, in particular the flattening of the wage and the evolution of sorting as the conversion rate τ increases from zero to one, closely parallel those presented in the next Subsection and, hence, are not reported here.

4.2 Unbundling of Skills

In this Subsection, we allow workers to perform their skills outside employment relationships. We assume the availability of a technology that enables workers to unbundle their skills and allows workers and firms to trade these skills as commodities. In particular, a worker hired by a main employing firm can sell intermediary inputs to other firms, most likely by incurring a private cost to have his skills unbundled. If unbundling comes from an innovation (such as Uber which creates a market for driving skills), workers and/or users are likely to have to pay a fee corresponding to the platform' margin.

To be specific, a worker with skills $x = (x_1, x_2)$ can decide to unbundle and sell amounts m_1 and m_2 of skills 1 and 2, with $0 \leq m_1 \leq x_1$ and $0 \leq m_2 \leq x_2$. Setting $m = (m_1, m_2)$, the worker is then left with a skill bundle $x - m$ that represents the amounts of skill 1 and 2 available for her employing firm. We assume that unbundling m_i units of skill i entails a cost proportional to m_i, namely $c_i^w m_i$, with $c_i^w \geq 0$ for $i = 1, 2$. Setting $c^w = (c_1^w, c_2^w)$, the total unbundling cost incurred by the worker is $c^w.m = c_1^w m_1 + c_2^w m_2$.

Similarly, on the firms' side, acquiring amounts m_1 and m_2 of stand-alone skills involves a cost $c^f.m = c_1^f m_1 + c_2^f m_2$, with $m = (m_1, m_2)$, $c_1^f \geq 0$ and $c_2^f \geq 0$. The vector $c = c^f + c^w$ thus represents the total cost incurred by workers and firms per unit of unbundled skill for skill 1 and skill 2.

An allocation of skills in the economy consists of a triplet (N^d, m^d, m^s), where $N^d(\mathrm{d}x; \phi)$ is an assignment of workers to firms and the functions m^d and m^s specify the amounts of skills $m^d(\phi)$ and $m^s(x)$ purchased by firms of type ϕ and sold by workers of type x.

We define market-clearing allocations as allocations (N^d, m^d, m^s) that clear both the labor market and the markets for stand-alone skills, i.e., allocations such that the assignment N^d satisfies (2) and the functions m^d and m^s satisfy

$$\int m^d(\phi)H^f(\mathrm{d}\phi) = \int m^s(x)H^w(\mathrm{d}x). \tag{33}$$

The total output in the economy, net of unbundling costs, is

$$Y = \int F\left(T(\phi);\phi\right)H^f(d\phi) - \int c.m^s(x)H^w(\mathrm{d}x), \tag{34}$$

where the firm-aggregated vector of tasks $T(\phi)$ is given by

$$T(\phi) = m^d(\phi) + \int [x - m^s(x)]N^d(\mathrm{d}x;\phi). \tag{35}$$

Equation (35), which replaces equation (1), shows how the unbundling of skills endogenously affects the link between skills and firm-aggregated tasks within firms. Maximizing the net output (34) over all market-clearing allocations (N^d, m^d, m^s) is the equivalent under unbundling of the primal problem (5).

In a competitive environment, let $w(x)$ denote the wage schedule and $p = (p_1, p_2)$ denote the vector of market prices for stand-alone skills. An equilibrium is defined by a market-clearing allocation (N^d, m^d, m^s) and a price system (w, p) such that:

(i) The assignment N^d and the function m^d reflect the demand for skill bundles and for stand-alone skills under the wage schedule w and the market price vector p, i.e., N^d and m^d solve

$$\Pi(\phi; w, p) = \max_{m^d, N^d} F\left(T(\phi);\phi\right) - \int w(x - m^s(x))N^d(\mathrm{d}x;\phi) - (p + c^f).m^d, \tag{36}$$

where $T(\phi)$ is given by (35);

(ii) The function m^s reflects the supply of stand-alone skills by workers:

$$U(x; w, p) = \max_{m^s} w(x - m^s) + (p - c^w).m^s. \tag{37}$$

By the same argument as in Proposition 2, there is no loss of generality in restricting to wage schedules $w(t)$ that are convex and homogenous of degree one. (Recall that these properties derive from the cost minimization carried out by firms.) The possibility for workers to unbundle and sell stand-alone skills yields the following additional restrictions on the shape of equilibrium wage schedules.

Proposition 9. *When workers and firms can trade stand-alone skills, the range of implicit prices for skill $i \in \{1, 2\}$ cannot exceed c_i in equilibrium*

$$\max_x w_i(x) - \min_x w_i(x) \leq c_i, \tag{38}$$

where c_i is the total per-unit cost associated with the unbundling of skill i. If the inequality is strict, the market for skill i is inactive. If it holds as an equality, the prices perceived by firms and workers on the market for skill i are respectively $\max w_i = p_i + c_i^f$ and $\min w_i = p_i - c_i^w$, where p_i is the market price of that skill.

The bundling environment corresponds to infinite unbundling costs for all skills. Let us denote by w^b the equilibrium wage schedule under bundling. Suppose that for some skill i it becomes possible to unbundle skill i at a per-unit cost c_i that is lower than the difference $\max_x w_i^b(x) - \min_x w_i^b(x)$. 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 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 deviation that breaks the bundling equilibrium.

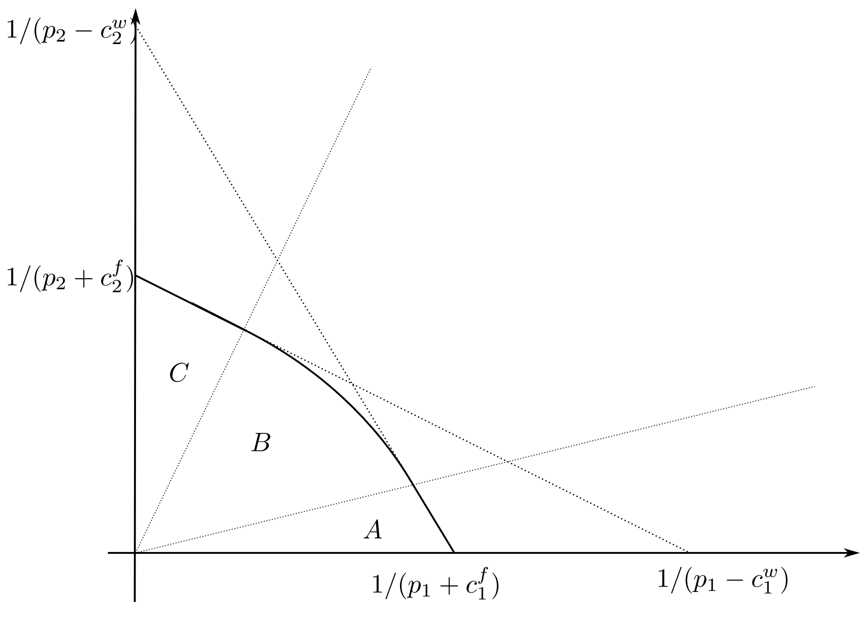


Figure 12: Iso-wage line under unbundling

The wage schedule and implicit prices are shown on Figure 12. 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 profile, 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.

Only a subset of skills may be traded on markets: The configuration shown on Figure 12 is compatible with only one market being active. Suppose for instance that the ranges of the implicit prices for skill 1 and 2 satisfy $\max w_1 - \min w_1 = c_1$ and $\max w_2 - \min w_2 < c_2$. In this case, the market for skill 2 remains inactive. The firm and worker prices $p_2 + c_2^f$ and $p_2 - c_2^w$, which do not exist, must be replaced with $\max w_2$ and $\min w_2$ on Figure 12. 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 firms hiring workers in Region 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.

Costless unbundling We now examine in greater detail the special case with no unbundling costs, $c = 0$. We use a superscript u to indicate costless unbundling. According to Proposition 9, any equilibrium wage schedule is fully linear and market prices coincide with implicit prices, $p_i^u = w_i^u$ for $i \in \{1, 2\}$. According to Proposition 11 and the first-order equation (12), all firms share the same marginal productivity for all skill types, i.e., $F_i(T(\phi);\phi)$ does not depend on the technological parameter ϕ. This situation corresponds full efficiency, i.e., the maximization of output with complete markets, see the primal problem (34) with $c = 0$.

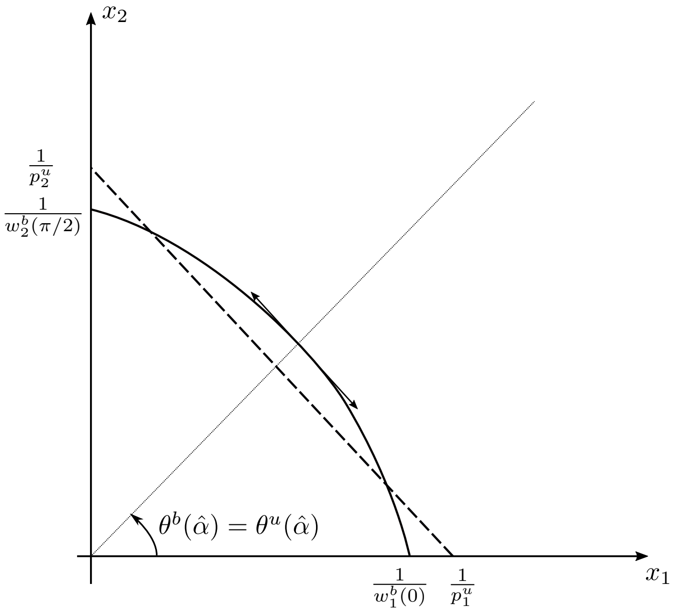


Figure 13: Iso-wage lines under bundling (solid) and costless unbundling (dashed)

Proposition 10 (Costless unbundling and polarization)**.** *Assume no unbundling costs, $c = 0$. Assume furthermore that the production function is of the form $zF(T;\alpha)$, where F is homogenous in T and F_2/F_1 increases with α on $[0,1]$.*

After unbundling some generalist workers are better off and if tasks are complementary inputs, some specialist workers are worse off. Specialized firms tend to specialize further, with their skill mixes being better aligned with their technologies.

Figure 13 represents the iso-wage curves under bundling and unbundling, $w^b(x) = 1$ and $w^u(x) = 1$. Figure 14 shows the corresponding matching maps. For firms of type $\hat{\alpha}$, the total amount of skill 2 divided by the total amount of skill 1 is the same under bundling and unbundling, $\theta^b(\hat{\alpha}) = \theta^u(\hat{\alpha})$. Workers with this skill profile are those who benefit the most from the unbundling of skills in the sense that the ratio

$$r(\alpha) = \frac{w^u(\tilde{X}^b(\alpha))}{w^b(\tilde{X}^b(\alpha))} = \frac{p_1^u \cos\theta^b(\alpha) + p_2^u \sin\theta^b(\alpha)}{w^b(\tilde{X}^b(\alpha))} \tag{39}$$

is maximal for $\alpha = \hat{\alpha}$. The ratio $r(\alpha)$ indicates how the unbundling of skills affects the earnings of the workers that are employed by firms of type α under bundling. We show in the Appendix that workers with skill profile $\tilde{X}^b(\hat{\alpha})$ are better off after unbundling,

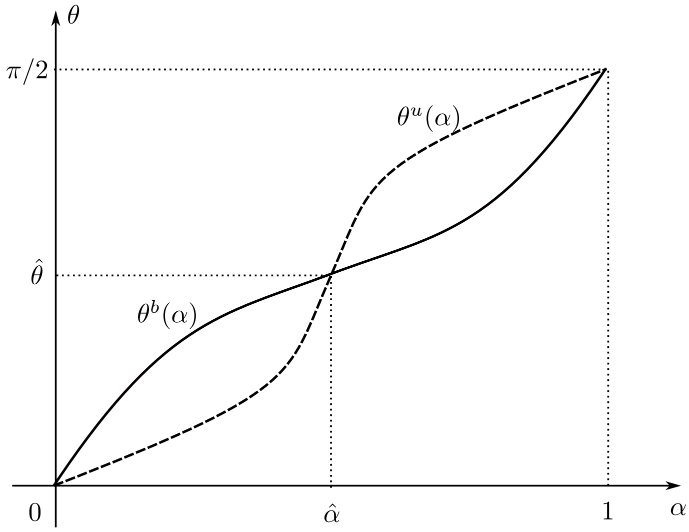


Figure 14: Polarization: Matching maps under bundling (solid) and under costless unbundling (dashed)

i.e., $r(\hat{\alpha}) > 1$, except in the case where the wage schedule under bundling is linear, $w^b = w^u$, and the two equilibria coincide.

By contrast, specialist workers tend to be harmed by the unbundling of skills because they face increased competition from the markets of stand-alone skills. A sufficient condition for some specialist workers to be worse off after unbundling is that the two skills are complementary inputs in the firms' production process.[34] In this case, the demands for each of the two skills are decreasing in both p_1 and in p_2 and therefore $p_1^u \geq \max w_1^b$ and $p_2^u \geq \max w_2^b$ would imply that all firms would reduce their demand for both skills after unbundling, which is impossible. As a consequence, except if the bundling and unbundling equilibria coincide, it must be the case that some specialist workers are harmed: $p_1^u < \max w_1^b$ or $p_2^u < \max w_2^b$.

For $\alpha \geq \hat{\alpha}$, the first-order conditions (14) under bundling and unbundling show that

$$\frac{F_1(T^b(\phi);\alpha)}{F_2(T^b(\phi);\alpha)} = \frac{w_1^b(\tilde{X}^b(\alpha))}{w_2^b(\tilde{X}^b(\alpha))} \leq \frac{p_1^u}{p_2^u} = \frac{F_1(T^u(\phi);\alpha)}{F_2(T^u(\phi);\alpha)},$$

which implies $\theta^b(\alpha) \leq \theta^u(\alpha)$. We may think of skill 2 as the core skill of firms with high values of the technological parameter α. The inequality $\theta^b(\alpha) \leq \theta^u(\alpha)$ shows that the composition of the skills used by firms is better aligned with their core skill after unbundling. Symmetrically, firms with low intensity for skill 2 ($\alpha < \hat{\alpha}$) tend to use relatively more of skill 1 after unbundling. This entails a polarization phenomenon.

[34] In the CES example (7), skills are complements if and only if $\rho < \eta$.

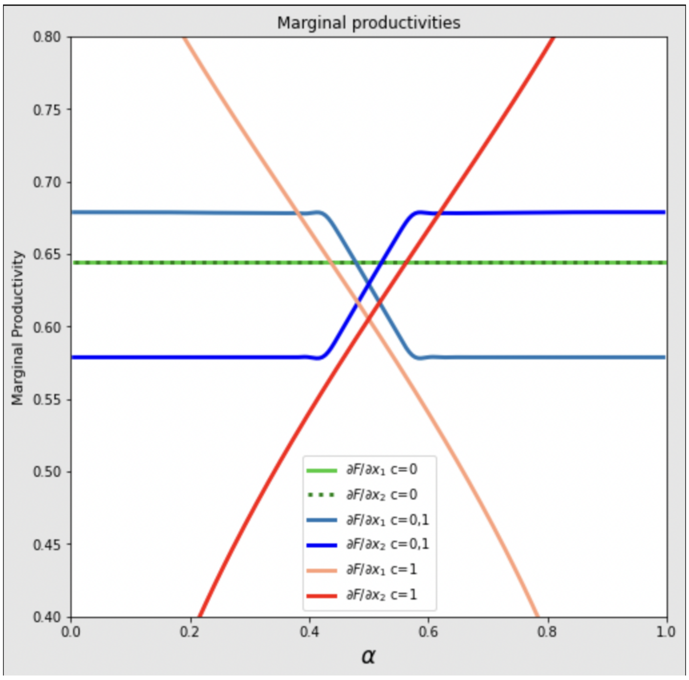
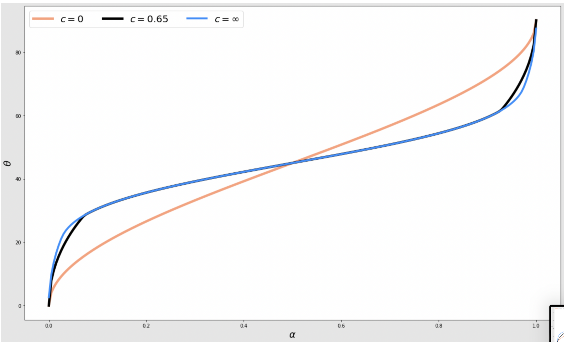


Figure 15: Implicit prices of skills under bundling (in Red), costly unbundling (Blue, $c_1 = c_2 = .1$), costless unbundling (Green). The implicit prices under bundling (in Red) have been truncated for readability.

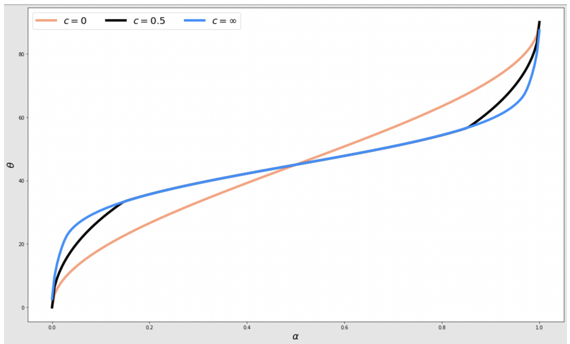
Firms with a high relative intensity in a skill use relatively more of that skill after unbundling than in the bundling equilibrium. The unbundling of skills allows specialized firms to specialize even further on their core skill.

From bundling to unbundling Using the algorithms developed in Paty, Choné, and Kramarz (2022), we simulate the transition from bundling to unbundling as the unbundling cost decreases to zero. We assume that the workers' skill profiles are distributed as a Beta(9,9) random variable, so that specialist profiles are rare in the economy, and therefore expensive under bundling. The technological intensity in skill 2 – the parameter α – is uniformly distributed on $[0, 1]$. There is no heterogeneity in workers' quality or in firms' total factor productivity. The two skills are complements ($\rho = -1$) and the returns to scale are decreasing ($\eta = .5$).

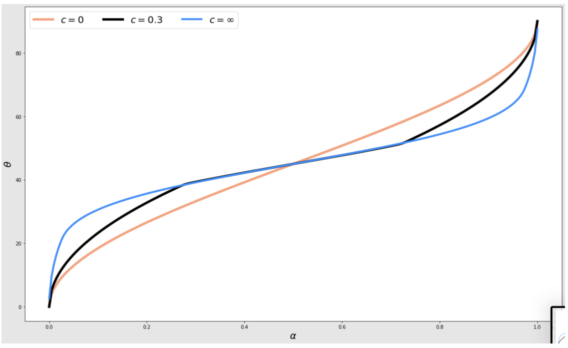
Under bundling, the law of one price does not apply. The range of implicit prices is $[.16, .92]$. The implicit price of skill 1 strictly decreases with the employing firm's type $\alpha \in [0, 1]$ while that of skill 2 increases, as shown on Figure 15. Hence the iso-wage curve is strictly concave and there is no bunching in the horizontal dimension. The sorting function $\theta(\alpha)$ is given by the blue curve shown on the four panels of Figure 16. Due to the symmetry of the workers' and firms' distributions, workers endowed with the same amount of the two skills, i.e., with skill profile $\theta = \pi/4$, are employed by firms



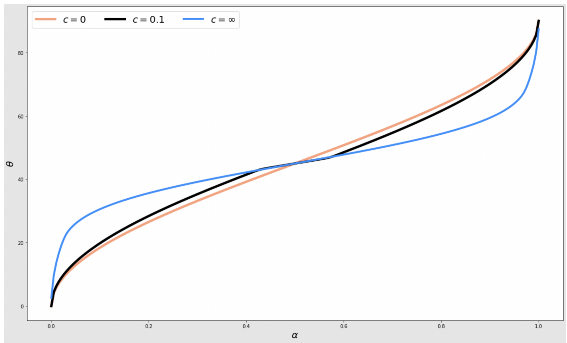
(a) Unbundling cost $c = .65$



(b) Unbundling cost $c = .5$



(c) Unbundling cost $c = .3$



(d) Unbundling cost $c = .1$

Figure 16: Sorting under bundling (Blue), costless unbundling (Orange), and costly unbundling (Black, for various levels of the unbundling cost c)

with balanced technology $\alpha = .5$. They are paid approximately .60 per unit of skill for each of the two skills.

Under costless unbundling, by definition, there is one single price per skill and by symmetry this price is common to the two skills. It is approximately equal to .64 in this example. As Proposition 10 predicts, generalist workers ($\theta = \pi/4$) are paid a higher price for their skills after unbundling, namely .64 rather than .60, implying a 6.7% gain in earnings. The sorting is represented by the orange line on Figure 16, with specialized firms (α close to zero and one) using more of their core skills than under bundling.

If each of the two skills can be unbundled at the same cost $c = .1$, then the range of implicit prices shrinks to $[.58, .68]$, with the spread $.68 - .58$ coinciding to the unbundling cost $c = .1$, as predicted by Proposition 9. The contraction of the range of implicit prices reflects the flattening of the wage schedule as skill unbundling becomes less costly. Figure 16 shows that the sorting shifts continuously from the bundling environment to the costless unbundling environment as the unbundling cost decreases.

5 The Empirical Content of Bundling and Unbundling

In this Section, we discuss the empirical content of our model, a question central to our theoretical quest. Indeed, even though we did not systematically present motivating facts for our theoretical choices, such choices are firmly grounded into an accumulation of research-based evidence on how work has been evolving through time, as discussed in the Introduction. Furthermore, we willingly centered our theory on objects that can be directly observed and measured in the data sources we already started to use (as shown in the next Section): workers' individual skills on one side, (the same) workers' employers on the other. [35]

We present in turn the main empirical consequences derived from our theory. We start by the matching patterns under bundling and the associated structure of workers' sorting to firms. We then turn to the structural wage equation that the theory delivers. We conclude this Section by discussing the empirical effects of unbundling and of endogenizing skills supply. All such empirical consequences should be understood as applying occupation by occupation (nurses, computer scientist,...) with potentially diverse skills, employed in a restricted set of firms, with a demand for skills and for the ensuing tasks that may vary from firm to firm.

We assume hereafter that $k = 2$, skill 1 comprises all *Cognitive* skills, x_C, and skill 2 comprises all *Non-Cognitive* skills, x_N, as is measured in the Swedish data. The skill profile of a worker with skill vector (x_C, x_N) is defined by $\tan\theta = x_N/x_C$.

5.1 Matching under Bundling: Within-Firm Heterogeneity in Workers' Profiles

Two sets of results imply workers' heterogeneity within firms at the matching equilibrium. The first, presented in Subsection 3.2, focuses on aggregated skills, $T(\phi)$, and its properties. The second, presented in Subsections 3.3 and 3.4, focus on the sorting equilibrium between individual workers and firms. We review them in turn and spell out their consequences in terms of workers' heterogeneity in matching workers and firms.

Proposition 3 proves the uniqueness of the firm-aggregated skill vector $T(\phi) = \int xN^d(\mathrm{d}x;\phi)$. Furthermore, by writing this skill vector as $T(\phi) = \Lambda(\phi)\tilde{X}^d(\phi)$, where $\Lambda(\phi) = |T(\phi)|$ is the total quality of the firm's employees and $\tilde{X}^d(\phi)$ is their average skill profile, we have shown that $\Lambda(\phi)$ increases with z. Hence, high-z firms, which are also high-Λ, can achieve this high total quality through a large number of employees

[35] Even though we come back to this issue a bit later in this Section, essential to note here that no equivalent data allow analysts to measure tasks performed by individual workers. Not to mention that tasks are likely to change for individuals performing similar occupations when they change employers. Hence, tasks are not a "personal attribute" as skills are.

or/and a large average quality of its bundled workers. As a consequence, our model does not imply homogenous quality of workers within a firm.

Turning now to sorting (using the two-skills – C and N – example), i.e. to the matching between individual workers and individual firms, this equilibrium is defined by two equations, (20) and (22). Equation (22) relates the matching map $\theta^d(\alpha_N)$ (with α_N, firm's preference for skill N) implicitly given by (20) and its derivative $\mathrm{d}\theta^d/\mathrm{d}\alpha_N$ 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*, assuming a CES production function (with two skills), it becomes possible to identify the distribution of αs using the link between workers' profile and α_N. This link underlines how workers with different qualities but similar skill profiles may be employed within the same firm. Hence, this sorting pattern confirms the presence of within-firm workers' heterogeneity.

All the above points hold in the absence of bunching. Importantly, bunching generates additional within-firm worker heterogeneity. We will come back to this point. But, first let us see how the above results contrast with other approaches and their consequences, as offered in the literature.

Absolute advantage: 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, recent path-breaking advances in the identification and estimation of "sorting" patterns in job-search models confirm the existence of *some* positive sorting (workers-to-firms matching, more precisely). However, 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: Deviations from perfect/absolute workers-to-firms matching exist and matter. However, even *without* assuming imperfect or asymmetric information, there are deeper reasons 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 identical skills profiles x_N/x_C but endowed with different quality levels (λ) may well work with the same employer. *Second*, when specific supply conditions prevail as demonstrated above, "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.

5.2 Wages under Bundling and Unbundling

Wage equation under bundling: Assume again that $k = 2$, skill 1 comprising all *Cognitive* skills, x_C, and skill 2 comprising all *Non-Cognitive* skills, x_N. From Proposition 2 we know that the wage schedule is homogenous of degree one. Using the function $\tilde{w}$ defined in (10), we can write the log-wage of workers with skills (x_C, x_N) as

$$\ln w(x_C, x_N) = \ln \lambda + \ln \tilde{w}(\theta), \tag{40}$$

where $\lambda = |(x_C, x_N)|$ and θ are, respectively, worker's quality and skill profile. In the absence of bunching, there is pure sorting in the horizontal dimension, recall Section 3.3, meaning that θ depends only on the technology (α_N, z) of the worker's employing firm. This property is reminiscent of the additive decomposition of the log-wage into a person and a firm effect contained in Abowd, Kramarz, and Margolis (1999).

If the production function has homothetic isoquants, the implied firm-effect is independent of z, the firm's total factor productivity. But this is not true in general. Under non-homotheticity and assuming that the marginal rate of technical substitution F_C/F_N, evaluated at $(\Lambda \cos \theta, \Lambda \sin \theta)$, increases with Λ, the equality $F_C/F_N = w_C/w_N$ implies that θ decreases with z (see Appendix A.4 for detail). Put differently, when 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. The "firm" effect now becomes linked to firm's productivity.

Recall now that from Proposition 4 the total quality of workers employed by a firm increases with the firm's total factor productivity z.[36] Hence, under non-homotheticity, the firm effect (which captures the intensity of the relative use of the two skills) and total quality of the firms' workers will be correlated. This will translate also at the individual level. Indeed, the strength of the correlation between individual worker quality and the

[36]This result is very general and does not require non-homotheticity of the production function, just the concavity of $zF(T; \alpha)$ in T.

firm effect will vary: zero under homotheticity whereas, under non-homotheticity, this individual-level correlation will be positive and small when the productive firms employ many average workers but positive and large when the productive firms employ a small number of very high-quality workers.

Identification of the wage equation: 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 immediate with no associated mobility, both in presence or in 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 within a given occupation. 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. 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.

5.3 Endogenous Skills Supply and Unbundling: Specialization and Polarization

As we discuss now, the two relaxations of the bundling constraints provide a very unified view of how bundling operates, whom it constrains the most, and who benefits most from increased opportunities. We examine them in turn.

Endogenous skills supply: Because one skill can be partly converted into another, while still being employed in a firm which buys the whole set of converted skills (hence keeping the bundling constraint) workers will tend to specialize into their better compensated skill. Hence, workers who benefit from this new opportunity are generalists who can specialize away from their less compensated skill. By contrast, specialist workers are harmed through increased competition and cannot compensate by transforming skill 1 into skill 2 since they are essentially deprived of skill 2. Endogenous skills supply therefore helps generalist workers increase their wage through increased specialization whereas it hurts specialists. The converse holds for firms. Specialist firms benefit from the relaxed supply environment.

Unbundling: Despite a very different institutional environment, unbundling has consequences that are essentially similar to those ensuing from relaxing skills supply. First, generalists – workers initially most constrained by the bundling environment – benefit from full unbundling when specialists are harmed. Second, firms employing the former are hurt when firms employing the latter benefit from this opening of markets. Furthermore, because firms can use all skills freely, they tend to increase their specialization in the direction of their comparative advantage, their preferred technology, potentially employing both salaried and contracted workers (see below when unbundling is costly).

Figure 14 presents this tendency to specialization of generalists, which is akin to a *polarization*. Generalists were constrained by bundling in their ability to sell their skills. Now, they are much less restricted in this full unbundling environment 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.

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 implicit price for skill 1 received by those employed at the firm is larger than the market price received by contracted workers for the exact same skill at the exact same firm.

5.4 From Skills to Tasks

As mentioned multiple times, skills are individual-specific. And, skills are aggregated within firms in order to produce tasks that firms will use for production. However, and until now, we have equated skills and tasks. There are many ways of aggregating workers' skills within a firm and within a skill. The most natural generalization of (1) is to consider additively separable specifications of the form:

$$T = \int g(x_C, x_N) N^d(\mathrm{d}x; \phi),$$

where g an exogenous, occupation-specific, one-to-one relationship between skills and tasks, $t = g(x)$. In Appendix A.2, we show that, *in the space of tasks*, the wage is convex and homogenous of degree one, i.e., $w_t(t_C, t_N) = w(g^{-1}(t_C, t_N))$ is convex and homogenous in (t_C, t_N). However, the link, g, between skills and tasks is unobserved, as are (t_C, t_N) and the wage function w_t. Hence, to know whether the observed wage schedule w, i.e., the wage as a function of the observed skills, inherits the properties of w_t becomes crucial.

To answer this question, we start from the simplest and most intuitive way to characterize a connection between skills and tasks and assume that each task uses each of the worker's skills in fixed quantities. An example of such a skills-to-tasks relationship is:

$$(t_C, t_N) = g(x_C, x_N) = (2/3x_C + 1/2x_N, 1/3x_C + 1/2x_N).$$

When g is linear, as in the above example, straightforwardly $w(x) = w_t(g(x))$ is also convex and homogenous of degree (as w_t). More generally, when the skills-to-tasks relationship is homogenous of degree $\gamma > 0$, as in

$$(t_C, t_N) = g(x_C, x_N) = (x_C^\gamma, x_N^\gamma),$$

the wage schedule is itself homogenous of the same degree.[37]

One may also consider aggregation technologies that are not additively separable in the workers skills. An often used aggregation scheme is CES:

$$T = \left(\left[\int x_C^\gamma N^d(\mathrm{d}x) \right]^{1/\gamma}, \left[\int x_N^\gamma N^d(\mathrm{d}x) \right]^{1/\gamma} \right),$$

with a substitution parameter $\gamma < 1$. In our leading example (7) where the production function $F(T_C, T_N)$ is itself CES, such a skills-aggregation scheme leads to a two-level nested CES. For our theoretical results to apply, we need $F(T)$ to be concave in the assignment N^d, i.e. we need the modified production function $\tilde{F}(T_C, T_N) = F(T_C^{1/\gamma}, T_N^{1/\gamma})$ to be concave in T, which obtains if $\gamma > \max(\rho, \eta)$.[38]

The number of skills needs not be equal to the number of tasks. Suppose two types of cognitive skills and two types of non-cognitive skills are used to produce two tasks according to

$$T = (T_1, T_2) = (\mathrm{CES}(C_1, N_1; \beta_1), \mathrm{CES}(C_2, N_2; \beta_2))$$

[37] Moreover, in the two-skills example above, there is a one-to-one relationship between a worker's skill profile x_C/x_N and her task profile t_C/t_N.

[38] $\tilde{F}$ is quasi-concave ($\rho/\gamma < 1$) and homogenous of degree $\eta/\gamma < 1$.

Task 1 uses skills C_1 and N_1, with β_1 representing the (potentially firm-specific) technical intensity in C_1 with an equivalent formulation holding for Task 2. The final output is then produced by combining the two tasks according to $F(T;\alpha)$. As above, the production function can be rewritten as $\tilde{F}(C_1;C_2,N_1,N_2;\alpha,\beta_1,\beta_2)$, where capital letters represent firm-aggregated quantities (for instance $C_1 = \int x_{C1}N^d(\mathrm{d}x)$) and (α,β_1,β_2) is a *firm-specific* vector of technical parameters.

A distinctive feature of the setup examined in the present paper is that firm-specific parameters interact only with firm-aggregated quantities. Using the first-order condition (14), aggregate sorting properties can be derived as in Proposition 4 noticing that $\tilde{F}_{C_i}/\tilde{F}_{N_i}$ increases with β_i and $\tilde{F}_{T_1}/\tilde{F}_{T_2}$ increases with α. This class of production functions strikingly differs from settings where a firm's set of technical characteristics interact with *individual* workers' characteristics, which pushes to individual rather than aggregate sorting.[39] The above formulation that connects skills and tasks, when compared with Haanwinckel (2020) or Teulings (2005), offer much more between-firms heterogeneity or, when compared with Eeckhout and Kircher (2018), possess a clear within-firm aggregation scheme.

6 Some Empirical Evidence

In this Section, we provide *preliminary* empirical evidence taken from Skans, Choné, and Kramarz (2022). This analysis is directly inspired by our theory. Full testing of its various components, both descriptive and structural, is left for future research as explained in our Conclusion.

6.1 The Data

6.1.1 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.[40] 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 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

[39] Choné and Kramarz (2022) considers the case where tasks are produced by interacting individual firm and worker characteristics which are then aggregated within each firm.

[40] 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.

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).

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. To be more specific, 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.[41] 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.[42]

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.

6.1.2 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.

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).

[41] 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.

[42] 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.

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.[43] 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.[44] 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.[45]

6.1.3 Descriptive Statistics

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

[43]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).

[44]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.

[45]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.

half of the sample is classified as generalists 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, "matched"). 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.

6.2 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 6.1.2). 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 6.2.3. Thus, we estimate models of the following form:

$$Y_{ij}^{\tau} = \alpha + \lambda^{C,\tau} * C_j^{-i} + \lambda^{N,\tau} * N_j^{-i} + \epsilon_{ij} \tag{41}$$

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 $\lambda^{C,C}$ and negative values on $\lambda^{C,N}$.

6.2.1 Simulating Assignment Principles

In the following paragraphs, 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				
Co-worker share of N-specialists	0.224	0.009	0.283	0.987
	(0.006)	(0.007)	(0.006)	(0.000)
Co-worker 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.

Looking at the first column in Table 2, workers appear to be systematically sorted across establishments (actual sorting). However, when comparing with the three simulated scenarios, the actual outcome is less extreme than those suggested by the 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 the theory above and not by the absolute one.

6.2.2 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 (41) 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 (2022)).

Table 3 shows the resulting estimates for workers with high total ability (full results are given in Skans, Choné, and Kramarz (2022)). As clearly appears in column (1), 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) and for high-level C-specialists in Column (3). Similar patterns also appear for mid- and low-level workers (see again Skans, Choné, and Kramarz (2022)) although horizontal sorting appears to be stronger for the high total ability workers.

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

Workers with High total ability : Dependent variable type	(1) High-ability N-Specialist	(2) High-ability Generalist	(3) High-ability 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

Notes: The results come from 9 different regressions (the full table is given in Skans, Choné, and Kramarz (2022)) 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. We report results from the 3 regressions for high total ability workers. 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.
* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

6.2.3 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 (41) 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:

$$Y_{ijt}^{\tau} = \alpha + \theta^{C,\tau} * t * C_{jt}^{-i} + \theta^{N,\tau} * t * N_{jt}^{-i} + \lambda^{C,\tau} * C_{jt}^{-i} + \lambda^{N,\tau} * N_{jt}^{-i} + D_t + X_{ijt}\beta^{\tau} + \epsilon_{ijt}^{\tau} \tag{42}$$

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.[46]

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 4 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. 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 $\theta^{C,N}$.

[46] Here, we refer to unbundling as all changes induced by endogenous labor supply as well as opening of markets, as in Section4.

The estimates are displayed in Table 4. Panel A shows the estimates for the outcome Y^C_{ijt} and panel B for Y^N_{ijt}. 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. Even more evidence is presented in Skans, Choné, and Kramarz (2022), with results fully consistent with those given just above.

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

Panel A: Being a C-specialist (dep. var.)	(1) Base	(2) Control for Occupation	(3) Co-workers in Job
C-specialists interacted with time	0.008***	0.008***	0.006***
	(0.001)	(0.001)	(0.001)
N-specialists interacted with time	-0.003***	-0.002*	-0.003***
	(0.001)	(0.001)	(0.001)
C-specialists	0.415***	0.269***	0.455***
	(0.006)	(0.008)	(0.008)
N-specialists	-0.203***	-0.122***	-0.241***
	(0.004)	(0.007)	(0.006)
N	2,317,898	1,255,003	896,931

Panel B: Being a N-specialist (dep. var.)	(1) Base	(2) Control for Occupation	(3) Coworkers in Job
N-specialists interacted with time	0.004***	0.002*	0.003**
	(0.001)	(0.001)	(0.001)
C-specialists interacted with time	-0.003***	-0.004***	-0.002*
	(0.001)	(0.001)	(0.001)
N-specialists	0.227***	0.144***	0.264***
	(0.005)	(0.008)	(0.008)
C-specialists	-0.251***	-0.147***	-0.261***
	(0.004)	(0.006)	(0.005)
N	2,317,898	1,255,003	896,931

Notes: The 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 the share of co-workers that are C/N-specialists interacted with time, normalized so that the main effects of co-workers reflect 2005. All specifications include year indicators. Col (2) also controls for occupation indicators 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). Standard errors clustered at the establishment level. Data cover 1996-2013.
* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

6.3 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.[47]

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 6.1.2). 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:

$$\ln W_{ijt} = \alpha^{C}_{C(i)} + \alpha^{N}_{N(i)} + D^{N-plant}_{jt} + \lambda^{N}_{j} * D^{N-in-N}_{ijt} + \lambda^{C}_{j} * D^{C-in-C}_{ijt} + X_{ijt}\beta \qquad (43)$$

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 (other results with a similar flavor are given in Skans, Choné, and Kramarz (2022). 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

[47] 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.

observations rather than just the half for whom we observe wages. All these results are consistent with workers being better paid when optimally matched (as in shown in (9)).

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

	(1) Base	(2) Control for Occupation	(3) Co-workers in Job
Panel A: Wages as a function of co-workers skills			
C-specialists in C-establishment	0.027***	0.009***	0.040***
	(0.003)	(0.002)	(0.003)
N-specialists in N-establishment	0.016***	0.005*	0.023***
	(0.004)	(0.003)	(0.004)
C-establishment	0.087***	0.020***	0.126***
	(0.004)	(0.003)	(0.005)
N	1,458,790	1,432,159	1,259,521
Panel B: Earnings as a function of co-workers skills			
C-specialists in C-establishment	0.036***	0.009***	0.044***
	(0.003)	(0.003)	(0.004)
N-specialists in N-establishment	0.023***	0.005*	0.026***
	(0.003)	(0.003)	(0.004)
C-establishment	0.081***	-0.002	0.108***
	(0.003)	(0.003)	(0.005)
N	2,945,409	1,432,159	1,259,521

Notes: The dependent variable is log of wages. Control variables are the indicators for each C-skill (1 to 9) and N-skill (1 to 9), indicators for being a C- or an N-specialist, as well as year indicators, an age polynomial and eight plant size indicators. 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. 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.

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

6.4 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 6.1.2). 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). Other specifications are presented in Skans, Choné, and Kramarz (2022) with similar results. 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 wage grows over time

Panel A All workers	(1) Base	(2) Control for Occupation	(3) Additional Controls
Generalist (indicator function) interacted with time	0.0012***	0.0007***	0.0007***
	(0.0002)	(0.0001)	(0.0002)
N	1,281,151	1,255,003	1,281,151
Panel B Matched sample only			
Generalist (indicator function) interacted with time	0.0031***	0.0020***	0.0018***
	(0.0006)	(0.0004)	(0.0006)
N	654,687	641,005	654,687

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. Standard errors clustered at the establishment level. Data cover 1996-2013.

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

7 Conclusion

Our paper, albeit very theoretical, has almost uniquely an applied motivation. It starts from important empirical questions on the deep structure of labor markets as it operated until recently and as it is being transformed today. Going from a world in which workers' skills are bundled to a world in which unbundling is getting easier through platforms or temp agencies, among others, provides a rich theoretical perspective on labor markets, at least we hope. For instance, the sorting patterns under bundling appear to be very different from those our theory predicts after (some) unbundling. The respective roles of firms and platforms get clearer. The structure of wages provides another striking example of contrasts between the old and the new world. Under bundling, the law of one price virtually never obtains. Log Wages tend to be the sum of a person effect and of a firm effect (due to sorting and/or productivity, under non-homotheticity of the production function). Associated to markets opening and the associated unbundling, our paper demonstrates a "flattening" of wage schedules, inducing a potential attenuation of what the literature calls, after AKM, firm effects. With the example of temp agencies in mind, Goldschmidt and Schmieder (2017) show how firms outsource tasks such as cleaning or canteens from within the firm, resulting in an elimination of the (associated) firm effect that used to be paid to those workers which accomplished these tasks in-house. Resulting also in wage losses for those workers displaced from their origin firm.

Hence, we believe that this paper has contributed to labor market theory but also to its empirics. However, we also strive to contribute to a better understanding of the productive role of workers within firms, both in their theoretical and empirical sides. Structural estimation of our model constitutes a natural way to make the two coincide.

This is our next step.[48] This new development of our research will get inspiration from recent structural contributions. More precisely, and because the sorting patterns and the matching between workers and firms are unlikely to be as clear-cut as those predicted by our theory, these contributions will guide us in our modelling of unobserved heterogeneity. These elements, observable by the workers and firms but not known to the econometrician, might account for the difference between what we predict and what is observed. For instance, in the spirit of Dupuy and Galichon (2014), workers may have idiosyncratic preferences for firms or may meet only a finite sample of them, potentially explaining some of the above difference. Alternatively or simultaneously, in the spirit of Chernozhukov, Galichon, Henry, and Pass (2021), some relevant components of the workers' skills may be observed by firms but not by the analyst, again rationalizing the distance between predictions and observations. Clearly, the methods developed there

[48] Together with Oskar Skans.

need to be adapted to our many-to-one framework where firms hire many workers and aggregate their multidimensional skills in order to deliver output using a production function framework, a mainstay of the empirical IO literature since Olley and Pakes (1996), at least.

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.

Berger, D., K. Herkenhoff, and S. Mongey (2022): "Labor Market Power," *American Economic Review*, 112(4), 1147–1193.

Boerma, J., A. Tsyvinski, and A. Zimin (2021): "Sorting with Team Formation," *Working Paper*.

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.

Chernozhukov, V., A. Galichon, M. Henry, and B. Pass (2021): "Identification of hedonic equilibrium and nonseparable simultaneous equations," *Journal of Political Economy*, 129(3), 842–870.

Chiappori, P.-A., R. McCann, and B. Pass (2016): "Multidimensional matching," *arXiv preprint arXiv:1604.05771*.

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.

Choné, P., N. Gozlan, and F. Kramarz (2022): "Weak optimal transport with unnormalized kernels," arXiv:2203.16227.

Choné, P., and F. Kramarz (2022): "Transporting superstars," 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.

Dupuy, A., and A. Galichon (2014): "Personality traits and the marriage market," *Journal of Political Economy*, 122(6), 1271–1319.

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.

Galichon, A., and B. Salanié (forthcoming): "Cupid's invisible hand: Social surplus and identification in matching models," *Review of Economic Studies*.

Goldschmidt, D., and J. F. Schmieder (2017): "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure," *Quarterly Journal of Economics*, 132(3), 1165–1217.

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.

Haanwinckel, D. (2020): "Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution," *Working paper*.

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*, 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.

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*.

Lise, J., and F. Postel-Vinay (2020): “Multidimensional skills, sorting, and human capital accumulation,” *American Economic Review*, 110(8), 2328–76.

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.

Ohnsorge, F., and D. Trefler (2007): “Sorting It Out: International Trade with Heterogeneous Workers,” *The Journal of Political Economy*, 115(5), 868–892.

Olley, G. S., and A. Pakes (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64(6), 1263–1297.

Paty, F.-P., P. Choné, and F. Kramarz (2022): “Algorithms for Weak Optimal Transport with an Application to Economics,” *arXiv preprint arXiv:2205.09825*.

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.

Roy, A. D. (1951): “Some thoughts on the distribution of earnings,” *Oxford economic papers*, 3(2), 135–146.

Sato, R. (1977): “Homothetic and non-homothetic CES production functions,” *The American Economic Review*, 67(4), 559–569.

Skans, O. N., P. Choné, and F. Kramarz (2022): “The Empirics of Labor Market Sorting When Workers Have Bundles of Skills and Jobs Are Heterogeneous,” IFAU Working Paper 2022:6.

Teulings, C. N. (2005): “Comparative advantage, relative wages, and the accumulation of human capital,” *Journal of Political Economy*, 113(2), 425–461.

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 Proof of Proposition 1

Choné, Gozlan, and Kramarz (2022) introduce a dual version of the primal problem

$$I^* \stackrel{\mathrm{d}}{\equiv} \inf_{w \in \mathcal{C}_b(\mathcal{X})} \int \Pi(\phi; w) H^f(\mathrm{d}\phi) + \int w(x) H^w(\mathrm{d}x), \tag{A.1}$$

where $\mathcal{C}_b(\mathcal{X})$ denotes the set of bounded continuous function on $\mathcal{X}$. Theorem 3.3 of the above paper establishes the duality formula $Y^* = I^*$ as well as the existence of solutions to the primal and dual problems (5) and (A.1). On the one hand, there exists a family of positive measures $N^d(\mathrm{d}x, \phi)$ satisfying $N^d H^f = H^w$ that achieves the upper bound in (5), hence the existence of an optimal assignment of workers to firms. On the other hand, there exists a bounded continuous w that achieves the lower bound in (A.1).

First, consider an equilibrium (w, N^d). Let us denote by T the firm-aggregated skill vector corresponding to the assignment N^d, i.e., $T(\phi) = \int x N^d(\mathrm{d}x; \phi)$. Using the market clearing condition $N^d H^f = H^w$, we have

$$\begin{aligned} I^* &\leq \int \Pi(\phi; w) H^f(\mathrm{d}\phi) + \int w(x) H^w(\mathrm{d}x) \\ &= \int F(T(\phi); \phi) H^f(\mathrm{d}\phi) - \iint w(x) N^d(\mathrm{d}x; \phi) H^f(\mathrm{d}\phi) + \int w(x) H^w(\mathrm{d}x) \\ &= \int F(T(\phi); \phi) H^f(\mathrm{d}\phi) \leq Y^*. \end{aligned}$$

Because $Y^* = I^*$, the last inequality is an equality, implying that the equilibrium assignment N^d is optimal.

Conversely, consider an optimal market clearing assignment N^d. As above, we denote by $T(\phi)$ the corresponding firm-aggregated skill vector. Then, for any dual optimizer w, we have by definition of the profit function

$$F(T(\phi); \phi) - \int w(x) N^d(\mathrm{d}x; \phi) \leq \Pi(\phi; w). \tag{A.2}$$

Integrating with respect to $H^f(\mathrm{d}\phi)$ and using $N^d H^f = H^w$ yields

$$Y^* = \int F(T(\phi); \phi) H^f(\mathrm{d}\phi) \leq \int \Pi(\phi; w) H^f(\mathrm{d}\phi) + \int w(x) H^w(\mathrm{d}x) = I^*. \tag{A.3}$$

The equality $Y^* = I^*$ shows that we must have equality in (A.2) for H^f-almost every $\phi \in \varPhi$, meaning that the optimal market clearing assignment N^d is decentralized by the wage schedule w.

A.2 Proof of Proposition 2

The firms' problem (6) can be broken down into two subproblems that consist respectively in finding the firm-aggregated skill vector T and in achieving that aggregate vector in the most economical way. Formally, the former problem is given by

$$\Pi(\phi; w) = \max_{T \in \mathcal{Z}} F(T; \phi) - \bar{w}(T), \tag{A.4}$$

where $\mathcal{Z}$ is the conical hull of $\mathcal{X}$: $\mathcal{Z} = \left\{ \sum_{j=1}^{k} a_j x_j, a_1, \ldots a_n \in \mathbb{R}_+, x_1, \ldots, x_n \in \mathcal{X} \right\}$. The latter problem (minimizing the wage bill at given firm-aggregated skill) is given by

$$\bar{w}(T) = \inf \left\{ \int w(x) N^d(\mathrm{d}x) : N^d \in \mathcal{M}(\mathcal{X}), \int x N^d(\mathrm{d}x) = T \right\}. \tag{A.5}$$

It is easy to check that the function $\bar{w}$ defined in (A.5) is convex and homogenous of degree one. For any $x \in \mathcal{X}$, we can take the $N^d(\mathrm{d}x)$ as the mass point at x, thus showing that $\bar{w}(x) \leq w(x)$. The map $\bar{w} : \mathcal{Z} \to \mathbb{R}_+$ is therefore the greatest convex and homogenous function such that $\bar{w} \leq w$ on $\mathcal{X}$.

By construction of $\bar{w}$, we have: $\Pi(\phi; w) = \Pi(\phi; \bar{w})$. Moreover, because $\bar{w} \leq w$, we have: $\int \bar{w}(x) H^w(\mathrm{d}x) \leq \int w(x) H^w(\mathrm{d}x)$. It follows that if w is a dual optimizer, i.e., a solution of Problem (A.1), so is $\bar{w}$. Using $\bar{w}$ instead of w in (A.2) and (A.3) shows that the optimal market clearing assignment N^d is decentralized by the convex and positively homogenous wage schedule $\bar{w}$. $\square$

Lemma A.1. *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 $\partial\mathcal{C}$.*

Relation between skills and tasks We present the change of variables $t = g(x)$ mentioned in Subsection 5.4. We define the probability distribution over tasks: $\tilde{H}^w(\mathrm{d}t) = g_\# H^w(\mathrm{d}x)$. To any assignment $N^d(\mathrm{d}x; \phi)$, we associate the assignment in the task space $M^d(\mathrm{d}t; \phi) = g_\# N^d(\mathrm{d}x; \phi)$. Because g is one-to-one, the market clearing conditions $N^d H^f = H^w$ and $M^d H^f = \tilde{H}^w$ are equivalent. The primal problem (5) that defines the optimal output under bundling can be rewritten as

$$Y^* = \sup_{M^d | M^d H^f = \tilde{H}^w} \int F\left(\int t M^d(\mathrm{d}t;\phi)\right) H^f(\mathrm{d}\phi).$$

Starting from any wage schedule $w(x)$, we define the price of task t as $p(t) = w(g^{-1}(t))$ and rewrite the firms' profit (6) as

$$\tilde{\Pi}(\phi;p) = \max_{M^d(\mathrm{d}t;\phi)} F\left(\int t M^d(\mathrm{d}t;\phi)\right) - \int p(t) M^d(\mathrm{d}t;\phi).$$

We can also the dual problem (A.1) as

$$I^* = \inf_{p \in \mathcal{C}_b(g(\mathcal{X}))} \int \tilde{\Pi}(\phi;p) H^f(\mathrm{d}\phi) + \int p(t) \tilde{H}^w(\mathrm{d}t).$$

We can thus apply the Fundamental Theorems in the tasks space $g(\mathcal{X})$ equipped with the probability measure $\tilde{H}^w(\mathrm{d}t)$ and the firm space Φ with the probability $H^f(\mathrm{d}\phi)$.

A.3 Proof of Proposition 3

Consider two optimal market clearing assignments of workers to firms, N_1^d and N_2^d. Let $T_i = \int x N_i^d(\mathrm{d}x;\phi)$, $i = 1,2$ denote the corresponding firm-aggregated skill vectors. We have seen in the proof of Proposition 1 that there exists a dual optimizer w, i.e., a solution to Problem (A.1), that is convex and homogenous of degree one. We know that T_1 and T_2 are solutions to Problem (11), recall (A.4) above. Because F is strictly concave and w is convex, the problem is strictly concave, which yields $T_1 = T_2$. □

CES technology and twist conditions For the CES production function (7) and $\phi = (z, \alpha_1, \ldots, \alpha_{k-1})$, we have

$$\nabla_\phi F(T;\phi) = (Y_0, Y_1, \ldots, Y_{k-1})',$$

with

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

for $j = 1, \ldots, k-1$. It follows that $T_j^\rho = (\rho/z)\,(\eta Y_0)^{\rho/\eta - 1} Y_j$ for $j = 1, \ldots, k$. The map $T \to \nabla_\phi F(T;\phi)$ is therefore invertible.

Proof of Corollary 2 From (12), the marginal rate of technical substitution (MRTS) equals the ratio of implicit prices across skills:

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

where $T = \Lambda\tilde{X}^d$, $\Lambda > 0$. Because the wage schedule is positively homogenous, the wage isolines are homothetic, and the ratios w_j/w_k depend only on $\tilde{X}^d$. If the production functions have homothetic isoquants, the same is true for the MRTS F_j/F_k.

Proof of Corollary 3 From Corollary 2, we know that the average skill profile $\tilde{X}^d$ does not depend on z. The total quality of a firm ϕ's employees, $\Lambda(\phi)$, is determined by maximizing its profit:

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

Using that F is homogenous of degree $\eta < 1$, we find that the total quality of workers employed by firm $\phi = (\alpha, z)$:

$$\Lambda^d(\alpha,z) = \left[\frac{\eta\, z\, F(\tilde{X}^d(\alpha);\alpha)}{w(\tilde{X}^d(\alpha))}\right]^{\frac{1}{1-\eta}}. \tag{A.6}$$

The firm's aggregate skill is $T(\phi) = \Lambda^d(\alpha,z)\tilde{X}^d(\alpha)$. Using that F is homogenous of degree η, we can write its wage bill as

$$w(T(\phi)) = \Lambda^d(\alpha,z)w(\tilde{X}^d(\alpha)) = \left[\eta\, z\, F\left(\frac{\tilde{X}^d(\alpha)}{w(\tilde{X}^d(\alpha))};\alpha\right)\right]^{\frac{1}{1-\eta}}. \tag{A.7}$$

The firm's profit is

$$\begin{aligned}
\Pi(\phi;w) &= (1-\eta)\,(z\eta^{\eta})^{\frac{1}{1-\eta}} \left[F\left(\frac{\tilde{X}^d(\alpha)}{w(\tilde{X}^d(\alpha))};\alpha\right)\right]^{\frac{1}{1-\eta}} \\
&= (1-\eta)\,(z\eta^{\eta})^{\frac{1}{1-\eta}}\, w(\tilde{X}^d(\alpha)) \left[\frac{F\left(\tilde{X}^d(\alpha);\alpha\right)}{w(\tilde{X}^d(\alpha))}\right]^{\frac{1}{1-\eta}}.
\end{aligned} \tag{A.8}$$

All the above quantities depend on the TFP parameter z through $z^{1/(1-\eta)}$.

A.4 Proof of Proposition 4

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

$$K_1(\theta, \Lambda) \stackrel{\mathrm{d}}{=} zF_1(\Lambda \cos\theta, \Lambda \sin\theta; \alpha) - w_1(\theta) \quad = \quad 0 \tag{A.9}$$
$$K_2(\theta, \Lambda) \stackrel{\mathrm{d}}{=} zF_2(\Lambda \cos\theta, \Lambda \sin\theta; \alpha) - w_2(\theta) \quad = \quad 0. \tag{A.10}$$

where K_1 and K_2 are the first derivatives of the firm's objective $F(T; \phi) - w(T)$. Differentiating the first-order conditions (A.9) and (A.10) and inverting the Jacobian of K yields

$$\begin{pmatrix} \frac{\partial \theta}{\partial \alpha} & \frac{\partial \theta}{\partial z} \\ \frac{\partial \Lambda}{\partial \alpha} & \frac{\partial \Lambda}{\partial z} \end{pmatrix} = -\frac{1}{d} \begin{pmatrix} z\frac{\partial F_2}{\partial \Lambda} & -z\frac{\partial F_1}{\partial \Lambda} \\ -\left(z\frac{\partial F_2}{\partial \theta} - w_2'\right) & z\frac{\partial F_1}{\partial \theta} - w_1' \end{pmatrix} \begin{pmatrix} z\frac{\partial F_1}{\partial \alpha} & F_1 \\ z\frac{\partial F_2}{\partial \alpha} & F_2 \end{pmatrix}, \tag{A.11}$$

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

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

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

To prove the first part of the proposition, we compute the derivative of total quality with respect to total factor productivity

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

Consider the above bracketed terms. 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.

To prove the second part – the PAM property –, we need to show that the determinant of the sorting matrix is positive and that θ increases with α. Regarding the former point, the determinant of the sorting matrix at left-hand side of (A.11) is positive because by concavity of the firm problem and the Assumption that F_2/F_1 increases with α the two matrices at the right-hand side have a negative determinant. Regarding the latter point, the derivative of the skill profile with respect to technological intensity is

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

Hence θ increases with α of and only if

$$\frac{\partial F_1}{\partial \alpha}\frac{\partial F_2}{\partial \Lambda} - \frac{\partial F_2}{\partial \alpha}\frac{\partial F_1}{\partial \Lambda} \geq 0. \tag{A.12}$$

It follows from the above analysis that (A.12), together with F_2/F_1 increasing in α, is a sufficient condition for PAM. Condition (A.12) holds in particular if production isoquants are homothetic. Indeed, we have in this case that $F_1 \partial F_2/\partial\Lambda = F_2 \partial F_1/\partial\Lambda$ and hence $(\partial F_1/\partial\Lambda, \partial F_2/\partial\Lambda) = -\kappa(F_1, F_2)$ for some constant $\kappa > 0$, which, together with F_2/F_1 increasing in α, guarantees that (A.12) holds.

Non-homothetic isoquants We now provide detail about the sorting pattern when production isoquants are non-homothetic, see the discussion in Section 5.2. From (A.11), we have

$$\frac{\partial \theta}{\partial z} = -(z/d)\left\{F_1\frac{\partial F_2}{\partial \Lambda} - F_2\frac{\partial F_1}{\partial \Lambda}\right\}.$$

where $d < 0$. It follows θ is independent of z when the production isoquants are homothetic and decreases with z if $\partial(F_1/F_2)/\partial\Lambda > 0$. Adapting notations $F_1 = F_C$ and $F_2 = F_N$ yields the results announced in Section 5.2. The latter condition holds for instance for the CES function modified in the spirit of Sato (1977):

$$zF(T;\alpha) = (z/\eta)\ \left[\alpha_C(T_C + \bar{T}_C)^\rho + \alpha_N T_N^\rho\right]^{\eta/\rho}, \tag{A.13}$$

where $\bar{T}_C$ is a positive constant. Indeed here

$$\frac{F_C}{F_N} = \frac{\alpha_C}{\alpha_N}\left[\frac{N}{C + \bar{T}_C}\right]^{1-\rho}$$

and hence F_C/F_N evaluated at $(\Lambda\cos\theta, \Lambda\sin\theta)$ increases with Λ.

A.5 Proof of Proposition 5

Let w be an equilibrium wage schedule that is convex and homogenous of degree one. We have, for any firm type ϕ

$$\frac{w\left(\int xN^d(\mathrm{d}x;\phi)\right)}{\int w(x)N^d(\mathrm{d}x;\phi)} = w\left(\frac{\int [x/w(x)]w(x)N^d(\mathrm{d}x;\phi)}{\int w(x)N^d(\mathrm{d}x;\phi)}\right) \leq \frac{\int w(x)N^d(\mathrm{d}x;\phi)}{\int w(x)N^d(\mathrm{d}x;\phi)} = 1. \tag{A.14}$$

When the iso-wage surface $\partial_+\mathcal{W}$ is strictly concave, the equality in (A.14) imposes that $x/w(x)$ is constant for N^d-almost every x, i.e., that all the workers employed by firms of type ϕ have the same skill profile.

Recall that 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 $\mathcal{Y}$

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

In the particular case of the operator W, we have

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

for any test function h. It follows that

$$\begin{aligned}
< W_{\#}T_{\#}H^f, h > &= \int_{\phi} h\left(\frac{T(\phi)}{w(T(\phi))}\right) w(T(\phi))H^f(\mathrm{d}\phi) \\
&= \int_{\phi} h\left(\frac{T(\phi)}{w(T(\phi))}\right) \int_x w(x)\mathrm{d}N^d(x;\phi)H^f(\mathrm{d}\phi) && \text{(A.15)} \\
&= \iint h\left(\frac{x}{w(x)}\right) w(x)\mathrm{d}N^d(x;\phi)H^f(\mathrm{d}\phi) && \text{(A.16)} \\
&= \int_x h\left(\frac{x}{w(x)}\right) w(x)H^w(\mathrm{d}x) && \text{(A.17)} \\
&= < W_{\#}H^w, h > .
\end{aligned}$$

Equation (A.15) follows from the equality in (A.14). Equation (A.16) uses that $x/w(x) = T(\phi)/w(T(\phi))$ for all x in the support of $N^d(\mathrm{d}x;\phi)$, i.e., for all x proportional to $\tilde{X}^d(\alpha)$. Equation (A.17) uses the equilibrium condition (2).

A.6 Proof of Proposition 6

For any convex test function h, we have, using the equality in (A.14) for w and Jensen inequality for h

$$h\left(\frac{T(\phi)}{w(T(\phi))}\right) = h\left(\frac{\int [x/w(x)]w(x)N^d(\mathrm{d}x;\phi)}{w(T(\phi))}\right) \leq \frac{1}{w(T(\phi))}\int h\left(\frac{x}{w(x)}\right) w(x)N^d(\mathrm{d}x;\phi),$$

which yields

$$\begin{aligned}
< W_\# T_\# H^f, h > &= \int_\phi h\left(\frac{T(\phi)}{w(T(\phi))}\right) w(T(\phi)) H^f(\mathrm{d}\phi) \\
&\leq \iint_\phi h\left(\frac{x}{w(x)}\right) w(x)\mathrm{d}N^d(x;\phi) H^f(\mathrm{d}\phi) \\
&= \int_x h\left(\frac{x}{w(x)}\right) w(x) H^w(\mathrm{d}x) \\
&= < W_\# H^w, h > .
\end{aligned}$$

A.7 Proof of Proposition 7

Consider a market clearing assignment N^d such that $T(\phi)$ is the firm-aggregated skill vector $T(\phi) = \int xN^d(\mathrm{d}x;\phi)$. Because any convex and positively 1-homogenous function is sub-additive, we have

$$\begin{aligned}
\int h(x) T_\# H^f(\mathrm{d}x) &= \int h(T(\phi)) H^f(\mathrm{d}\phi) \\
&= \int h\left(\int xN^d(\mathrm{d}x;\phi)\right) H^f(\mathrm{d}\phi) \\
&\leq \iint h(x) N^d(\mathrm{d}x;\phi) H^f(\mathrm{d}\phi) = \int h(x) H^w(\mathrm{d}x),
\end{aligned}$$

which proves $T_\# H^f \leq_{phc} H^w$.

The converse property follows from the new variant of Strassen Theorem established by Choné, Gozlan, and Kramarz (2022). Theorem 4.2 in their paper establishes that for any distribution γ more "generalist" than H^w in the sense that $\gamma \leq_{phc} H^w$, there exists a market clearing assignment $N^d(\mathrm{d}x;\phi)$ such that $N^d\gamma = H^w$ and $y = \int xN^d(\mathrm{d}x;\phi)$ for γ-almost every y. Applying this result to the distribution $\gamma = T_\# H^f$ yields the desired property. The equality (26) follows from Theorem 4.5 of Choné, Gozlan, and Kramarz (2022).

A.8 Proof of Proposition 8

We prove that we can restrict attention to price schedules satisfying (32). We can write the dual version of Problem (27) with endogenous supply of skills

$$I^* = \inf_{w \in \mathcal{C}_b(\mathcal{X})} \int \Pi(\phi; w) H^f(\mathrm{d}\phi) + \int U(x; w) H^w(\mathrm{d}x), \tag{A.18}$$

The function $U(x_1 - x, x_2 + \tau x; w)$ is non-increasing on $[0, x_1]$ because $g(x_1 - x, x_2 + \tau x) \subset g(x_1, x_2)$. It follows that $\tau U_2 - U_1 \leq 0$ or $U_1/U_2 \geq \tau$. Replacing $w(x)$ with

$U(x;w)$ does not alter the workers' utilities and decreases the firms' profit because $U(x;w) \geq w(x)$. Hence if w is solution to the dual problem, so is $U(x;w)$. It follows that without loss of generality we may restrict attention to wage schedules that satisfy (32) and the dual version of the problem can be rewritten as

$$I^* = \inf_{w \in \mathcal{C}_b(\mathcal{X}) \,|\, w_1 \geq \tau w_2} \int \Pi(\phi;w) H^f(\mathrm{d}\phi) + \int w(x) H^w(\mathrm{d}x). \tag{A.19}$$

A.9 Proof of Proposition 9

From the first-order conditions of the firms' problem (36), we have

$$F_i(T(\phi);\phi) = w_i(T(\phi)) \leq p_i + c_i^f$$

for any technology ϕ, with equality if firms of type ϕ purchase a positive amount of task $i \in \{1,2\}$, i.e., if $m_i^d > 0$. From the first-order conditions of the workers' problem (37), we have

$$p_i - c_i^w - w_i(x - m_s) \leq 0$$

for any skill vector x, with equality if workers of type x sell a positive amount of task i, i.e., if $m_i^s > 0$. It follows that

$$\max_x w_i \leq p_i + c_i^f \leq \min_x w_i + c_i^w + c_i^f = \min_x w_i + c_i,$$

which yields (38) and confirms that equality holds when a positive amount of task i is traded.

A.10 Proof of Proposition 10

In this Subsection, we prove that $r(\hat{\alpha}) > 1$, where r is defined by (39). Recall that the superscripts b and u refer to the polar cases of bundling ($c_i^b = \infty$ for $i \in \{1,2\}$) and costless unbundling ($c_i^u = 0$ for $i \in \{1,2\}$). We denote by $\tilde{X}^b(\alpha) = (\cos\theta^b(\alpha), \sin\theta^b(\alpha))$ the average skill profile of the workers hired by firms with technological parameter α under bundling.

From (10) and the envelope theorem, we have: $w'(\theta) = -w_1(\theta)\sin\theta + w_2(\theta)\cos\theta$, which yields the derivatives

$$r'(\alpha) = (\theta^b)'(\alpha)\, \frac{p_2^u w_1^b(\tilde{X}^b(\alpha)) - p_1^u w_2^b(\tilde{X}^b(\alpha))}{w^b(\tilde{X}^b(\alpha))^2} \tag{A.20}$$

Because by assumption F_2/F_1 increases in α, the matching map θ^b is increasing and therefore the implicit prices $w_1^d(\tilde{X}^b(\alpha))$ and $w_2^d(\tilde{X}^b(\alpha))$ respectively decrease and increase with α for $d = b$ and $d = u$. The numerator of the above fraction is decreasing in α. It is zero for firms $\hat{\alpha}$ that have the same average skill profile $\tilde{X}^b(\hat{\alpha}) = \tilde{X}^u(\hat{\alpha})$ under bundling and unbundling. The function $r(\alpha)$ is quasi-concave and achieves its maximum at $\hat{\alpha}$ and local minima at $\alpha = 0$ and $\alpha = 1$.

According to Lemma A.1 below, some weighted average of $r(\alpha)^{1/(1-\eta)}$ is larger than one. Given the shape of $r(\alpha)$, the former property guarantees that the workers of skill profile $\theta^b(\hat{\alpha}) = \theta^u(\hat{\alpha})$ are indeed strictly better off under unbundling, $r(\hat{\alpha}) > 1$.

Lemma A.1. *Let r be the ratio defined by (39) There exists a nonnegative functions $\mu(\alpha)$ such that $\int_0^1 \mu(\alpha)\mathrm{d}\alpha = 1$ and*

$$\int \mu(\alpha) r(\alpha)^{1/(1-\eta)} \mathrm{d}\alpha \geq 1, \tag{A.21}$$

with equality if and only if the bundling and unbundling equilibria are the same.

Proof. The proof proceeds by computing the quantity $\int w^u(x) H^w(\mathrm{d}x)$ in two different ways, where $w^u(x) = p_1^u x + p_2^u x_2$ is the wage schedule under costless unbundling.

First, we interpret this quantity as the sum of the wage bills of all firms under unbundling. In this situation, the firm's problem (36) writes $\max_T F(T;\phi) - w^u(T)$ which is the same problem as (11). We denote by $T^u(\phi)$ the solution of that problem and set $\tilde{X}^u(\alpha) = (\cos\theta^u(\alpha), \sin\theta^u(\alpha)) = T^u(\phi)/|T^u(\phi)|$. Using the expression (A.7) for the wage bill, we get

$$\int w^u(x) H^w(\mathrm{d}x) = \int w^u(\tilde{X}^u(\alpha)) \left[\frac{F(\tilde{X}^u(\alpha);\alpha)}{w^u(\tilde{X}^u(\alpha))}\right]^{1/(1-\eta)} Z^f(\alpha) H^f(\mathrm{d}\alpha), \tag{A.22}$$

where $Z^f(\alpha)$ is defined below (21).

Second, we use the linearity of w^u and the equilibrium condition $N^d H^f = H^w$ to get

$$\int w^u(T^b(\phi)) H^f(\mathrm{d}\phi) = \int w^u\left(\int x N^d(\mathrm{d}x;\phi)\right) H^f(\mathrm{d}\phi) = \int w^u(x) H^w(\mathrm{d}x). \tag{A.23}$$

Using the size of the firms under bundling given by (A.6), we then compute

$$\begin{aligned}
\int w^u(x)H^w(\mathrm{d}x) &= \int w^u(T^b(\phi))H^f(\mathrm{d}\phi) \\
&= \int w^u(\tilde{X}^b(\alpha)) \left[\frac{F(\tilde{X}^b(\alpha);\alpha)}{w^b(\tilde{X}^b(\alpha))}\right]^{\frac{1}{1-\eta}} Z^f(\alpha)H^f(\mathrm{d}\alpha) \\
&= \int w^u(\tilde{X}^b(\alpha)) \left[\frac{F(\tilde{X}^b(\alpha);\alpha)}{w^u(\tilde{X}^b(\alpha))}\right]^{\frac{1}{1-\eta}} r(\alpha)^{\frac{1}{1-\eta}} Z^f(\alpha)H^f(\mathrm{d}\alpha) \\
&\le \int w^u(\tilde{X}^u(\alpha)) \left[\frac{F(\tilde{X}^u(\alpha);\alpha)}{w^u(\tilde{X}^u(\alpha))}\right]^{\frac{1}{1-\eta}} r(\alpha)^{\frac{1}{1-\eta}} Z^f(\alpha)H^f(\mathrm{d}\alpha),
\end{aligned}$$

with the above inequality coming from the profit optimization of firm α under unbundling (recall the firm profit is given by (A.8)). Combining this inequality with (A.22) and (A.23) yields (A.21). The equality occurs if and only if the bundling and unbundling equilibria are the same, i.e., if and only if $\tilde{X}^b = \tilde{X}^u$. □

A.11 Connection to optimal transport theory

In this section, we explain how our setup is related to optimal transport theory.

Weak optimal transport (WOT) Given two probability measures μ and ν, and a cost function $c(\phi, m)$ that is convex in m, Gozlan, Roberto, Samson, and Tetali (2017) consider the problem of minimizing

$$\inf_{\pi\in\Pi(\mu,\nu)} \int c(\phi, p^\phi)\mathrm{d}\mu(\phi), \tag{A.24}$$

where $\Pi(\mu, \nu)$ is the set of all couplings π of μ and ν (i.e., the set of probability measures over $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν) and p^ϕ is the (μ-almost surely unique) probability kernel such that

$$\mathrm{d}\pi(x, \phi) = \mathrm{d}p^\phi(x)\,\mathrm{d}\mu(\phi). \tag{A.25}$$

Gozlan, Roberto, Samson, and Tetali (2017) prove existence and duality results for Problem (A.24) 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 (5), has the same form as (A.24), 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 (A.24) is equivalent to maximizing (5) 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.

Unnormalized kernels and endogenous firms' sizes As mentioned in Section 2, the framework developed in the present article has an important difference with the WOT problem described above. Specifically, we do not impose that the workers-to-firms assignments, $N^d(\mathrm{d}x;\phi)$, are *probability* measures, as is required in the kernel disintegration (A.25). Accordingly, Choné, Gozlan, and Kramarz (2022) relax the assumption that π_x in (A.24) is a *probability* measure. Denoting by $\mathcal{M}(\mathcal{Y})$ the set of positive measures over $\mathcal{Y}$, they introduce the weak optimal transport problem with unnormalized kernel (WOTUK) as

$$\mathrm{WOTUK}(\mu,\nu) \overset{\mathrm{d}}{=} \sup_{\substack{q\in\mathcal{M}(\mathcal{Y})^{\mathcal{X}} \\ \int q_x\, d\mu(x)=\nu}} \int_{\mathcal{X}} \mathcal{F}(x,q_x)\, d\mu(x), \tag{A.26}$$

where $\mathcal{F}:\mathcal{X}\times\mathcal{M}(\mathcal{Y})\to\mathbb{R}$. The constraint $\int q_x\, d\mu(x)=\nu$ expresses that the unnormalized kernel (q_x) transports μ onto ν. They connect the WOTUK problem (A.26) to a WOT problem as follows. Letting

$$\Pi(\ll\mu,\nu) \overset{\mathrm{d}}{=} \{P\in\Pi(\eta,\nu)\,,\, \eta\in\mathcal{P}(\mathcal{X}), \eta\ll\mu\},$$

denote the set of probability measure over $\mathcal{X}$ that are absolutely continuous with respect to μ, they show that

$$\mathrm{WOTUK}(\mu,\nu) = \sup_{\Pi(\ll\mu,\nu)} \sup_{\pi\in\Pi(\eta,\nu)} \int_{\mathcal{X}} \mathcal{F}\left(x,\frac{d\eta}{d\mu}(x)\pi_x\right) d\mu(x) \tag{A.27}$$

where $\pi_x\in\mathcal{P}(\mathcal{Y})$ is the unique disintegration of π with respect to η, *i.e.* such that $d\pi(x,y)=d\eta(x)d\pi_x(y)$. At given η, we thus get back the WOT problem. Instead of constraining the first marginal of π to be μ, the WOTUK problem only imposes that the first marginal is absolutely continuous with respect to μ. They show that the density of η with respect to μ is nothing else than the mass of q_x, i.e., $d\eta/d\mu = q_x(\mathcal{Y})$. Choné, Gozlan, and Kramarz (2022) prove the existence of a solution of the primal problem and a Kantorovich type duality formula that yields (A.1).

In the economic setting of this paper, $q_x(\mathcal{Y})$ represents the number of employees (i.e., the size) of firms with type x, which we have denoted by $N(x)$, so we have $N(x) \overset{\mathrm{d}}{=} \frac{d\eta}{d\mu}(x)\in\mathbb{R}_+$. Allowing q_x to be an unnormalized positive measure instead of a probability measure avoids having to assume that all firms have the same size.

Conical WOTUK problems The specification studied in the present paper corresponds to a special class of WOTUK problems, which Choné, Gozlan, and Kramarz (2022) call conical WOTUK problem. It corresponds to the case where

$$\mathcal{F}(x,p) = F\left(x, \int_{\mathcal{Y}} y\, dp(y)\right)$$

for some $F : \mathcal{X} \times \mathrm{cone}(\mathcal{Y}) \to \mathbb{R}$, where the conical hull of $\mathcal{Y}$ is given by

$$\mathrm{cone}(\mathcal{Y}) \stackrel{\mathrm{d}}{=} \left\{ \sum_{i=1}^{n} \lambda_i y_i \,,\, \lambda_1, \ldots, \lambda_n \in \mathbb{R}_+, y_1, \ldots, y_n \in \mathcal{Y}, n \geq 1 \right\}.$$

Choné, Gozlan, and Kramarz (2022) establish the existence of solutions for the dual problem, which guarantee the existence of a competitive equilibria in our setting where a firm's output depends on the *conical* combination of its employees' types, $\int y\, dq_x(y)$. The combination is said to be "conical" because the mass of q_x is not necessarily equal to one. In other words, the aggregate skill of the workers hired by a firm is not their average skills as in the WOT setting, but their average skills *scaled by the positive factor* $q_x(\mathcal{Y})$ that represents the number of employees.

B A Dixit-Stiglitz Environment

In the environment presented in main text, the price of the final good is exogenous and normalized to one, and quantities (output, labor demand, etc.) are determined by decreasing returns to scale. We now present a different framework where firms operate under constant returns to scale and quantities are set by monopolistic competition à la Dixit-Stiglitz. This framework is used for instance by Costinot and Vogel (2010) where they consider one-dimensional skills.

For simplicity of exposition, we assume in the following that skills are two-dimensional. Firms indexed by (α, z) produce a differentiated good under constant returns to scale. The production function takes the form $y(\alpha, z) = zF(X_1, X_2; \alpha)$, where F is homogenous of degree one. A representative consumer has income I and preferences over baskets $\mathbf{y} = (y(\alpha, z))$ given by

$$U(\mathbf{y}) = \left(\int y(\alpha, z)^{\frac{\sigma-1}{\sigma}} H^f(\mathrm{d}\alpha, \mathrm{d}z) \right)^{\frac{\sigma}{\sigma-1}},$$

with $\sigma > 1$. Let $p(\alpha, z)$ denote the price of good (α, z). The Marshallian demand is given by

$$y(\alpha, z) = I \frac{p(\alpha, z)^{-\sigma}}{\int [p(\alpha, z)]^{1-\sigma} H^f(\mathrm{d}\alpha, \mathrm{d}z)}. \tag{B.1}$$

As in the text, we parameterize aggregate skill vectors as $T = (X_1, X_2) = \Lambda(\cos\theta, \sin\theta)$, where θ is the firm's aggregate skill profile and Λ is the aggregate quality of its employees. The output can thus be rewritten as

$$y(\alpha, z) = z\Lambda F(\cos\theta, \sin\theta; \alpha).$$

The wage schedule is denoted $w(x_1, x_2)$. By sub-additivity of the wage schedule, we know that the wage bill of a firm using aggregate skill T is

$$w(T) = \min_N(\mathrm{d}x) \left\{ \int w(x)N(\mathrm{d}x) \mid \int xN(\mathrm{d}x) = T \right\}.$$

There is monopolistic competition on the downstream market and firm (α, z) chooses its aggregate skill vector $T = \int xN(\mathrm{d}x)$ to maximize its profit

$$\begin{aligned} p(\alpha, z)y - w(T) &= y\left[p(\alpha, z) - \frac{w(T)}{y}\right] \\ &= y\left[p(\alpha, z) - \frac{w(\cos\theta, \sin\theta)}{z\ F(\cos\theta; \sin\theta; \alpha)}\right] \end{aligned}$$

subject to the demand equation (B.1).

For any aggregate skill profile θ, the firm chooses its aggregate worker quality Λ (or equivalently its output y) under the demand equation, which yields the standard mark-up condition

$$p(\alpha, z) = c(\theta; \alpha, z)\ \frac{\sigma}{\sigma - 1}, \tag{B.2}$$

and the output

$$y(\alpha, z) = I\ \frac{\sigma - 1}{\sigma}\ \frac{c(\theta; \alpha, z)^{-\sigma}}{\int c(\theta; \alpha, z)^{1-\sigma} H^f(\mathrm{d}\alpha, \mathrm{d}z)}, \tag{B.3}$$

where

$$c(\theta; \alpha, z) = \frac{w(\cos\theta, \sin\theta)}{z\ F(\cos\theta; \sin\theta; \alpha)}$$

is the firm's constant marginal cost. The firms' profit is decreasing in the unit cost, hence the aggregate skill profile θ is chosen to minimize the cost:

$$\tilde{c}(\alpha, z) = \min_\theta c(\theta; \alpha, z) = \min_\theta \frac{w(\cos\theta, \sin\theta)}{z\ F(\cos\theta; \sin\theta; \alpha)}. \tag{B.4}$$

The above determination of the aggregate profile θ is exactly the same as in the main text, see for instance Figure 2. However, the resulting sorting differs from the one of the main text essentially because the equations that define it under Dixit-Stiglitz (DS) are not similar to those obtained for our Bundling analysis in ways that we do not study here. However, our AKM decomposition still holds in this DS environment.

From output (B.3), we obtain the resulting labor demand:

$$\begin{aligned}\Lambda(\alpha, z) &= \frac{y(\alpha, z)}{zF(\cos\theta, \sin\theta; \alpha)} \\ &= I\frac{\sigma - 1}{\sigma}\frac{1}{\int \tilde{c}(\alpha, z)^{1-\sigma} H^f(\mathrm{d}\alpha, \mathrm{d}z)}\frac{[zF(\cos\theta, \sin\theta; \alpha)]^{\sigma-1}}{w(\cos\theta, \sin\theta)^{\sigma}},\end{aligned} \tag{B.5}$$

and the wage bill:

$$\begin{aligned}w(T) &= \Lambda(\alpha, z)\, w(\cos\theta, \sin\theta) \\ &= I\frac{\sigma - 1}{\sigma}\frac{1}{\int \tilde{c}(\alpha, z)^{1-\sigma} H^f(\mathrm{d}\alpha, \mathrm{d}z)}\left[\frac{zF(\cos\theta, \sin\theta; \alpha)}{w(\cos\theta, \sin\theta)}\right]^{\sigma-1}\end{aligned} \tag{B.6}$$

The two environments, the one presented in this Appendix and the competitive one from the main text, have deep similarities. Equations (B.5) and (B.6) replace Equations (A.6) and (A.7). The labor demand elasticity is $\sigma > 1$ here and $1/(1-\eta) > 1$ in the main text. The equilibrium conditions (21) and (22) must be modified according to (B.5) and (B.6). The primal problem consists here in maximizing total welfare (consumer utility minus total costs) instead of total output. But the resulting welfare in this Dixit-Stiglitz environment must be lower than that of the purely competitive environment, because of monopolistic competition. All results obtained under Bundling have their counterpart in the DS world despite differences in the resulting formulas. Furthermore, our analysis of unbundling can also be carried out within the DS setting.