

## ECONOMICS

## Unpacking the polarization of workplace skills

Ahmad Alabdulkareem<sup>1,2\*</sup>, Morgan R. Frank<sup>3\*</sup>, Lijun Sun<sup>4</sup>, Bedoor AIShebli<sup>5</sup>,  
César Hidalgo<sup>3</sup>, Iyad Rahwan<sup>1,3†</sup>

Economic inequality is one of the biggest challenges facing society today. Inequality has been recently exacerbated by growth in high- and low-wage occupations at the expense of middle-wage occupations, leading to a “hollowing” of the middle class. Yet, our understanding of how workplace skills drive this process is limited. Specifically, how do skill requirements distinguish high- and low-wage occupations, and does this distinction constrain the mobility of individuals and urban labor markets? Using unsupervised clustering techniques from network science, we show that skills exhibit a striking polarization into two clusters that highlight the specific social-cognitive skills and sensory-physical skills of high- and low-wage occupations, respectively. The connections between skills explain various dynamics: how workers transition between occupations, how cities acquire comparative advantage in new skills, and how individual occupations change their skill requirements. We also show that the polarized skill topology constrains the career mobility of individual workers, with low-skill workers “stuck” relying on the low-wage skill set. Together, these results provide a new explanation for the persistence of occupational polarization and inform strategies to mitigate the negative effects of automation and offshoring of employment. In addition to our analysis, we provide an online tool for the public and policy makers to explore the skill network: [skillscape.mit.edu](https://skillscape.mit.edu).

## INTRODUCTION

Economic inequality is on the rise, making it one of the central challenges facing U.S. policy makers today (1). For example, absolute income mobility—the fraction of children who earn more than their parents—has fallen markedly in the United States, from 90% for children born in 1940 to 50% for children born in 1980 (2). Some declared that the diminishing opportunity for prosperity and success marks the fading of the “American dream” (3, 4), an ideal that is intimately associated with the U.S. national identity and ethos.

In contemporary political debate, one of the main culprits behind economic inequality has been the lack of “good jobs.” Both nationally and in a majority of U.S. metropolitan areas (5), economists have identified occupational polarization: an increasing proportion of high- and low-wage employment, accompanied by a relative decrease in employment share in middle-wage occupations (6–8). The result is a “hollowing” of the middle class. Mechanisms driving this trend include the offshoring of work (9), something that has triggered recent shifts in international trade policy. Another mechanism is the automation of routine work, something that has sparked major concerns about the impact of automation on the future of work (10–12).

However, while mechanisms like offshoring and automation ultimately affect people’s jobs, they do not typically operate at the level of occupations. Rather, they alter the demand for specific workplace skills, tasks, knowledge, and abilities (hereafter referred to as “skills”). If individual workers—or even entire cities—are unable to appropriately adapt their own skills, then their ability to compete in the national and global labor market may be diminished.

Despite the important role of skills in occupational polarization, existing studies have explained the hollowing of the middle class in terms of annual wages (13) and broad, subjectively defined occupational

categories, such as “cognitive” versus “physical” or “routine” versus “non-routine” (6). For example, suppose we use wage as a proxy for skill—that is, high-wage occupations are considered high-skilled occupations, etc. Then, if we find that growth in employment in middle-wage occupations is slower than that in low- and high-wage occupations, we may conclude that the demand for high and low skills is driving economic inequality. But this coarse-grained distinction may miss important relationships between skills that affect how workers adapt. This motivates the first set of questions we wish to explore in this study:

Q1. Can we recover occupational polarization, at the finer-grained level of underlying skills, using an objective (unsupervised) data-driven clustering? How many distinct clusters, if any, does this skill structure contain? And does the skill structure exhibit smooth or abrupt transition between skill clusters?

To answer these questions, we apply data-driven methods to map skill complementarity as a network. We then use techniques from network science to identify distinct clusters of skills. Since we use an unsupervised methodology, we demonstrate the usefulness of the resulting skill network by relating its structure to important real-world labor dynamics. Workers leverage skill complementarity between their existing skills to make career changes (14). Similarly, cities leverage complementarity between industries to optimize productivity and increase their competitiveness in a global economy (15–18). We find that the structure of skill complementarity explains many stylized observations about occupational polarization and the hollowing of the middle class.

Having mapped the structure of skills and identified aggregate structure, the next obvious question to ask is, “Does the granular structure matter?” Studies have identified the aggregate effects of skill complementarity on labor dynamics, such as the redefinition of skills comprising each occupation (12). We unpack the role of skill complementarity in labor dynamics by exploring the following additional questions:

Q2. Can the skill topology predict changes in the latent skills of different urban labor markets (cities)? That is, given the skills used effectively in a given city at time  $t$ , can the network structure help us predict which new skills will become competitive in that city at time  $t + 1$ ?

Q3. Can the skill topology help us predict changes in the skill requirements of a given job—that is, how the job’s requirements change over time?

<sup>1</sup>Institute for Data, Systems, and Society, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. <sup>2</sup>Center for Complex Engineering Systems at MIT and King Abdulaziz City for Science and Technology, Riyadh 12371, Saudi Arabia. <sup>3</sup>Media Laboratory, MIT, Cambridge, MA 02139, USA. <sup>4</sup>Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Québec H3A 0C3, Canada. <sup>5</sup>Electrical Engineering and Computer Science, Masdar Institute of Science and Technology, Abu Dhabi, UAE.

\*These authors contributed equally to this work.

†Corresponding author. Email: [irahwan@mit.edu](mailto:irahwan@mit.edu)

Q4. Can the skill topology help us predict changes in the skills of individual workers as they transition from one job to another?

Having shown that skill polarization exists and affects some key dynamics, we ask:

Q5. Is the mobility of individual workers between skill sets (as they change jobs) consistent with the polarized structure of skills?

Our analysis suggests that the answer is “yes.” We provide three types of evidence: (i) Workers tend to transition between occupations relying on the same skill set; (ii) workers are unable to switch away from occupations relying equally on cognitive and physical labor; and (iii) this constraining effect is reflected in the national employment statistics.

In the next section, we describe our methodology in detail. We then present our analysis and discuss its implications and potential weaknesses before concluding the paper.

## MATERIALS AND METHODS

The O\*NET program by the U.S. Department of Labor annually produces the publicly available O\*NET database detailing the importance of 161 workplace skills, knowledge, and abilities for the completion of each of the 672 occupations recognized under the Standard Occupational Classification (SOC) System. The O\*NET database is updated regularly, allowing for annual snapshots of the relationships between occupations and skills through continual survey of workers from each occupation. We used annual O\*NET data from the years 2010 through 2015. We denoted the importance of skill  $s \in S$  to occupation  $j \in J$  using  $\text{onet}(j, s) \in [0, 1]$ , where  $\text{onet}(j, s) = 1$  indicates that  $s$  is essential to  $j$ , while  $\text{onet}(j, s) = 0$  indicates that workers of occupation  $j$  need not possess or perform  $s$ .

The Bureau of Labor Statistics (BLS) annually produces publicly available data detailing the distribution of SOC occupations in each U.S. metropolitan statistical area (MSA). MSAs represent an entire urban system, including areas with large proportions of commuters employed in the city proper. We interchangeably used the terms “MSA” and “city.” Along with the numbers of workers of each occupation, the BLS provides additional details about the annual salary of each occupation in each city.

The U.S. Census Bureau and the BLS produce a monthly Current Population Survey (CPS) through a continuous survey process that produces representative samples of the U.S. population. Providing high-resolution labor statistics is one of the primary goals of CPS; in particular, CPS records changes in occupations of survey participants over the 1.5-year period for which that participant is an active contributor to the survey. For our purpose, we are interested only in participants who reported one occupation when they were first surveyed in 2014 and reported working a different occupation when they were surveyed 1 year later in 2015. There are several methods for joining different time periods of the CPS data (19), so we used a strict merging criteria, including participant ID, gender, sex, state of residency, and age to verify the validity of our occupational transitions. The result was a data set of 5400 occupational transitions for individual U.S. workers from 2014 to 2015.

## RESULTS

### Mapping skill complementarity

Typically, occupations are the units of interest in labor dynamics. However, in other situations, occupations are broken down even further because the labor requirements that define an occupation are reflected in

the skills possessed by workers of that occupation (see Fig. 1A). These skill requirements represent key features that uniquely identify occupations, and so, we seek a data-driven methodology that maximizes the information about each occupation while minimizing the potential bias that can accompany investigations through ad hoc skill aggregations. However, raw O\*NET data do not control for ubiquitous skills, such as “Identifying Objects” and “Communicating with Supervisors and Peers” (see fig. S1). Therefore, we focus on skills that are overexpressed in an occupation by calculating the revealed comparative advantage (RCA) (20–22) of each skill in an occupation according to

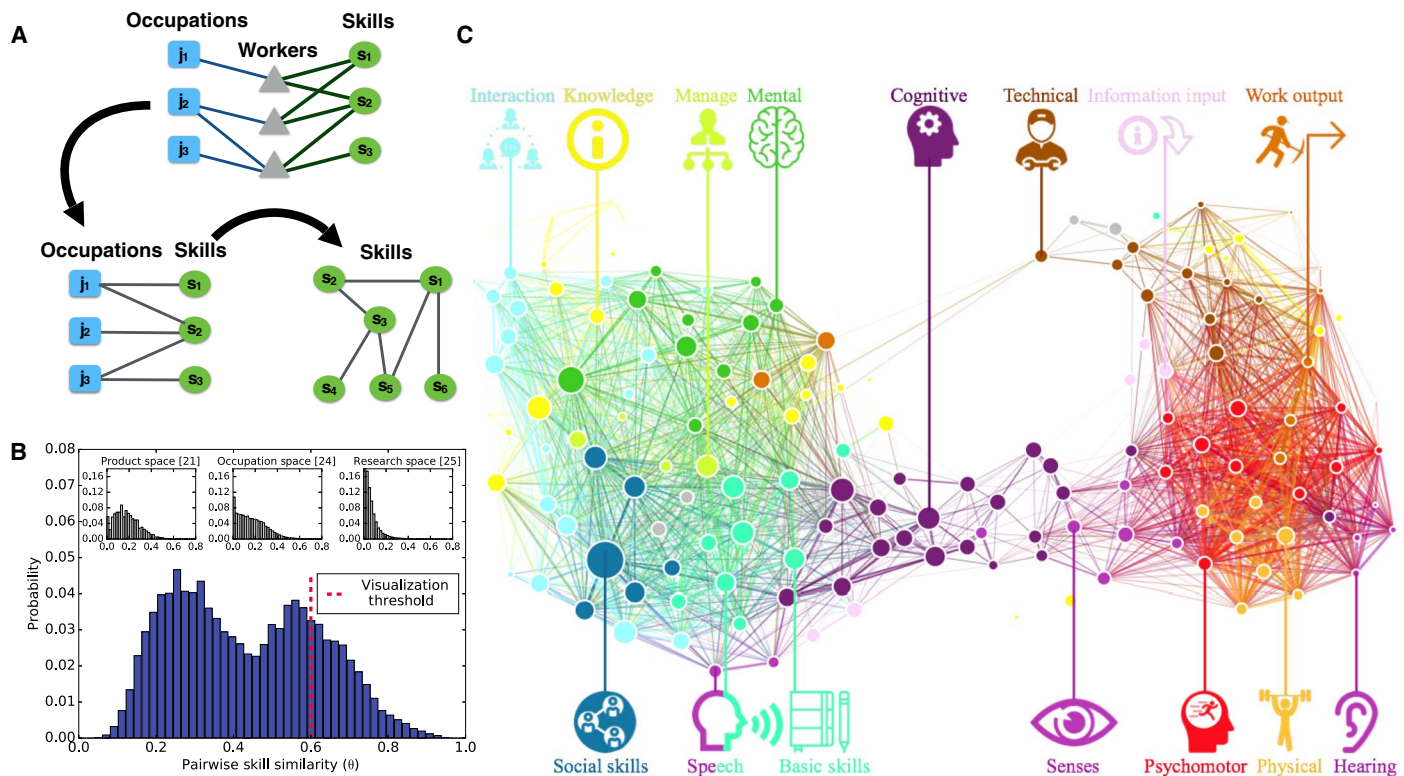
$$\text{rca}(j, s) = \frac{\text{onet}(j, s) / \sum_{s' \in S} \text{onet}(j, s')}{\sum_{j' \in J} \text{onet}(j', s) / \sum_{j' \in J, s' \in S} \text{onet}(j', s')} \quad (1)$$

RCA (also known as “location quotient”) has been used in a variety of applications, including identifying the key industries in cities (23–25), key exports of nations (20, 26), and key features in the labor distributions of industries (27). Similarly, occupations are distinguishable from each other according to their “effective use” of skills; we denote effective use of skills using  $e(j, s) = 1$  if  $\text{rca}(j, s) > 1$ , and  $e(j, s) = 0$  otherwise. Here, RCA normalization compares the relative importance of a skill to an occupation (that is, the numerator in Eq. 1) to the expected relative importance of a skill on aggregate (that is, the denominator);  $\text{rca}(j, s) > 1$  indicates that occupation  $j$  relies on skill  $s$  more than expected on aggregate. Skill complementarity (denoted  $\theta$ ) (14, 17) is then the minimum of the conditional probabilities of a pair of skills being effectively used by the same occupation

$$\theta(s, s') = \frac{\sum_{j \in J} e(j, s) \cdot e(j, s')}{\max\left(\sum_{j \in J} e(j, s), \sum_{j \in J} e(j, s')\right)} \quad (2)$$

The distribution of complementarity values is provided in Fig. 1B. This methodology identifies skill pairs that co-occur across occupations and represent key occupational features. Co-occurrence captures how a pair of skills supports each other, either by boosting the productivity of a worker who possesses both skills or by the ease of simultaneously acquiring both skills. Our definition of complementarity is agnostic to the exact source of the complementarity. We call the resulting network of skill complementarity the “Skillscape” (see Fig. 1C and also section S1 for visualizations of this methodology and a visualization of the Skillscape as a skill-to-skill complementarity matrix).

Ideally, the aggregate structure in the skill network should correspond to meaningful labor dynamics. For example, node communities in the skill network represent clusters of complementary skills that define important types of labor. To this end, we identify skill types using the Louvain community detection (28). This method greedily identifies node communities by comparing the density of connections within a community to the density of connections between communities. This method requires no assumptions about the number of communities to be found. This community detection method has been widely used in a variety of fields, including neuroscience (29, 30), transportation research (31), social science (32), business/management research (33), climatology (34), and cybersecurity (35).



**Fig. 1. Constructing the Skillscape.** (A) An occupation is identified through the skills of workers of that occupation. The bipartite network connecting occupations to required skills is a result of an underlying tripartite network containing workers as a conduit between occupations and skills. Relationships between skills are determined from their co-occurring importance across occupations. (B) Unlike previous applications of RCA (insets), the Skillscape contains a bimodal distribution of pairwise skill complementarity. (C) The Skillscape thresholded according to a minimum skill similarity (that is,  $\theta > 0.6$ ) visibly reveals two communities of complementary skills and respects expertly derived O\*NET categories (colors). Node sizes reflect the total skill similarity shared between that skill and all other skills.

### Identifying skill polarization from the bottom-up

Existing studies have explained the hollowing of the middle class in terms of annual wages (13) and broad, subjectively defined occupational categories, such as cognitive versus physical or routine versus nonroutine (6). For example, it has been shown that some decades are marked by a relative increase in the share of employment in high- and low-wage jobs at the expense of workers in middle-wage jobs. While these results identify the outcome of labor polarization, they do not relate this polarization to the underlying topology of skills. The limitations discussed above have led researchers to call for new high-resolution models that more accurately account for raw workplace tasks and skills (8).

On aggregate, our cluster analysis reveals that the skill network is highly polarized into a sociocognitive cluster of skills and a sensory-physical cluster (see Fig. 1C). This polarization is not an artifact of the methods we used (see Fig. 1B) and is significantly different from comparisons to a null model (see section S4). This divide between traditionally “technical” and “nontechnical” skills largely supports previous findings characterizing the U.S. occupational polarization. For example, let SocioCog denote the set of sociocognitive skills according to the community detection algorithm (see Fig. 2A). We measure the cognitive skill fraction of job  $j$  according to

$$\text{cognitive}_j = \frac{\sum_{s \in \text{SocioCog}} \text{onet}(j, s)}{\sum_{s \in S} \text{onet}(j, s)} \quad (3)$$

Jobs with higher cognitive<sub>j</sub> tend to yield higher annual wages (see Fig. 2B; Pearson correlation  $\rho = 0.42$ ,  $P < 10^{-26}$ ). This result demonstrates the direct link between the skill polarization we have identified and the occupational polarization, which is characterized by growing employment share for high- and low-wage occupations (13).

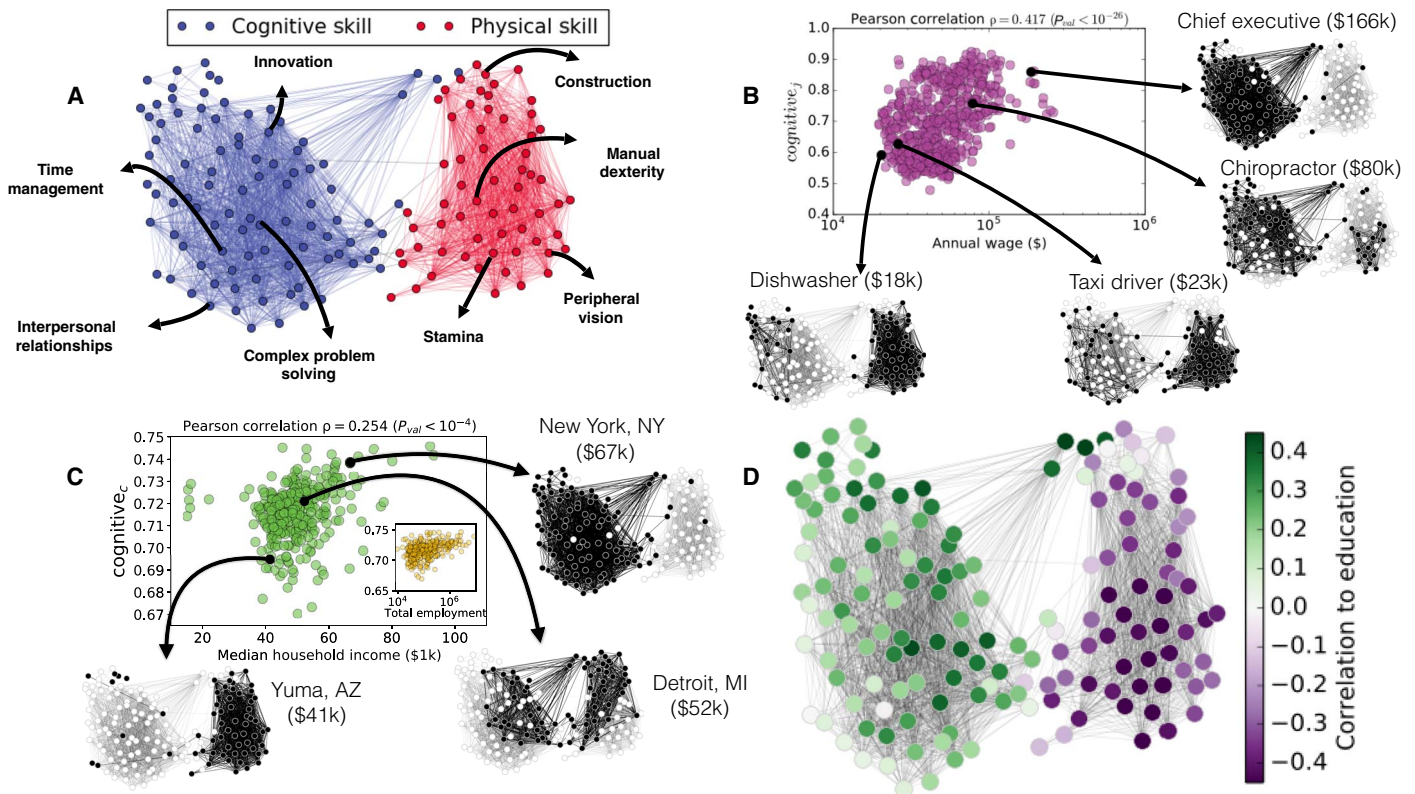
### Comparison with top-down categorization

One might wonder whether our approach to skill polarization captures factors beyond those well known in the literature. Previous work has leveraged ad hoc distinctions between occupations based on their reliance on routine versus nonroutine skills to study occupational polarization (8, 36). Does our approach to skill polarization add further predictive power?

In agreement with the existing work, our investigation of skills should incorporate known worker-related variables, such as education. Education level is a key factor in determining wages (13, 37) as educational institutions act as a social “sorting machine” (37) when students begin their careers. The skill polarization we observe respects the educational requirements of occupations. If we correlate  $\text{onet}(j, s)$  and the average degree requirement for each occupation, we find that skills in the sociocognitive cluster indicate higher education requirements across occupations. Conversely, occupations with more lenient degree requirements tend to rely on sensory-physical skills (see Fig. 2D).

Although the aggregate polarization of skills captures known features that determine worker wages, it remains to show the added predictive power gained from the granularity of our model. In particular,





**Fig. 2. The polarized Skillscape explains occupational wage polarization and economic well-being of urban workforces.** (A) Community detection on the complete Skillscape network (that is, no minimum  $\theta$ ) reveals two communities of complementary skills: sociocognitive skills (blue) and sensory-physical skills (red). The displayed network is filtered ( $\theta > 0.6$ ) for visualization purposes. (B) Occupations relying on sociocognitive skills tend to make higher annual salaries. (C) Larger cities rely more strongly on sociocognitive skills (inset), yielding higher median household income by comparison to smaller cities. In (B) and (C), example occupations (cities), along with their annual wages (median household income), are projected onto the Skillscape using black nodes for effectively used skills. (D) The skill network colored by correlation between  $\text{onet}(j, s)$  and the average educational degree requirement across occupations.

do the existing ad hoc distinction between routine versus nonroutine skills, and the level of education, completely explain the differences in wages? Or does the polarized structure of the skill network we have identified play an independent role? We investigate this question by comparing different regression models in Fig. 3.

In model 1, we consider the relative importance of routine labor by combining the O\*NET data with the routine O\*NET variables defined in (38) [that is,  $\sum_{s \in R} \text{onet}(j, s) / \sum_{s \in S} \text{onet}(j, s)$ , where  $R$  are routine O\*NET variables,  $R^2 = 0.12$ ]. Model 2 demonstrates the superior performance of cognitive<sub>j</sub> ( $R^2 = 0.15$ ). In addition, we consider the total skill content required by each occupation [that is,  $\sum_{s \in S} \text{onet}(j, s)$ ] in model 3 ( $R^2 = 0.30$ ). Models 4 to 6 demonstrate that total skill content and cognitive skill fraction outperform models using the variable for routine labor (model 6 has  $R^2 = 0.46$ ) and that total skill content is largely orthogonal to reliance on cognitive skills. In model 5, we consider variables for each occupation's total employment whose highest educational attainment was a high school diploma, a bachelor's degree, etc. Modeling with these educational variables alone performs worse than using cognitive<sub>j</sub> ( $R^2 = 0.12$ ). Finally, model 8 demonstrates the improved performance from including the variable for routine labor and total skill content ( $R^2 = 0.42$ ), but maximum performance is achieved when including cognitive<sub>j</sub> as well (model 9 has  $R^2 = 0.49$ ). We provide out-of-sample testing to demonstrate the robustness of our models' performance; we find that the inclusion of skill-related variables in models 8 and 9 reduces the variance in model performance. In ad-

dition, the SE and statistical significance of coefficient estimates are reported in the regression table.

In summary, we find that cognitive skill fraction (cognitive<sub>j</sub>) explains the annual wages of occupations better than models using routine labor or educational variables alone. Additional regression analyses detailing occupation wages and the median household income of cities are provided in section S6.

### Skills of urban workforces

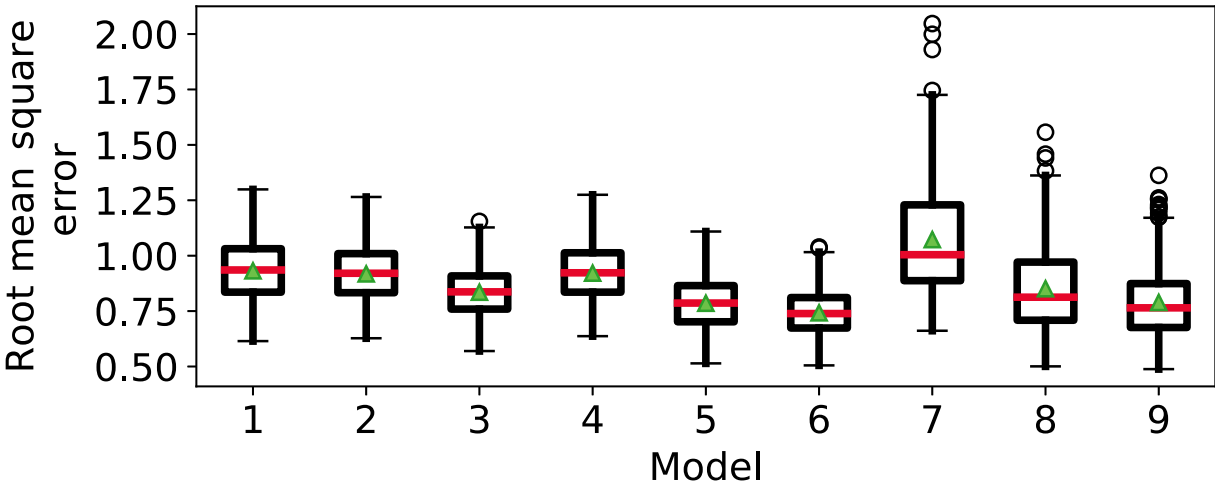
We combine the O\*NET database with employment distributions in U.S. cities according to the BLS to approximate the importance of each workplace skill to each urban workforce. Denoting the number of workers in city  $c$  with occupation  $j$  using  $\text{bls}(c, j)$ , we combine the two data sets according to

$$CS(c, s) = \sum_{j \in J} \text{bls}(c, j) \cdot \text{onet}(j, s) \quad (4)$$

where  $CS(c, s)$  denotes city  $c$ 's reliance on workplace skill  $s$  (see section S5). As with the raw O\*NET data, certain jobs and certain skills are ubiquitous across many cities. We again apply RCA on  $CS(c, s)$  to calculate  $\text{rca}(c, s)$  (as in Eq. 1) and identify which skills are effectively used in each city. Similar to occupations,  $\text{rca}(c, s) > 1$  indicates the effective use of  $s$  in  $c$ . Additional explanatory visualizations are shown in section S5.

Dependent variable: Annual wage of occupations									
Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
cognitive <sub>j</sub>		0.387*** (0.036)		0.352*** (0.078)		0.614*** (0.062)			0.572*** (0.063)
Routine labor	−0.352*** (0.037)			−0.040 (0.078)	−0.288*** (0.031)	0.263*** (0.067)		−0.268*** (0.032)	0.233*** (0.063)
Σ <sub>s</sub> onet( <i>j</i> , <i>s</i> )			0.548*** (0.033)		0.512*** (0.031)	0.574*** (0.040)		0.485*** (0.032)	0.550*** (0.031)
No GED							−0.094* (0.045)	−0.006 (0.037)	0.009 (0.035)
H.S. diploma							−0.170*** (0.047)	−0.086* (0.038)	−0.078* (0.036)
Bachelor's degree							0.099* (0.042)	0.062* (0.035)	0.035 (0.033)
Master's degree							0.064 (0.041)	−0.049 (0.034)	−0.056* (0.032)
Doctoral degree							0.240*** (0.037)	0.191*** (0.031)	0.161*** (0.029)
R <sup>2</sup>	0.124	0.150	0.301	0.151	0.382	0.460	0.126	0.424	0.490
Adj. R <sup>2</sup>	0.122	0.149	0.299	0.148	0.380	0.458	0.119	0.418	0.483

*p* < 0.1\*, *p* < 0.01\*\*, *p* < 0.001\*\*\*



**Fig. 3. Reliance on cognitive skills predicts increased annual wages according to OLS regression.** As a baseline, we consider the relative importance of routine labor using routine O\*NET variables from (38). In addition to cognitive skill fraction (cognitive<sub>j</sub>), we calculate the total skill content [Σ<sub>s</sub> onet(*j*, *s*)] of each occupation. Each educational variable represents the total employment in that occupation whose highest educational degree is a high school diploma, a bachelor's degree, etc. All variables were standardized before regression. SEs are reported in parentheses, and asterisks indicate the statistical significance of coefficient approximations. We perform out-of-sample testing for each model through 1000 trails of randomly selecting 75% of the occupations as training data and measuring the root mean square error of the resulting model applied to the remaining 25% of occupations. We represent the resulting model performance as box plots. Red lines represent median error, while triangles represent the mean error. GED, General Education Diploma.

By considering onet(*c*, *s*) in place of onet(*j*, *s*) in Eq. 3, we can compute the same cognitive skill fraction (denoted cognitive<sub>*c*</sub>) for entire cities. Analogously, Fig. 2C shows that cities with higher median household incomes ( $\rho = 0.25$ ,  $P < 10^{-4}$ ) also tend to rely on sociocognitive skills. We also find a significant correlation between city size and the degree to which the city's local labor market relies on sociocognitive skills: Larger cities are more sociocognitive (see inset in Fig. 2C). Together, these results suggest that inequality between cities may be driven by processes that operate at the level of skill supply and the ability of cities to effectively exploit skill complementary within the sociocognitive niche.

**Skillscape proximity and skill acquisition**

Does skill complementarity (that is,  $\theta$ ) correspond to “nearby” skills in practice? We capture this using a measure for the network “proximity” between each pair of skills based on the network topology and an empirical measure for skill acquisition. Let  $E_t^\lambda(j)$  represent the set of skills

that job *j* effectively uses at time *t* according to some threshold  $\lambda \geq 0$ , that is

$$E_t^\lambda(j) = \{s \in S | rca_t(j, s) > \lambda\}$$
 (5)

We say that a skill is “acquired” if it was not effectively used at time *t*<sub>1</sub> and becomes effectively used at *t*<sub>2</sub>. Specifically, we denote the set of occupation *j*’s acquired skills using

$$\text{Acquired}_{t_1, t_2}^{\lambda_1, \lambda_2}(j) = \{s \in S | s \notin E_{t_1}^{\lambda_1}(j), s \in E_{t_2}^{\lambda_2}(j)\}$$
 (6)

According to this definition, two different thresholds,  $\lambda_1$  and  $\lambda_2$ , are selected for time steps *t*<sub>1</sub> and *t*<sub>2</sub>, respectively. This allows us to vary the magnitude of skill change we are interested in; that is,  $\lambda_2 - \lambda_1$  determines the severity of the skill change in order for a skill to be acquired for  $\lambda_2 > \lambda_1$ . Notice that if  $\lambda_1 > \lambda_2$ , then this would be skill loss instead of acquisition. For the analysis in the main text, we consider discrete choices of  $\lambda$ .

according to each percentile of empirical RCA values (that is,  $\lambda_1, \lambda_2 = 0, 1, \dots, 99, 100\%$  such that  $\lambda_1 < \lambda_2$ ).

For a measure to be predictive of skill acquisition, skills with high scores (for example, in O\*NET) should have higher probability of being acquired for each choice of  $\lambda_1$  and  $\lambda_2$ . For example, if we consider the raw O\*NET values [that is,  $\text{onet}(j, s)$ ] as a proxy for skill acquisition, then skills that are not effectively used by an occupation [that is,  $s \notin E_{t_1}^{\lambda_1}(j)$ ] but have a high score [that is,  $\text{onet}(j, s) \rightarrow 1$ ] should have higher probability of being acquired. We capture this by ordering pairs of occupations and skills by their O\*NET value such that the skill is not effectively used by that occupation [that is,  $s \notin E_{t_1}^{\lambda_1}(j)$ ] and binning these pairs into 30 quantiles according to associated O\*NET value [that is,  $\text{onet}(j, s)$ ]. For each pair, we calculate the probability that the skill is acquired in  $t_2$  (that is,  $s \in \text{Acquired}_{t_1, t_2}^{\lambda_1, \lambda_2}$ ) across all choices of  $\lambda_1$  and  $\lambda_2$ . This produces several points for each quantile; we use the average and the 95% confidence interval for each quantile to simplify the data for visualization. This method is similar to previous studies using network topology to predict the regional acquisition of new industries (17). In the main text, we consider a LOWESS interpolation through the averages of each quantile. In addition to the raw O\*NET as a proxy for skill acquisition, we also consider RCA values and a measure of network skill proximity (described below). In addition to the interpolated plots of the main text, we provide bar plots with the associated error bars in fig. S27.

For noneffectively used skills [that is,  $s \notin E_{t_1}^{\lambda_1}(j)$ ], we say that a skill is nearby to occupation  $j$  if that skill has strong average complementarity with the effectively used skills of  $j$  (that is,  $E_{t_1}^{\lambda_1}$ ). We capture this by introducing a topological measure for proximity according to

$$\text{proximity}(j, s) = \frac{\sum_{s' \in E_{t_1}^{\lambda_1}(j)} \theta(s, s')}{\sum_{s' \in S} \theta(s, s')} \quad (7)$$

This proximity measure only uses information at  $t_1$  to evaluate the status of all skills. Note that analogous calculations can determine Skillscape proximity from urban workforces by considering  $\text{rca}(c, s)$  instead of  $\text{rca}(j, s)$ , and similarly for individual workers. Figures S17 to S21 provide an alternative analysis using receiver operating characteristic curves.

### Dynamics: Skill polarization and transition between jobs

Skill acquisition through explicit education can be costly and time-consuming, so more commonly, workers transition between occupations based on the similarity of their skill set and the skill requirements of each occupation (36). Ideally, the granular network topology of the Skillscape should capture this dynamic. In combination with the aggregate polarization of skills, we also expect that worker mobility between skill categories should be constrained. This hypothesis is not directly testable because we do not understand the precise mechanisms for worker adaptation, nor do we understand the mechanism's interplay with other market equilibrium dynamics (8, 12).

However, the hypothesis reveals three labor trends that the skill network should relate to. First, the topological proximity of skills on the network should relate to skill-related trends, including the changing skill requirements of individual workers, the dynamic skill requirements of occupations, and the changes in the latent skill sets of urban labor markets. Second, if the connections between skills represent skill complementarity, then workers are more likely to transition to occupa-

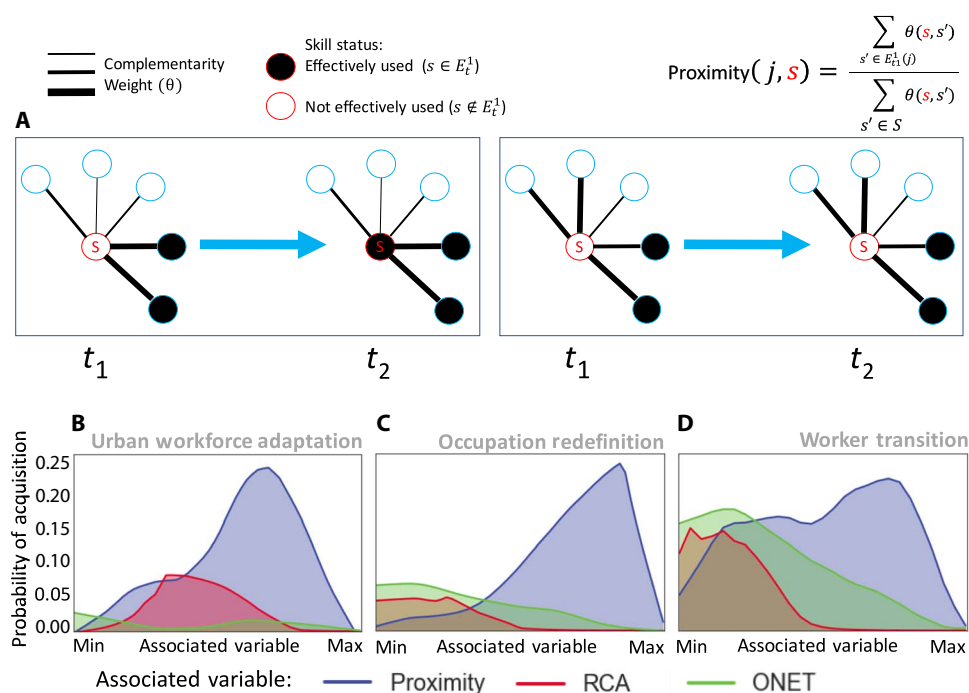
tions relying on skills in the same skill cluster. Third, skill polarization represents a bottleneck in workers' upward mobility toward high-wage occupations. This should lead to disproportionately high employment below a certain cognitive<sub>j</sub> threshold, rather than a smooth distribution of employment across the range of cognitive<sub>j</sub> values. In the remainder, we demonstrate how the Skillscape relates to these important features of the U.S. labor market.

We validate our first prediction in Fig. 4 using a topological measure for skill proximity [that is,  $\text{proximity}(j, s)$ ; see Fig. 4A for an example of Skillscape proximity]. A worker's skill set can be approximated from the skill requirements of his or her occupation, and we suppose that skills that are nearby to these skill sets in terms of network topology are more attainable by that worker. Analogously, nearby skills to a city's local labor market are more likely to be obtained by workers in that city. We empirically validate our proximity measure by comparison to the probability that a skill is acquired (that is,  $s \in \text{Acquired}_{t_1, t_2}^{\lambda_1, \lambda_2}$ ) by a city (see Fig. 4B), an occupation (see Fig. 4C), or an individual worker (see Fig. 4D). In each case, network proximity most strongly indicates newly acquired skills, thus demonstrating the highly granular relationship between the skill network topology and labor dynamics. We provide an alternative analysis in section S7, and bar plots including 95% confidence intervals in section S7.4.

For our second prediction, since occupational transitions represent local changes in workers' skill requirements, the polarized network of skills should constrain mobility between low-wage sensory-physical occupations and high-wage sociocognitive occupations. We capture this explicitly by binning occupational transitions into quantiles (each representing 780 transitions) according to the cognitive skill fraction of the workers' starting occupation (cognitive<sub>j<sub>A</sub></sub>) and examining the average cognitive change (that is,  $\Delta \text{cognitive} = \text{cognitive}_{j_B} - \text{cognitive}_{j_A}$ ; see Fig. 5A) and the average magnitude of cognitive change (Fig. 5B) for each bin. We consider workers selecting their new occupations at random as a null model for comparison (see section S7.1 for a discussion of alternative null models, including randomizing the selection of "cognitive skills"). Workers transitioning from sensory-physical occupations tend toward new occupations with higher sociocognitive skill fraction, but the magnitude of change is less than would be expected under random occupation selection (and vice versa for the other end of the spectrum). By contrast, workers transitioning from mid-quantile occupations, which represent starting occupations that effectively use cognitive and physical skills evenly, exhibit larger magnitudes of change in cognitive<sub>j</sub> compared to the null model. In conclusion, workers of occupations relying strongly on one skill community tend toward other occupations within the same skill community, thus validating the second prediction.

For our third prediction, note first that the definition of skill complementarity (14) indicates increasing returns to combining skills within each skill community. Therefore, skill communities may be explained by the easy acquisition of related skills or by production efficiencies offered by workers who have complementary skills. However, this also means that workers relying on sensory-physical skills will face difficulty acquiring sociocognitive occupations because they are unprepared to exploit large proportions of the sociocognitive skills. Until they have a sufficient proportion of sociocognitive skills, sensory-physical workers are bottlenecked by the polarized structure of skill complementarity. If true, then we expect disproportionately high employment in occupations under some threshold of cognitive<sub>j</sub>.

Binning national employment according to cognitive<sub>j</sub> yields a trimodal distribution (see Fig. 5C; additional years and binning, as well as



**Fig. 4. Skill proximity predicts worker transitions between occupations, skill redefinition of occupations, and skill acquisition in cities.** (A) An example demonstrating Skillscape proximity [that is,  $\text{proximity}(j, s)$ ] as a proxy for the connections between effectively used skills and other skills. (B) Skills with high proximity to the effectively used skills of an urban labor market in 2010 are more likely to be effectively used by that workforce in 2015. (C) Skills with high proximity to the effectively used skills of an occupation in 2010 are more likely to be effectively used by that occupation in 2015. (D) The effectively used skills of a worker's occupation in 2015 are more likely to be effectively used by the workers' next occupation in 2016. We provide bar plots including 95% confidence intervals for these probabilities in section S7.4, and we consider an alternative receiver operator curve analysis in section S7.

city employment distributions, are provided in section S7.2). The upper and lower modes of the distribution correspond to workers who effectively exploit the skill complementarity within each of their respective skill communities. The presence of a third mode in the middle suggests that skill polarization constrains workers from obtaining attractive sociocognitive skills, thus demonstrating the third prediction and adding more evidence toward our hypothesis that the network of skill complementarity constrains labor mobility.

Finally, Fig. 5D quantifies the average complementarity score of each skill as an approximation for that skill's network embeddedness. Considering our hypothesis and the strong relationship between skill proximity and skill acquisition, network embeddedness should correlate with increased labor mobility (individual skills are shown in fig. S6).

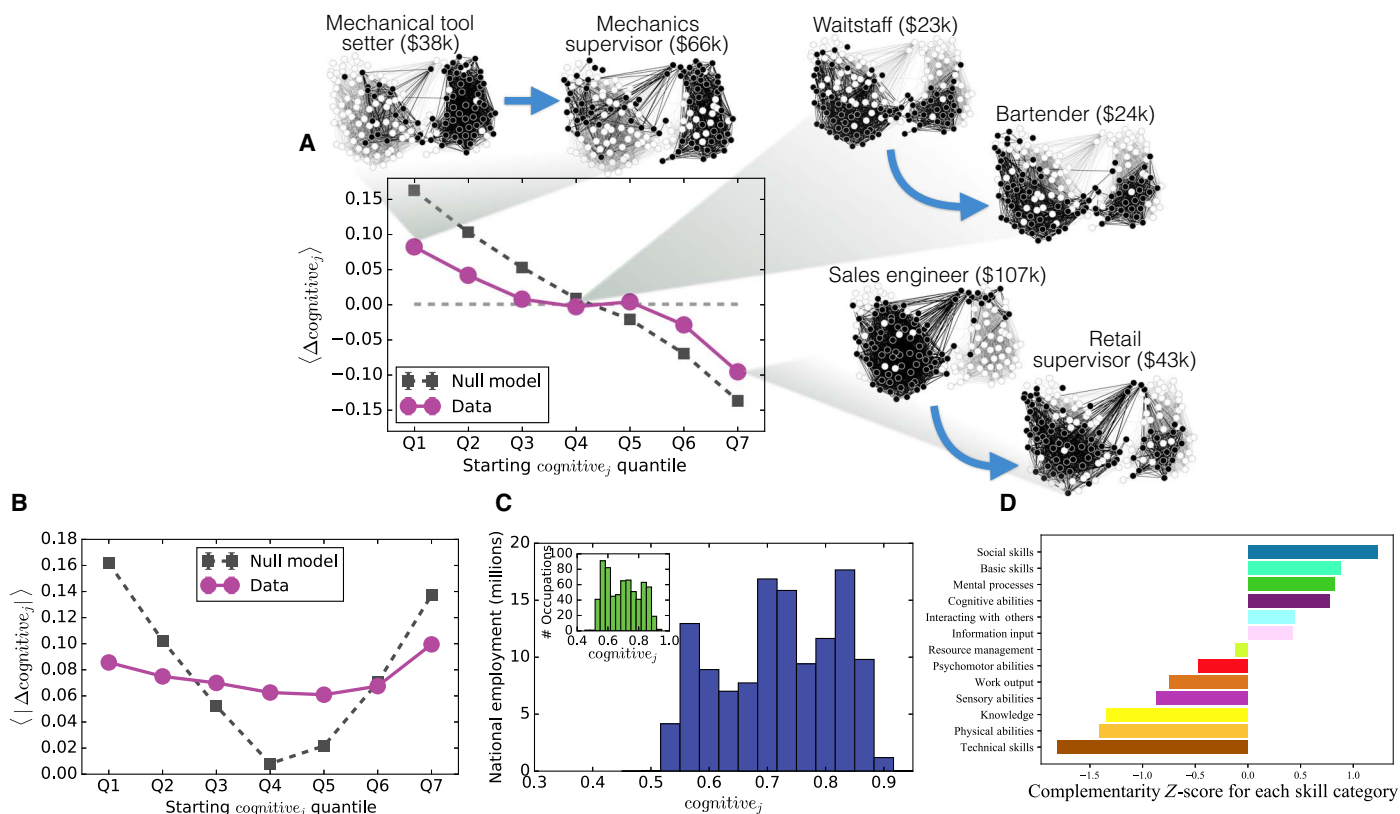
The Skillscape maps the structure of workplace skill complementarity and connects urban workforces and occupations to their constituent skills. While our analysis identifies the specific skill requirements of low- and high-skill occupations that characterize occupational polarization, our analysis does not reveal whether occupational polarization is a result of skill polarization, or vice versa. Many external factors, such as automation (10, 12) and offshoring, likely contribute to both effects. Nevertheless, the Skillscape comprehensively explains the polarization of high- and low-skill occupations as a separation between workers with sociocognitive and sensory-physical skills. This high-resolution framework for understanding workplace skill requirements provides policy makers with a new explanation for stymied career mobility while also providing a tool to workers and urban planners trying to traverse the space of workplace skills.

## DISCUSSION

We can summarize the paper's argument as follows: Occupational polarization has been studied using broad subjective occupation categories (that is, cognitive or physical and routine or nonroutine) that fail to capture the dynamics of workplace skills and decreased labor mobility between low- and high-wage occupations. Rather than subjective occupational categories determined entirely by annual wages, we propose a purely data-driven methodology to map the space of workplace skills based on skill complementarity. The resulting network of skills is polarized in a way that respects stylized facts about occupational polarization; in particular, skill communities distinguish between occupations of different annual wages, thus demonstrating the direct connection between skill polarization and the hollowing of the middle class [see Figs. 2 (A and B) and 3].

Beyond the aggregate structure of the skill network (that is, node communities), we demonstrate that the raw topology of the network corresponds to pathways along which labor dynamics can occur; specifically, we find that the network proximity between skills predicts (i) skill adaptation in cities, (ii) skill redefinition of occupations, and (iii) the changing skill requirements of individual workers as they transition between occupations (see Fig. 4). Finally, by combining our observations of skill polarization with the labor dynamics determined by the network topology, we hypothesize that worker mobility between physical and cognitive occupations will be constrained, and we provide three types of supporting evidence: (i) Workers tend to transition between occupations relying on the same skill set, (ii) workers are unable to switch away from occupations relying equally on cognitive and physical labor, and (iii) this constraining effect is reflected in the national employment statistics (see Fig. 5). Interesting





**Fig. 5. The polarized skill network constrains worker mobility.** Binning by the  $\text{cognitive}_j$  of the worker's occupation in 2014 reveals the (A) expected cognitive change and the (B) expected magnitude of cognitive change when workers change occupations. Random occupation selection is considered as a null model (gray). SE bars are provided but are small. Actual occupational transitions are provided as examples in (A). (C) The national distribution of employment by  $\text{cognitive}_j$  with the distribution of individual occupations as an inset. (D) The average complementarity strength that skills possess in each skill category; this measure corresponds to worker mobility because skill proximity is indicative of skill acquisition.

future work might use older sources for skills data, such as the *Dictionary of Occupational Titles*, in combination with our methodology to examine the larger temporal dynamics of skill polarization and their consequences on labor.

While our methods provide more texture to changing labor demands, they have some limitations. First, while the O\*NET database facilitates the improved resolution of our model, the taxonomy of O\*NET skills may not capture the real-time dynamics of skill categories. For example, consider that a job listing for a software developer in the 1990s may only require “programming” skill, while modern listings might require specific types of programming skill, including proficiency in Hadoop, Java, or Python as examples. The O\*NET database may miss this change in skill specificity until the taxonomy of skill categories is explicitly updated. External data sources, such as LinkedIn, provide user-defined skills that may allow the future study of skill category dynamics—although these data suffer from being non-representative.

Second, our analysis provides evidence that cities, occupations, and individual workers leverage the complementarity between skills to navigate changing labor demands and to facilitate career mobility. While our methods provide a data-driven view of the structure underlying these dynamics, they do not account for general market equilibrium dynamics that accompany changing skill demands, and our results demonstrate the need for refined theoretical work that incorporates the granularity of specific workplace tasks and skills. For example, how

would the advent of new technology that performs a specific workplace skill change the skill network? And how does the relative cost of capital equipment play into decisions to retrain workers or purchase software or hardware? Answering these types of questions requires knowledge of other mechanisms, such as demand elasticity or capital availability, in addition to knowledge about the skill's location in the skill network. Nevertheless, we hope that our framework inspires further investigation into how skill structure dynamics interact with economic equilibrium dynamics studied in traditional models.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/7/eaao6030/DC1>

- Section S1. Exploring occupations and their constituent skills
- Section S2. Skill complementarity propensities and clusters
- Section S3. How educational requirements relate to skill requirements for occupations
- Section S4. Validating skill polarization
- Section S5. Projecting urban workforces onto the Skillscape
- Section S6. Predicting economic well-being with sociocognitive skills
- Section S7. Using Skillscape proximity to predict labor dynamics
- Fig. S1. Transforming raw O\*NET data with RCA.
- Fig. S2. Distribution of aggregate skill importance by summing the raw O\*NET values of each occupation.
- Fig. S3. Projecting occupational skill requirements onto the polarized skill network.
- Fig. S4. A comparison of the raw O\*NET data (left column) and the resulting Skillscape matrix (right column) for 2010, 2013, and 2015.



Fig. S5. The Skillscape network respects skill categorization from the experts.  
 Fig. S6. Complementarity scores for every individual skill (node in the network).  
 Fig. S7. The skill requirements of an occupation indicate the education required.  
 Fig. S8. Testing the significance of Skillscape polarization.  
 Fig. S9. Identifying the skill sets of urban workforces.  
 Fig. S10. Example cities projected onto the Skillscape according to the effective use of skills.  
 Fig. S11. Distribution of expected annual wages across occupations.  
 Fig. S12. Out-of-sample testing of model performance from Table 3.  
 Fig. S13. Out-of-sample testing of model performance from Table 4.  
 Fig. S14. Out-of-sample testing of model performance from Table 5.  
 Fig. S15. Out-of-sample testing of model performance from Table 6.  
 Fig. S16. A cartoon example of Area Under the Receiver Operating Characteristic curve (AUROC) calculation.  
 Fig. S17. Worker mobility and occupation redefinition are constrained by skill complementarity and polarization.  
 Fig. S18. Predicting changes in cognitive skill fraction of individual workers binning transitions by the magnitude of change.  
 Fig. S19. Predicting changes in cognitive skill fraction of individual workers binning transitions by their starting cognitive skill fraction.  
 Fig. S20. Predicting changes to the cognitive skill fraction of occupations.  
 Fig. S21. Predicting the effectively used skills of cities over time.  
 Fig. S22. Workers exhibit greater career mobility when leveraging exclusively sociocognitive or sensory-physical skills.  
 Fig. S23. Effects of randomly selecting cognitive skills as a null model alternative to Louvain community detection.  
 Fig. S24. Distribution of national employment and individual occupations as an inset, after binning by cognitive.  
 Fig. S25. Distribution of national employment in 2015 and individual occupations as an inset, after binning by cognitive, while varying the number of bins.  
 Fig. S26. Binning employment according to cognitive skill fraction reveals a trimodal distribution across cities of all sizes.  
 Fig. S27. Skill proximity predicts skill acquisition for individual workers transitioning between occupations, for the skill requirements of occupations, and for labor markets of cities.  
 Table S1. Skills comprising each skill community on the Skillscape.  
 Table S2. Descriptions of each occupation type indicator variable used in regression models.  
 Table S3. Linear regression using standardized cognitive<sub>c</sub> for each occupation and occupation type indicator variables.  
 Table S4. Linear regression using cognitive<sub>c</sub> and employment in each occupation with a bachelor's degree (denoted B.D. Employment) and without a bachelor's degree (denoted No B.D. Employment).  
 Table S5. Linear regression using standardized cognitive<sub>c</sub> for each city and employment in that city of each occupation type.  
 Table S6. Linear regression using cognitive<sub>c</sub> and education variables.

## REFERENCES AND NOTES

- R. Kochhar, R. Fry, M. Rohal, *The American Middle Class is Losing Ground* (Pew Research Center, 2015).
- R. Chetty, D. Grusky, M. Hell, N. Hendren, R. Manduca, J. Narang, "The fading American dream: Trends in absolute income mobility since 1940" (Technical Report, National Bureau of Economic Research, 2016).
- R. D. Putnam, *Our Kids: The American Dream in Crisis* (Simon and Schuster, 2016).
- H. B. Johnson, *The American Dream and the Power of Wealth: Choosing Schools and Inheriting Inequality in The Land of Opportunity* (Routledge, 2014).
- "America's shrinking middle class: A close look at changes within metropolitan areas" (Technical Report, Pew Research Center, 2016).
- D. H. Autor, D. Dorn, The growth of low-skill service jobs and the polarization of the US labor market. *Am. Econ. Rev.* **103**, 1553–1597 (2013).
- D. H. Autor, L. F. Katz, M. S. Kearney, Trends in U.S. wage inequality: Revising the revisionists. *Rev. Econ. Stat.* **90**, 300–323 (2008).
- D. Acemoglu, D. Autor, Skills, tasks and technologies: Implications for employment and earnings. *Handb. Labor Econ.* **4**, 1043–1171 (2011).
- A. Ebenstein, A. Harrison, M. McMillan, S. Phillips, Estimating the impact of trade and offshoring on American workers using the current population surveys. *Rev. Econ. Stat.* **96**, 581–595 (2014).
- F. MacCrory, G. Westerman, Y. Alhammadi, E. Brynjolfsson, Racing with and against the machine: Changes in occupational skill composition in an era of rapid technological advance, in *Proceedings of the International Conference on Information Systems—Building a Better World through Information Systems (ICIS)*, 2014, Auckland, New Zealand, 14 to 17 December 2017.
- D. H. Autor, Why are there still so many jobs? The history and future of workplace automation. *J. Econ. Perspect.* **29**, 3–30 (2015).
- J. E. Bessen, *How Computer Automation Affects Occupations: Technology, Jobs, and Skills* (Boston Univ. School of Law, Law and Economics Research Paper, 2015), pp. 15–49.
- D. Autor, *The Polarization of Job Opportunities in the US Labor Market: Implications for Employment and Earnings* (Center for American Progress and The Hamilton Project, 2010).
- E. Brynjolfsson, P. Milgrom, Complementarity in organizations, in *The Handbook of Organizational Economics*, R. Gibbons, J. Roberts, Eds. (Princeton Univ. Press, 2013), pp. 11–55.
- M. E. Porter, Clusters and the new economics of competition. *Harv. Bus. Rev.* **76**, 77–90 (1998).
- M. E. Porter, Location, competition, and economic development: Local clusters in a global economy. *Econ. Dev. Q.* **14**, 15–34 (2000).
- F. Neffke, M. Henning, R. Boschma, How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Econ. Geogr.* **87**, 237–265 (2011).
- F. Neffke, M. Henning, Skill relatedness and firm diversification. *Strat. Mgmt. J.* **34**, 297–316 (2013).
- B. C. Madrian, L. J. Lefgren, An approach to longitudinally matching current population survey (CPS) respondents. *J. Econ. Soc. Meas.* **26**, 31–62 (2000).
- C. A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann, The product space conditions the development of nations. *Science* **317**, 482–487 (2007).
- C. A. Hidalgo, R. Hausmann, The building blocks of economic complexity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10570–10575 (2009).
- R. Hausmann, C. A. Hidalgo, The network structure of economic output. *J. Econ. Growth* **16**, 309–342 (2011).
- E. L. Glaeser, H. D. Kallal, J. A. Scheinkman, A. Shleifer, Growth in cities. *J. Polit. Econ.* **100**, 1126–1152 (1992).
- A. M. Isserman, The location quotient approach to estimating regional economic impacts. *J. Am. Inst. Plann.* **43**, 33–41 (1977).
- S. T. Shuttters, R. Muneeppeerakul, J. Lobo, Constrained pathways to a creative urban economy. *Urban Stud.* **53**, 3439–3454 (2016).
- T. L. Vollrath, A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage. *Weltwirtsch. Arch.* **127**, 265–280 (1991).
- F. Neffke, M. S. Henning, Revealed relatedness: Mapping industry space (Papers in Evolutionary Economic Geography 8:19, Urban and Regional Research Centre Utrecht, Utrecht Univ., 2008).
- V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
- M. Rubinov, O. Sporns, Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
- O. Sporns, R. F. Betzel, Modular brain networks. *Annu. Rev. Psychol.* **67**, 613–640 (2016).
- M. Barthélemy, Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
- M. Berest, R. Gera, Z. Lukens, N. Martinez, B. McCaleb, Predicting network evolution through temporal Twitter snapshots for Paris attacks of 2015, in *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation*, 2016.
- T. M. Devinney, J. Hohberger, The past is prologue: Moving on from *Cultures Consequences*. *J. Int. Business Stud.* **48**, 68–82 (2017).
- J. Fan, J. Meng, Y. Ashkenazy, S. Havlin, H. J. Schellnhuber, Network analysis reveals strongly localized impacts of El Niño. *Proc. Natl. Acad. Sci. U.S.A.* **201701214** (2017).
- Y. Cohen, D. Hendler, A. Rubin, Detection of malicious webmail attachments based on propagation patterns. *Knowl. Based Syst.* **141**, 67–79 (2018).
- C. Gathmann, U. Schönberg, How general is human capital? A task-based approach. *J. Labor Econ.* **28**, 1–49 (2010).
- A. C. Kerckhoff, Education and social stratification processes in comparative perspective. *Sociol. Educ.* **74**, 3–18 (2001).
- D. H. Autor, F. Levy, R. J. Murnane, The skill content of recent technological change: An empirical exploration. *Q. J. Econ.* **118**, 1279–1333 (2003).

## Acknowledgments

**Funding:** This work was supported by the Center for Complex Engineering Systems at King Abdulaziz City for Science and Technology (KACST), Massachusetts Institute of Technology, the Siegel Family Endowment, and the Ethics and Governance of AI Fund. **Author contributions:** A.A., M.R.F., and L.S. performed the calculations. A.A. and M.R.F. produced the figures. A.A., B.A., and L.S. constructed the online data visualization. A.A., M.R.F., I.R., and C.H. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** We provide an online interactive tool for exploring occupations and urban workforces on the Skillscape at [skillscape.mit.edu](https://skillscape.mit.edu) (password: workforce). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 8 August 2017

Accepted 11 June 2018

Published 18 July 2018

10.1126/sciadv.aao6030

**Citation:** A. Alabdulkareem, M. R. Frank, L. Sun, B. AIShebli, C. Hidalgo, I. Rahwan, Unpacking the polarization of workplace skills. *Sci. Adv.* **4**, eaao6030 (2018).