Commentary

BIG DATA & SOCIETY

Big Data & Society
January–June: 1–5

DOI: 10.1177/20539517211003120
journals.sagepub.com/home/bds
SAGE

One size does not fit all: Constructing complementary digital reskilling strategies using online labour market data

Fabian Stephany[1,2]

Abstract

Digital technologies are radically transforming our work environments and demand for skills, with certain jobs being automated away and others demanding mastery of new digital techniques. This global challenge of rapidly changing skill requirements due to task automation overwhelms workers. The digital skill gap widens further as technological and social transformation outpaces national education systems and precise skill requirements for mastering emerging technologies, such as Artificial Intelligence, remain opaque. Online labour platforms could help us to understand this grand challenge of reskilling en masse. Online labour platforms build a globally integrated market that mediates between millions of buyers and sellers of remotely deliverable cognitive work. This commentary argues that, over the last decade, online labour platforms have become the 'laboratories' of skill rebundling; the combination of skills from different occupational domains. Online labour platform data allows us to establish a new taxonomy on the individual complementarity of skills. For policy makers, education providers and recruiters, a continuous analysis of complementary reskilling trajectories enables automated, individual and far-sighted suggestions on the value of learning a new skill in a future of technological disruption.

Keywords

Artificial intelligence, automation, Big Data, networks, online labour platforms, skills

Technological change: A call for reskilling

Technology is changing the way we work, fundamentally. In economic literature, the periodic warning that automation and new technologies were going to terminate large numbers of jobs is an "evergreen" (Acemoglu and Autor, 2011; Brynjolfsson and McAfee, 2014; Frey and Osborne, 2017). In contrast to recurring fears of mass unemployment, current literature shows that the (digital) technology revolution, rather automates tasks than entire occupations (Autor, 2015). Technological and social transformation change the skill composition of professions (Acemoglu & Autor, 2011). The work that is thereby eliminated has different skill requirements than the newly created jobs, which leads to the paradox of simultaneous unemployment and labour shortage (Autor, 2015). Professional service or admin—"white collar"—jobs are particularly exposed to this trend, which is "hollowing-out" the middle employment spectrum of the labour market (Baldwin and Forslid, 2020). The strategies to reskill are often unclear, as the future benefits of learning a new capacity are uncertain and precise skill requirements for mastering emerging technologies remain opaque (De Mauro et al., 2018).

A conventional policy response, aligning national education systems with changing labour market demand, is increasingly ineffectual as technological and social transformation outpaces national education systems (Collins and Halverson, 2018). Likewise, large employers are struggling to keep their workforces' skills up to date (Illanes et al., 2018). At the same time, the COVID-19 pandemic has tightened company

[1]Oxford Internet Institute, University of Oxford, Oxford, UK
[2]Humboldt Institute for Internet and Society, Berlin, Germany

Corresponding author:
Fabian Stephany, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK.
Email: fabian.stephany@oii.ox.ac.uk

budgets, forced employees to work remotely, and facilitated the global need for reskilling (Stephany et al., 2020a). Workers have begun to assume greater personal responsibility for reskilling, via online courses, distance education tools, and entrepreneurial approaches to work (Allen and Seaman, 2015).

Most recently, just-in-time skills development, motivated by perceived market shifts, has emerged, as formal training courses are unaffordable for some workers (Kester et al., 2006). In addition, female participation is still hindered by cultural aspects in traditional science, technology, engineering, and math (STEM) education (Kahn and Ginther, 2017). Instead research shows that independent professionals, including women, prefer informal, digital, social learning resources like *Stack Overflow* and tutorial videos to develop new skills (Yin et al., 2018).

From songs to skills: The rebundling paradigm

In the field of reskilling, the individualistic approach of workers to learn new skills in a fragmented, incremental, and digital fashion has, once again, fostered one of the defining paradigms of the Internet: rebundling (McManus et al., 2018). One of the first sectors to witness the powerful emergence of rebundling was the music industry. First, in the early days of the commercial Internet, download platforms allowed music lovers to access songs individually without having to acquire the band's entire album. The single item (song) was unbundled from the original bundle (album). Later, at a second stage, streaming platforms, like *Spotify*, reversed the trick by allowing the (re)bundling of previously unrelated items. Users could listen to songs from different artists for one single price. The mastery of this strategy has made digital entertainment companies superstars' firms (Eriksson et al., 2019).

Similarly, this paradigm has affected the way we learn new skills. Initially, digital technologies allowed education providers to offer topical online courses (Wulf et al., 2014). In a second stage, e-learning platforms like *Udacity* performed the rebundling and offered a whole set of topical courses for one single price (Bates, 2019). The acquisition of individual skills (programming in *Python* or designing a logo) has been detached from its original domain of training (studying informatics or graphic design).

Online labour platforms: Laboratories for skills

In fact, newest findings show that independent IT professionals today develop new skills incrementally, adding closely related skills to their existing portfolio, as argued by Lehdonvirta et al. (2019). Their work examines the skill development of freelancers on online labour platforms (OLPs). These platforms are websites that mediate between buyers and sellers of remotely deliverable cognitive work (Horton, 2010).[1] OLPs can be further subdivided into microtask platforms, e.g. *Amazon Mechanical Turk*, where payment is on a piece rate basis or freelancing platforms, such as *UpWork*, where payment is on an hourly or milestone basis (Kässi and Lehdonvirta, 2018). In light of the COVID-19 pandemic and its significant economic repercussions across industries (Stephany et al., 2020a), OLPs continue to increase in popularity due to a general trend of work at distance (Stephany et al., 2020b).

In this role, OLPs might have become early "laboratories" for the de- and rebundling of incremental skills. It could be argued that, for work, OLPs have turned into what streaming providers are for music: Freelancers can sell previously unrelated skills in one single portfolio for one hourly price, as illustrated by Figure 1. The role of the Data Scientist is a prime example of the rebundling of skills from different domains, i.e., visualisation, programming and statistics, as an economically profitable offer.

The general success of the Data Scientist skill bundle is only one of many examples for the profitability of skill rebundling and cross-skilling strategies; i.e., the combination of skills from different occupational domains. Anderson (2017), for example, shows that freelancers with diverse skill portfolios are able to gain higher wages, on average. Similarly, Stephany (2020) illustrates that the acquisition of a new skill from a different, but adjacent skill domain is related to higher asking wages of online freelancers, while the benefit of the individual capacity depends on its complementarity with the skill bundle it is attached to.

The network perspective: Establishing new skill taxonomies

In the situation of rapidly changing skill requirements, systematic oversight is essential. However, individuals often lack foresight into which skills are rising and falling, which skills are most valuable, and, most importantly, which skills their existing portfolio is complementary to. They might get locked into path dependencies that may result in dead ends that prevent them from reskilling into new areas (Escobari et al., 2019).

The work by Anderson (2017) or Stephany (2020) shows how OLP data allows us to monitor skill rebundling in a global workforce by near real-time with up-to-date skill bundles on a granular level. They use the rich toolbox of network analysis for the

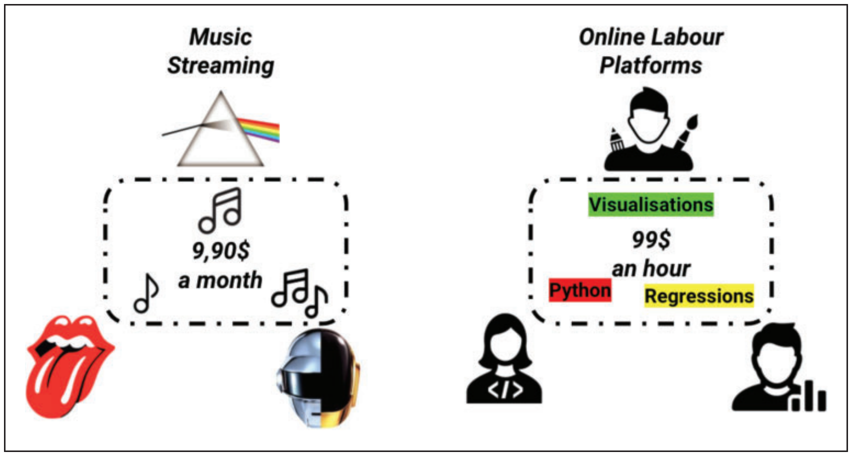

Figure 1. Music streaming platforms allow users to rebundle individual songs from different artists for a monthly charge (lhs). Via online labour platforms, previously unrelated domain specific skills appear now rebundled under a new profession, e.g. Data Scientist (rhs).

characterisation of skill relationships. Given a sample of freelancers with multidimensional skill portfolios, the authors construct a human capital network in which skills are nodes and two skills are connected by a link if a worker has both. This skill network provides the researchers with an endogenous categorisation of skills based on their relationship in application and the context dependency of human capital, as illustrated by Stephany (2020) shown in Figure 2.

Most importantly, the taxonomies emerging from the analysis of large-scale OLP data allow researchers to investigate the complementarities of learning. One and the same skill can inflict different costs and leverage different benefits depending on the learner's existing skill set that it complements. With the example of programming languages, Stephany (2020) illustrates that learning *Java* is of limited economic benefit in the field of Data Engineering where learning how to programme in *Python,* on the other hand, increases worker wages significantly. For the field of 3D design, however, the picture flips and *Java* yields much higher contributions to worker wages than the allegedly 'super star' programming language *Python* (Grus, 2019). Similarly, complementary costs in learning can be reduced as previously acquired skills may lower the entry barrier into new skill domains, e.g., via the underlying similarity of language logics across programming languages.

Pathways into the future: Sketching reskilling strategies

The global workforce is under constant pressure to reskill, as technological change accelerates and task automation reshuffles occupational skill requirements. Labour market mismatches must be avoided but traditional reskilling, via national education policies, is too slow for the fast pace of technological change and precise skill requirements for emerging new technologies are too fuzzy. In addition, economic tightening due to situations like the COVID-19 lockdown further accelerates digitalisation trends while forcing workers to reskill from home. Workers have already taken matters into their own hands by incremental and just-in-time skill development via social learning tutorial and online education providers.

The first markets witnessing this individualistic reskilling trend are OLPs. Here, entirely new skill compositions are offered, at times, with very profitable margins for workers. In fact, OLPs have become early "laboratories" for the de- and rebundling of skills from previously unrelated domains. They allow us to observe the rebundling of skills into yet unnamed occupations. The statistical analysis of OLP data holds a tremendous opportunity for revealing endogenous categorisation of skills via skill networks and for giving us insights into the value of complementary reskilling. First investigations reveal that diverse and innovative skill compositions can be extremely profitable for some workers, like Data Scientists (Anderson, 2017; Stephany, 2020).

The complementarity in costs and benefits of learning new skills make reskilling decisions a complex and highly individual endeavour with no single best choice for everybody. The analysis of OLP data can help estimating the complementarities of learning a new skill, which is key for finding individual and sustainable, i.e. continuously profitable, reskilling pathways. It supports establishing a new taxonomy of digital reskilling opportunities so that job market mismatches can be reduced, at scale. In cooperation with freelance platforms and online education providers, researchers

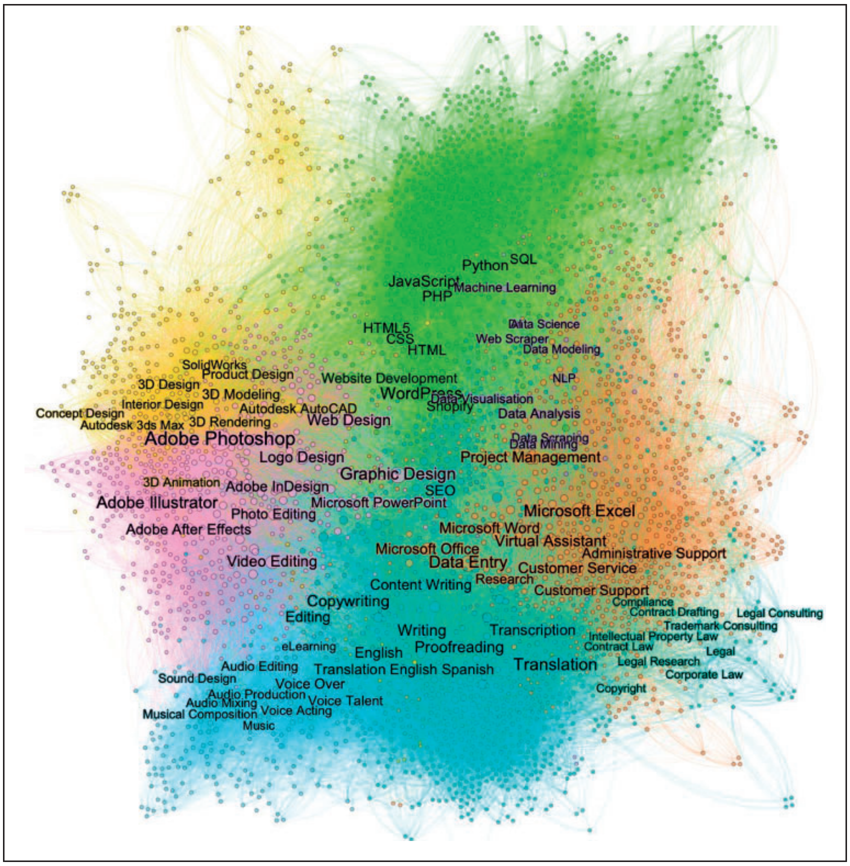

Figure 2. In a skill network, 3525 skills are connected if jointly advertised by the same worker. Skills group in eight clusters with different degrees of centrality (node size). Seven clusters emerge: 3D design (top left), admin support (middle right), audio design (bottom left), data engineering (top right), graphic design (middle left), legal (bottom right), Software and technology (top middle), and translation and writing (bottom middle). Recoloured source: Stephany (2020).

and policy makers should consider using this blueprint to give learners a personalised recommendation on which skills to develop complementing their own capacities. A continuous time series analysis of complementary cross-skilling trajectories (Stephany, 2020), via OLP data, enables automated, individual, and far-sighted suggestions on the value of learning a new skill in a future of technological disruption.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Fabian Stephany https://orcid.org/0000-0002-0713-6010

Note

1. The sellers of work on OLPs are either people in regular employment earning additional income by "moonlighting" via the Internet as freelancers or they are self-employed independent contractors. The buyers of work range from individuals and early-stage startups to Fortune 500 companies (Corporaal and Lehdonvirta, 2017).

References

Acemoglu D and Autor D. Skills, tasks and technologies: Implications for employment and earnings. In: *Handbook of labor economics*. Vol. 4. Amsterdam: Elsevier, 2011, pp. 1043–1171.

Allen IE and Seaman J (2015) *Grade level: Tracking online education in the United States*. Babson Survey Research Group, Babson College, USA.

Anderson KA (2017) Skill networks and measures of complex human capital. *Proceedings of the National Academy of Sciences of the United States of America* 114(48): 12720–12724.

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

Baldwin R and Forslid R (2020) *Globotics and development: When manufacturing is jobless and services are tradable* (No. w26731). National Bureau of Economic Research, USA.

Bates T (2019) What's Right and What's Wrong about Coursera-Style MOOCs. In Kimmons R (ed) *EdTech in the Wild*. EdTech Books. https://edtechbooks.org/wild/mooc_right_wrong

Brynjolfsson E and McAfee A (2014) *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, NY: WW Norton & Company.

Collins A and Halverson R (2018) *Rethinking Education in the Age of Technology: The Digital Revolution and Schooling in America*. New York, NY: Teachers College Press.

Corporaal GF and Lehdonvirta V (2017) *Platform Sourcing: How Fortune 500 Firms are Adopting Online Freelancing Platforms*. Oxford: Oxford Internet Institute.

De Mauro A, Greco M, Grimaldi M, et al. (2018) Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management* 54(5): 807–817.

Eriksson M, Fleischer R, Johansson A, et al. (2019) *Spotify Teardown: Inside the Black Box of Streaming Music*. MIT Press.

Escobari M, Seyal I and Meaney M (2019) *Realism about Reskilling: Upgrading the Career Prospects of America's Low-Wage Workers. Workforce of the Future Initiative*. Center for Universal Education at The Brookings Institution.

Frey CB and Osborne MA (2017) The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* 114: 254–280.

Grus J (2019) *Data Science from Scratch: First Principles with Python*. Newton, MA: O'Reilly Media.

Horton JJ (2010) Online Labor Markets. In Saberi A (ed) *Internet and Network Economics*. Springer: Berlin Heidelberg, pp.515–522.

Illanes P, Lund S, Mourshed M, et al. (2018) *Retraining and reskilling workers in the age of automation*. McKinsey Global Institute.

Kahn S and Ginther D (2017) *Women and STEM (No. w23525)*. National Bureau of Economic Research.

Kässi O and Lehdonvirta V (2018) Online labour index: Measuring the online gig economy for policy and research. *Technological Forecasting and Social Change* 137: 241–248.

Kester L, Lehnen C, Van Gerven PW, et al. (2006) Just-in-time, schematic supportive information presentation during cognitive skill acquisition. *Computers in Human Behavior* 22(1): 93–112.

Lehdonvirta V, Margaryan A and Davies HUW (2019) Skills formation and skills matching in online platform work: Policies and practices for promoting crowdworkers' continuous learning. CrowdLearn.

McManus B, Nevo A, Nolan Z, et al. (2018) *Steering Incentives and Bundling Practices in the Telecommunications Industry*. NET Institute Working Paper No. 18-12, Available at SSRN: https://ssrn.com/abstract = 3267060

Stephany F (2020) When does it pay off to learn a new skill? Revealing the complementary benefit of cross-skilling. *arXiv preprint arXiv:2010.11841*. https://arxiv.org/abs/2010.11841

Stephany F, Lehdonvirta V, Sawyer S, et al. (2020) Distancing bonus or downscaling loss? The changing livelihood of US online workers in times of COVID-19. *Tijdschrift Voor Economische En Sociale Geografie = Journal of Economic and Social Geography = Revue de Geographie Economique et Humaine = Zeitschrift Fur Okonomische Und Soziale Geographie = Revista de Geografia Economica y Social* 111(3): 561–573.

Stephany F, Stoehr N, Darius P, et al. (2020) The CoRisk-Index: A data-mining approach to identify industry-specific risk assessments related to COVID-19 in real-time. *arXiv preprint arXiv:2003.12432*. https://arxiv.org/abs/2003.12432

Wulf J, Blohm I, Leimeister JM, et al. (2014) Massive open online courses. *Business & Information Systems Engineering* 6(2): 111–114.

Yin P, Deng B, Chen E, et al. (May 2018) Learning to mine aligned code and natural language pairs from stack overflow. In: *2018 IEEE/ACM 15th international conference on mining software repositories (MSR)*, pp. 476–486. Piscataway, NJ: IEEE.