Small cities face greater impact from automation

Morgan R. Frank,[1] Lijun Sun,[1] Manuel Cebrian,[2] HyeJin Youn,[3] Iyad Rahwan[1*]

[1]Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Data61 Unit, Commonwealth Scientific and Industrial Research Organization, Melbourne, Victoria, Australia
[3]Institute for New Economic Thinking, University of Oxford, Oxford, UK

[*]To whom correspondence should be addressed; E-mail: irahwan@mit.edu.

The city has proven to be the most pervasive form of human agglomeration providing wide employment opportunities for its dwellers. As advances in robotics and Artificial Intelligence revive concerns about the impact of automation on jobs, a question looms: How will automation affect employment in cities? Here, we produce a comparative picture of the impact of automation across US urban areas revealing that small cities will undertake greater adjustments, such as worker displacement, and job content substitutions. To understand this trend, we demonstrate that large cities exhibit increased occupational and skill specialization, which allows us to better predict job impact from automation using a regression model. This increased specialization in large cities is due to an increased abundance of managerial jobs and specialized occupations, which require workers with technical skills who are prepared to use and improve cutting-edge technology. Our study provides the first empirical law connecting two societal forces: urban agglomeration and automation's impact on employment.

With over half of the world's population residing in cities (*1*), urbanization and humanity's transition to mega-cities (*2–4*) highlight the need for sustainable urban ecosystems. Despite their complexity, cities have revealed pervasive scaling relationships through the growing urban physics literature (*5*). In particular, urban scaling has revealed benefits to urbanization, such as increased innovation rates, returns to scale (*6, 7*), and increased labor diversity (*8*). However, these studies also reveal the negative externalities of urbanization, such as higher pollution levels, increased rates of contagious diseases, increased crime (*7, 9*), and lower fertility rates (*10*). In light of these trends, city, state, and federal governments must effectively balance the benefits and drawbacks of urbanization.

Job migration is a leading factor in urbanization (*1, 11*), and policy makers are increasingly concerned about the impact of automation on job availability in cities (*12, 13*). For example, building on assessments from machine learning experts, Frey and Osborne's model (*14*) estimates that 47% of U.S. employment is at "high risk of computerization" in the foreseeable future, while an alternative OECD study employing a task-based methodology concludes a more modest 9% of employment is at risk (*15*). The seminal work of Autor et al (*16, 17*) and Brynjolfsson and McAfee (*18*) highlight the two main effects of automation on workplace skills: substitution of routine skills, and complementarity of non-routine and communication skills. Additionally, there are dynamics that may alter demand for labor as a result of increased efficiency. Our purpose here is not to foresee those response dynamics; rather, our aim is to quantify the magnitude of the shocks across U.S. metropolitan areas. In particular, we strive to identify those urban features that may moderate or exacerbate the labor impact of automation. It is not clear *a priori* how cities, as adaptive complex systems (*19,20*), naturally respond to this threat.

Competing theories exist for the economic sustainability of cities. An influx of occupational diversity may explain the wealth-creation, innovation, and success of cities (*9, 21–23*), and is evidenced by the increased number of industries (*8*) and increased number of unique job types in large cities (Fig. S1B). On the other hand, cities connect people with greater efficiency (*9, 24*) which enables a division of labor that increases overall productivity (*25–27*) through occupational specialization. As evidence, R&D employment has been shown to scale superlinearly with city size (*7*). However, the division of labor can be a conduit for automation as it identifies routine tasks and encourages worker modularity. These observations pose a puzzle: are the forces of diversity, specialization, and the division of labor shaping a city's ability to accommodate automation?

We estimate the expected impact on jobs from automation in cities (i.e. U.S. Census Metropolitan Statistical Areas) according to

$$E_m = \frac{\sum_{j \in Jobs_m} p_{auto}(j) \cdot f_m(j)}{\sum_{j \in Jobs} f_m(j)} \tag{1}$$

where $Jobs_m$ denotes the set of job types in city m, $f_m(j)$ denotes the number of workers in city m with job j according to the U.S. Bureau of Labor Statistics (BLS), and $p_{auto}(j)$ denotes the probability of computerization for job j according to (*14*) (see S3 for more details). Each city should expect between one-half and three-quarters of their current jobs to be affected in the foreseeable future due to improvements in automation (see Fig. 1A. Also note that this estimate differs from (*14*) which focused on national statistics). While this calculation omits potential job creation or job redefinition which typically accompany innovation (*28, 29*), it highlights the differential impact of automation across cities and smooths potential noise in the predicted automation of individual jobs.

What differentiates cities' resilience to automation? Figure 1B demonstrates that expected job impact decreases according to $E_m \propto -3.2 \times \log_{10}(\text{city size})$, which suggests that larger cities are more resilient to the negative effects of automation. This relationship is significant with a Pearson correlation $\rho = -0.53$ ($p_{val} < 10^{-28}$), and shows that laborers in smaller cities are susceptible to the impact of automated methods ($R^2 = 0.28$). We confirm our finding using separate conservative skill-based estimates of the automatability of jobs (*15*) (Pearson $\rho = -0.26$ ($p_{val} < 10^{-7}$) and $E_m \propto -1.24 \times \log_{10}(\text{city size})$. See Fig.1B inset and S3.1). Despite the conservative nature of these alternative probabilities, we observe increased resilience with city size.

To identify the mechanisms underpinning resilience to automation, we examine the most distinctive characteristics of urban economies: diversification and specialization. Since automation typically targets workplace skills (*15*), we include the O*NET skill dataset in our examination. The U.S. BLS conducts annual surveys measuring the absolute importance of various workplace skills to the completion of each BLS job. We use normalized Shannon entropy, an information theoretic measure for distributions (see S2), to measure the specialization of the job distributions in city m according to

$$H_{job}(m) = -\sum_{j \in Jobs_m} p_m(j) \frac{\log(p_m(j))}{\log(\mid Jobs_m \mid)} \tag{2}$$

and to measure the distribution of skills in city m according to

$$H_{skill}(m) = -\sum_{s \in Skills_m} p_m(s) \frac{\log(p_m(s))}{\log(\mid Skills_m \mid)}, \tag{3}$$

where $p_m(j)$ is the relative abundance of job j in city m, $p_m(s)$ is the relative abundance of skill s in city m, and $Skills_m$ is the set of skills in city m. Normalizing the canonical Shannon entropy definition by the maximum possible Shannon entropy given the number of unique occupations or skills in a city (i.e. $\log(|Jobs_m|)$ or $\log(|Skills_m|)$, respectively) allows us to control for the size effects on the number of different occupations and skills present in a city (see S2 for complete details). Specialization (i.e. descreased Shannon entropy) appears in the distributions of jobs in cities (Fig. 2A) and, separately, through the aggregate distributions of skills in cities

(Fig. 2B). Additionally, we use Theil entropy to measure the proportion of specialized jobs (in terms of skills) in comparison to the skill specialization of the city on whole. Theil entropy for city m is calculated according to

$$T_m = \sum_{j \in Jobs_m} p_m(j) \cdot \frac{H_{skill}(m) - H_j}{H_{skill}(m)}, \tag{4}$$

where H_j is the Shannon entropy of the distribution of skills for job j according to O*NET data. We find an increased proportion of specialized jobs in large cities (i.e. $1 - T_m$ decreases, Fig. 2C).

Using these specialization measures in three separate linear regression models reveals that skill specialization is predictive of expected job impact in cities ($R^2 = 0.20$, see S4). Multiple linear regression produces the most predictive model accounting for 66% of the variance across cities (see Fig. 2D). This model relies most strongly on skill specialization while controlling for several generic urban factors, such as per capita GDP, city size, and education levels, and improves on the base model using only these generic urban factors without specialization measures ($R^2 = 0.53$). We confirm the stability of our regression results by alternatively training the regression model on half of the cities and measuring the performance of the regression on the remaining cities as validation (see S4). The residuals between the actual and modelled values highlight notably resilient cities (given the model), such as Boulder, C.O. and Warner Robins, G.A., and notably susceptible cities, such as Napa, C.A. and Carson City, N.V. (see Fig. 2). Examining these cities more closely may allow urban policy experts with a nuanced understanding of the policies in these cities to more easily identify causal mechanisms. The predictive power of this model and its reliance on workplace skills justifies our inclusion of skills data in addition to occupation data, and motivates us to characterize urban resilience to automation from the skills in cities.

The juxtaposition of evidence for both diversity and specialization in large cities is reconcilable through the division of labor theory (*25*). Specifically, we find that average number of workers per firm increases logarithmically with city size (see S1). Under the division of labor argument, large firms have better ability to support specialized workers along with the management required to coordinate them (*30*). At the same time, workers possessing specialized skills seek specific employment opportunities which maximize their financial return (*31, 32*). The demand for specific specialized jobs increases occupational specialization while also increasing the number of unique job types and industries in a city (*11*).

What do cities specialize in and why? Specialization alone is not enough to explain the resilience we observe across cities. For example, Detroit, which is famous for its specialization in automotive manufacturing, has experienced economic down turn (*33*), while the San Francisco Bay area, epicenter of the information technology industry, continues to flourish despite the dot-com bubble perhaps due to its support of a "creative class" of workers (*34*).

Our analysis reveals which skills receive greater attention when occupational specialization is observed in cities. In particular, given observations of resilience and observations of specialization, we want to know how managerial jobs and highly technical jobs change with city size according to urban scaling (*35*). We use K-means clustering algorithm (i.e. occupations are instances and raw O*NET skill importance are features) to identify five clusters of jobs according to skill similarity (see S6.3) and examine the scaling relationship between job clusters and city size according to (number of workers) $\propto$ (city size)$^{\beta}$ in Figure 3A. Note that the exponent, β, entirely describes the growth rate of these job clusters relative to city size. The job cluster comprised of highly specialized jobs, such as Mathematician and Chemist, exhibits a notably superlinear scaling relationship with city size ($\beta = 1.39$). This scaling exponent is similar to the scaling relationship observed for *Private R&D employment* ($\beta = 1.34$) found in (*7*) and is in good agreement with similar studies on job growth (*36*). Furthermore, our finding of one job cluster exhibiting notably larger scaling than the other job clusters is stable to sub-sampling occupations at various rates (see S6.3.2). Managerial jobs also grow superlinearly, but to a weaker extent ($\beta = 1.08$). The job cluster exhibiting the slowest growth ($\beta = 0.94$) is comprised of entertainment and service jobs.

We confirm that the fastest growing job cluster is indeed comprised of "technical" jobs from their constituent workplace skills. We employ K-means clustering (i.e. O*NET skills are instances and the correlation of raw O*NET importance of skills across occupations are features) to simplify the complete space of O*NET skills to ten skill types based on the co-occurrence of skills across jobs (see S5, and S6.6 for complete description of skill clusters). These simplified skill types allow us to intuitively explore which skills indicate specialization or indicate resilience in cities. Computational/Analytical skills and Management skills are more likely in faster growing (i.e. superlinear) jobs, while physical skills, such as Physical Coordination and Control/Perceptual skills, indicate notably slower job growth with city size (Fig. 3B). We confirm our findings using alternative definitions for aggregate workplace tasks and skills (see S5).

The skills which are relied on by fast-growing technical jobs suggest mechanisms for resilience and growth in cities. Existing work (*37*) identifies that individual workers can gain skills to compete with automation, gain skills to complement automation, or seek industries removed from the impacts of automation. Similar to individual workers, the division of labor in large cities allows them to specialize in skills removed from the threat of automation. Computational/Analytical, Managerial, Organization, and Relational skills are more likely to be present in specialized and resilient cities (Fig. 4A&C), while Physical Coordination and Control/Perceptual skills indicate both decreased specialization and decreased resilience in cities (Fig. 4B&D). We confirm our results using alternative groups of workplace tasks (*38*) provided by O*NET (see S5.1) and again by examining the routineness of workplace tasks (*16*) (see S5.2). Figure 4E reflects the same conclusion by comparing the relationship of each skill type to city size (right column) and expected job impact (middle column) (see S6.5 for comparison with raw O*NET skills). Effectively, large cities employ workers whose skills better prepare them to interface with automation technology, while small cities rely more prominently on physical workers, who are more susceptible to automation.

Despite being seemingly unrelated societal forces, we uncover a positive interplay between urbanization and automation. Larger cities not only tend to be more innovative (*6, 7*), but also harbor the workers who are prepared to both use and improve cutting-edge technology. In turn, these workers are more specialized in their workplace skills and less likely to be replaced by automated methods in the foreseeable future. It is difficult to concretely identify causal mechanisms at the scale of this investigation, but, despite this difficulty, we highlight evidence for the division of labor in large cities and show its importance as a piece of the automation and urbanization puzzle. These findings open the door for more controlled investigations with input from policy makers.

It is important to put our results in context. By quantifying relative *impact*, we do not make any explicit predictions about relative *technological unemployment* in cities. Historical effects on total employment have been driven by systemic effects, rather than solely through job-level effects (*17*). For example, the introduction of Automated Teller Machines (ATMs) suggested a likely decrease in human bank teller employment. However, contrary to this prediction, ATM technology cut the cost to banks for opening and operating new branches. This shifted the supply curve outwards, leading banks to open more branches at market equilibrium. The result was an *increase* in bank teller employment (*36, 39*), but those tellers performed different tasks, such as relationship management and investment advice, which required very different skills. Hence, by *impact*, we refer to the *magnitude* of the skill substitution shocks that cities must respond to. The actual response will be shaped both by free market dynamics (e.g. shifts in supply and demand curves) and by economic and educational policy (e.g. worker re-training, or skilled migration). Still, recognizing that small cities will experience larger adjustments to automation calls on policy-makers to pay special attention to the pronounced risks we have identified.

References and Notes

1. I. O. for Migration, World migration report 2015 - migrants and cities: New partnerships to manage mobility (2015).

2. F. Kraas, S. Aggarwal, M. Coy, G. Mertins, *Megacities: our global urban future* (Springer Science & Business Media, 2013).

3. C. Ash, B. R. Jasny, L. Roberts, R. Stone, A. M. Sugden, *Science* **319**, 739 (2008).

4. M. R. Montgomery, *Science* **319**, 761 (2008).

5. M. Batty, *Science* **319**, 769 (2008).

6. L. M. A. Bettencourt, J. Lobo, D. Strumsky, G. B. West, *PLoS ONE* **5**, e13541 (2010).

7. L. M. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, *Proceedings of the National Academy of Sciences* **104**, 7301 (2007).

8. H. Youn, *et al.*, *Journal of The Royal Society Interface* **13**, 20150937 (2016).

9. W. Pan, G. Ghoshal, C. Krumme, M. Cebrian, A. Pentland, *Nature communications* **4** (2013).

10. R. Mace, *Science* **319**, 764 (2008).

11. C. Rozenblat, D. Pumain, *Cities in Globalization: Practices, Policies, and Theories* (Routledge, 2007), pp. 130–156.

12. B. Carlsson, *Technological systems and economic performance: the case of factory automation*, vol. 5 (Springer Science & Business Media, 2012).

13. M. Olsen, D. Hemous, *2014 Meeting Papers*, no. 162 (Society for Economic Dynamics, 2014).

14. C. B. Frey, M. A. Osborne, *Technological Forecasting and Social Change* (2016).

15. M. Arntz, T. Gregory, U. Zierahn (2016).

16. D. H. Autor, F. Levy, R. J. Murnane, The skill content of recent technological change: An empirical exploration, *Tech. rep.*, National Bureau of Economic Research (2001).

17. D. H. Autor, *The Journal of Economic Perspectives* **29**, 3 (2015).

18. E. Brynjolfsson, A. McAfee, *The second machine age: work, progress, and prosperity in a time of brilliant technologies* (WW Norton & Company, 2014).

19. R. Albert, H. Jeong, A.-L. Barabási, *Nature* **406**, 378 (2000).

20. P. Holme, B. J. Kim, C. N. Yoon, S. K. Han, *Physical Review E* **65**, 056109 (2002).

21. E. Glaeser, *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier* (Penguin, 2011).

22. J. M. Quigley, *The Journal of Economic Perspectives* **12**, 127 (1998).

23. J. V. Henderson, *et al.*, *OUP Catalogue* (1991).

24. A. Sim, S. N. Yaliraki, M. Barahona, M. P. Stumpf, *Journal of The Royal Society Interface* **12**, 20150315 (2015).

25. A. Smith (1976).

26. L. M. Bettencourt, H. Samaniego, H. Youn, *Scientific Reports* **4** (2014).

27. L. Sveikauskas, *The Quarterly Journal of Economics* pp. 393–413 (1975).

28. M. Mazzucato, *Industrial and Corporate Change* p. dtt025 (2013).

29. D. Archibugi, A. Filippetti, M. Frenz, *Research Policy* **42**, 303 (2013).

30. R. H. Coase, *economica* **4**, 386 (1937).

31. D. E. Bloom, D. Canning, G. Fink, *Science* **319**, 772 (2008).

32. M. Schich, *et al.*, *Science* **345**, 558 (2014).

33. T. Klier, *Economic Perspectives* **33** (2009).

34. R. Florida, *The rise of the creative class* (Basic books New York, 2004).

35. D. Pumain, *Santa Fe Working Paper* (2004).

36. J. E. Bessen, *Boston Univ. School of Law, Law and Economics Research Paper* (2015).

37. F. MacCrory, G. Westerman, Y. Alhammadi, E. Brynjolfsson, *International Conference on Information Systems (ICIS)* (2014).

38. S. Kok, B. t. Weel, *Journal of Regional Science* **54**, 856 (2014).

39. J. Bessen, *Learning by doing: the real connection between innovation, wages, and wealth* (Yale University Press, 2015).

Acknowledgements

The authors would like to thank David Autor and Lorenzo Coviello for their feedback during the undertaking of this study.

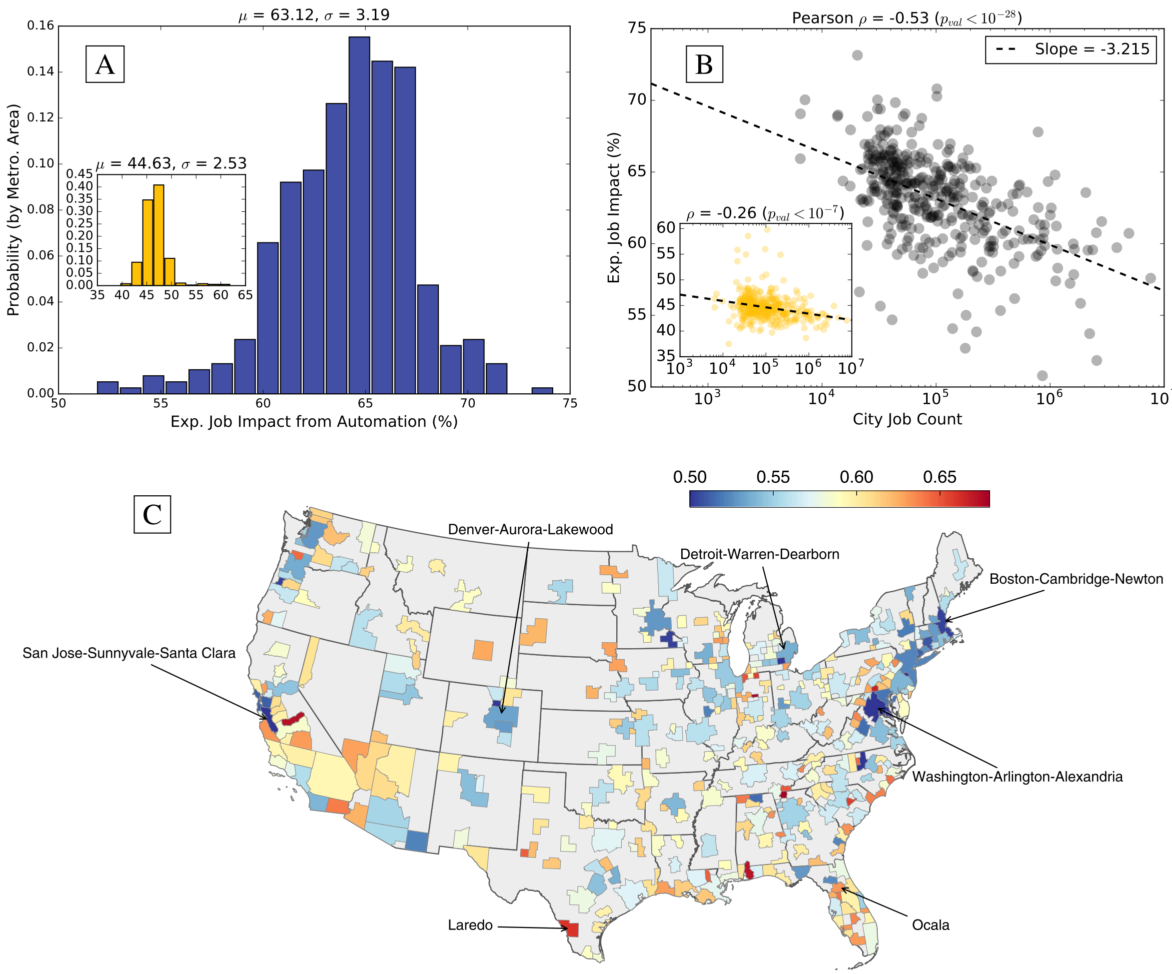


Figure 1: The impact of automation in U.S. cities. **(A)** The distribution of expected job impact (E_m) from automation across U.S. cities using predictions from (*14*). (Inset) The distribution of expected job impact using alternative conservative estimates from (*15*). **(B)** Expected job impact from automation decreases logarithmically with city size using predictions from (*14*). We provide the line of best fit (Slope $= -3.215$) and use Pearson correlation to demonstrate significance. (Inset) Decreased expected job impact with increased city size is again observed when using an alternative conservative prediction of job automatability produced in (*15*) (best fit line has slope -1.24, Pearson $\rho = -0.26$, $p_{val} < 10^{-7}$). **(C)** A map of U.S. metropolitan statistical areas colored according to expected job impact from automation.

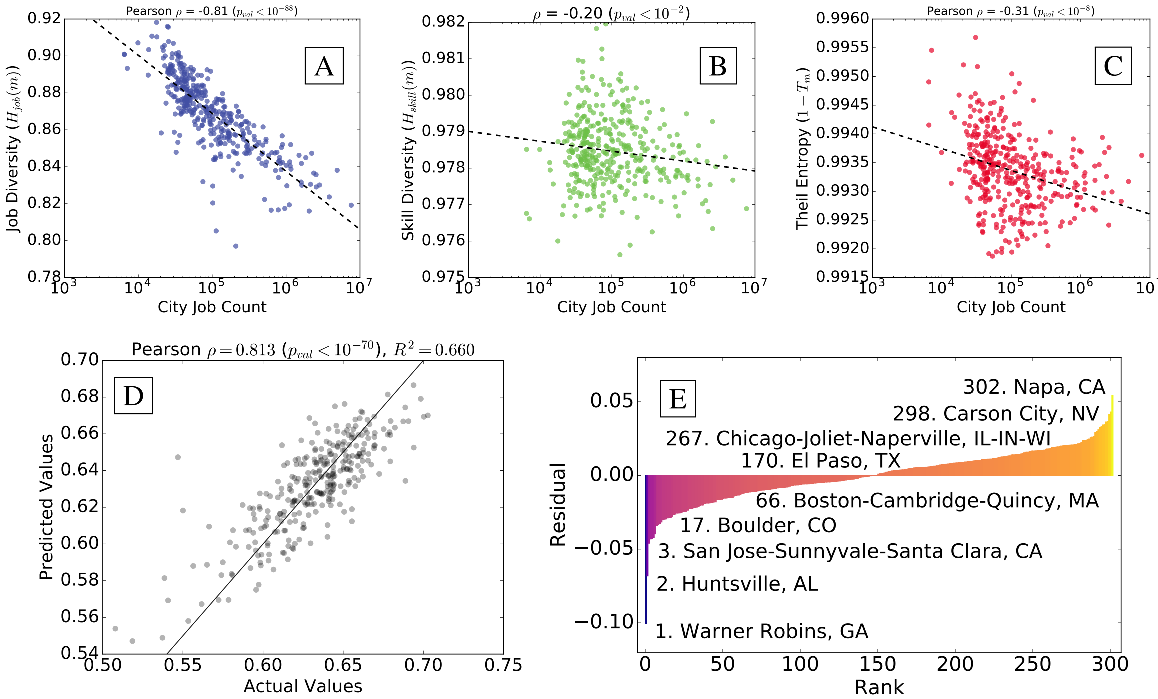


Figure 2: Large cities reveal increased occupational specialization through both job and skill distributions. **(A)** Shannon entropy of job distributions, $H_{job}(m)$, decreases with city size. **(B)** Shannon entropy of the O*NET skill distributions, $H_{skill}(m)$, decreases with city size. **(C)** Theil entropy, T_m, reveals the proportion of specialized jobs increases with city size. For plots (A), (B), & (C), we provide the line of best fit for reference. **(D)** A multiple linear regression model utilizing measures for labor specialization accounts for 66% of the variance in expected job impact from automation across U.S. cities (see S4 for details). **(E)** The distribution of residuals between the actual and predicted values of the model, and the rank of some example cities.

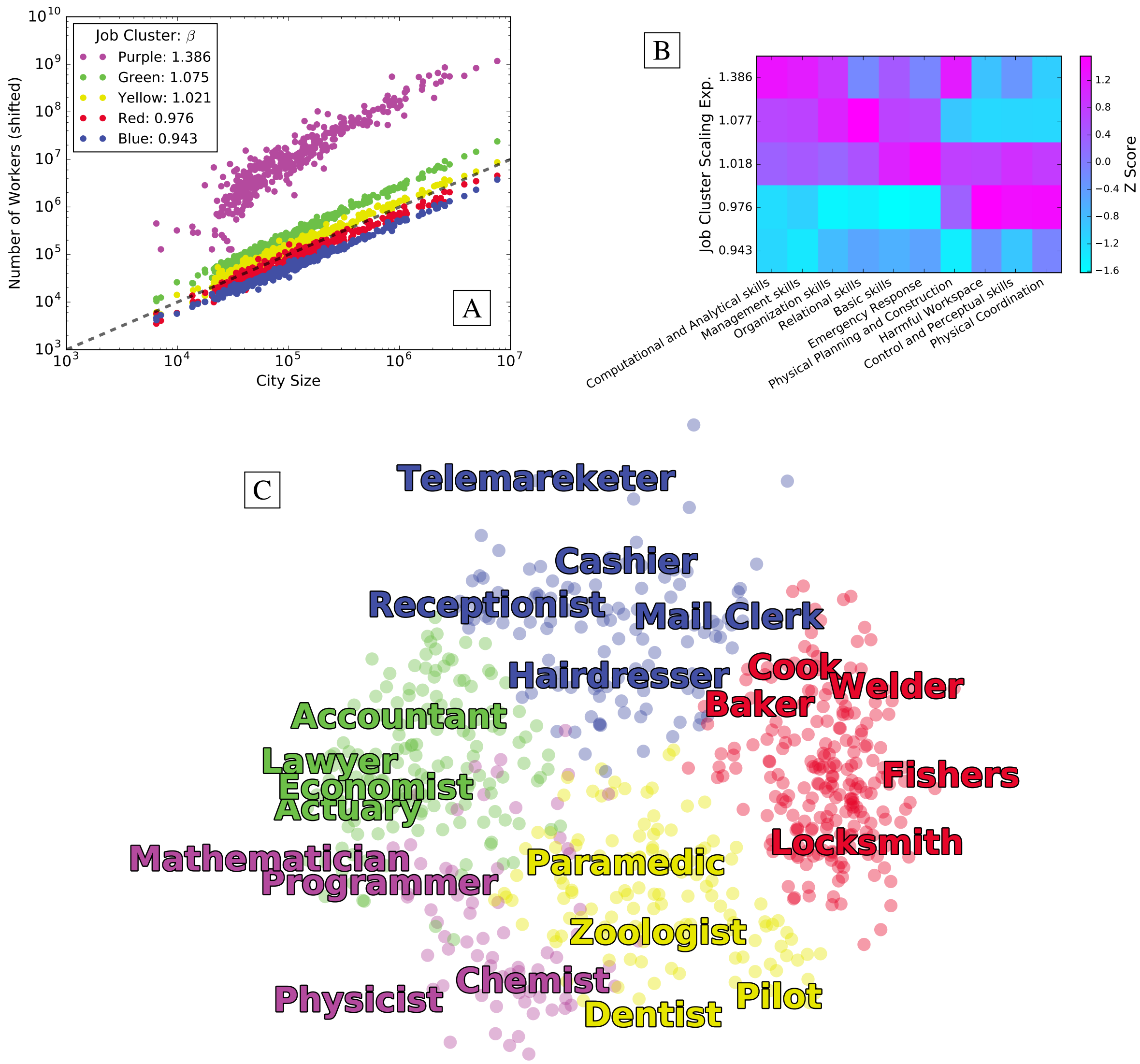
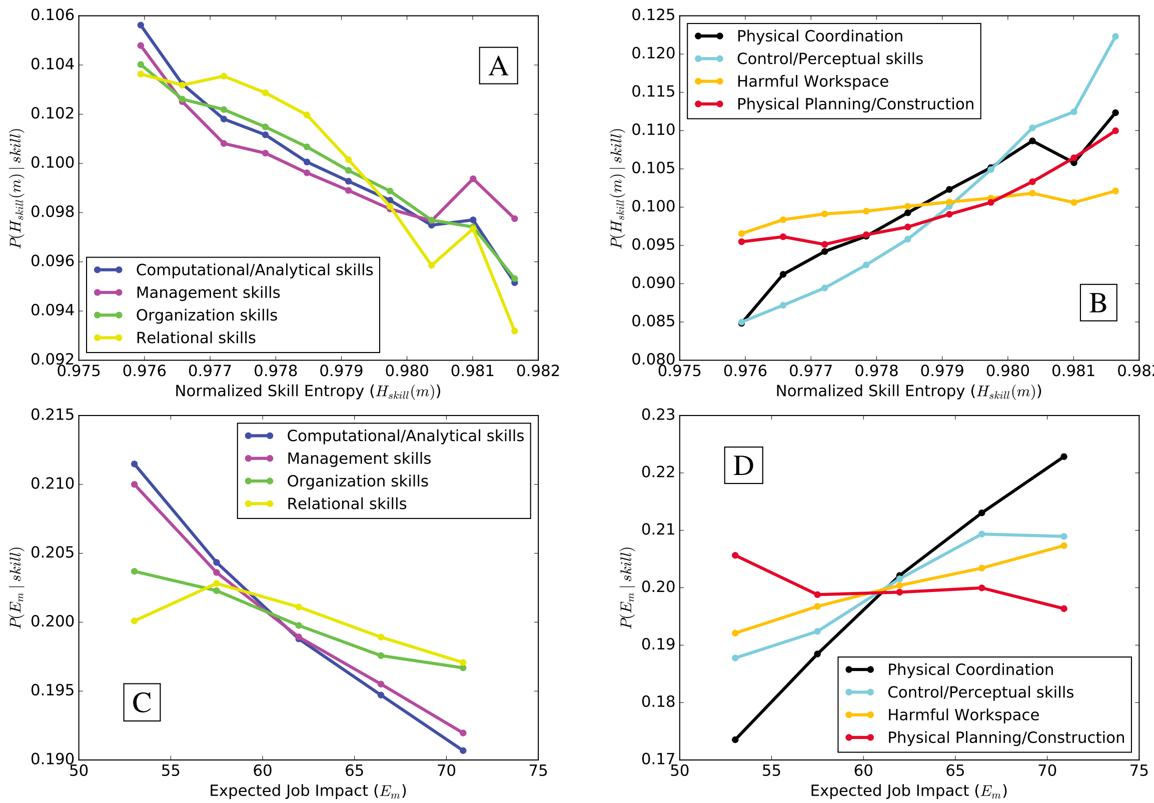


Figure 3: Scaling relationships for collections of similar jobs and their associated skills. **(A)** For each job cluster (color), we plot the total number of workers of jobs in that cluster (y-axis) versus the total number of workers in that city (x-axis). The black dashed line has a slope of 1 for reference. Each cluster is vertically shifted according to a linear fit in log scale to highlight the scaling difference. **(B)** By summing the importance of each skill type to each job cluster, we assess how strongly those skills indicate a scaling relationship according to its z score. For a given skill type, z scores are calculated according to the distribution of importance across job clusters. **(C)** We use principal component analysis to project jobs from the space of all skills to a 2-D plane. Example jobs from each cluster (indicated by color) are highlighted. The complete list of BLS jobs comprising each cluster is presented in S6.3.



E

Skill Type	Job Impact Corr.	Log$_{10}$ City Size Corr.
Computational/ Analytical	-0.88 ($< 10^{-124}$)	0.58 ($< 10^{-34}$)
Management	-0.87 ($< 10^{-120}$)	0.52 ($< 10^{-27}$)
Organization	-0.62 ($< 10^{-41}$)	0.35 ($< 10^{-11}$)
Relationship	-0.26 ($< 10^{-6}$)	-0.06 (0.3)
Physical Planning	-0.07 (0.21)	0.18 (0.0006)
Basic/General	0.13 (0.01)	-0.24 ($< 10^{-5}$)
Control & Perceptual	0.45 ($< 10^{-19}$)	-0.14 (0.005)
Emergency Response	0.46 ($< 10^{-20}$)	-0.32 ($< 10^{-9}$)
Physical Coordination	0.83 ($< 10^{-99}$)	-0.52 ($< 10^{-26}$)
Harmful Workspace	0.90 ($< 10^{-140}$)	-0.54 ($< 10^{-28}$)

Figure 4: Understanding occupational specialization and expected job impact through workplace skills. **(A) & (B)** Revealed skill importance as a function of skill specialization ($H_{skill}(m)$). Skills in (A) are more likely in cities exhibiting occupational specialization, while skills in (B) are more likely to co-occur with occupational diversity. **(C) & (D)** Revealed skill importance as a function of expected job impact (E_m). Skills in (C) are more likely in cities exhibiting resilience, while skills in (D) co-occur with increased job impact from automation. **(E)** The Pearson correlation of the proportion of each skill type to the expected job impact and to $\log_{10}$ city size with p-values in parentheses. A similar table for the raw O*NET skills is presented in S6.5 .