Author for correspondence:
Iyad Rahwan
e-mail: irahwan@mit.edu

THE ROYAL SOCIETY
PUBLISHING

Small cities face greater impact from automation

Morgan R. Frank[1], Lijun Sun[1], Manuel Cebrian[1,3], Hyejin Youn[4,5,6] and Iyad Rahwan[1,2]

[1]Media Laboratory, and [2]Institute for Data, Systems, & Society, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Data61 Unit, Commonwealth Scientific and Industrial Research Organization, Melbourne, Victoria, Australia
[4]Kellogg School of Management, Northwestern University, Evanston, IL, USA
[5]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208, USA
[6]London Mathematical Lab, London WC2N 6DF, UK

MRF, 0000-0001-9487-9359; HY, 0000-0002-6190-4412

The city has proved to be the most successful form of human agglomeration and provides 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 provide a comparative picture of the impact of automation across US urban areas. Small cities will undertake greater adjustments, such as worker displacement and job content substitutions. We demonstrate that large cities exhibit increased occupational and skill specialization due to increased abundance of managerial and technical professions. These occupations are not easily automatable, and, thus, reduce the potential impact of automation in large cities. Our results pass several robustness checks including potential errors in the estimation of occupational automation and subsampling of occupations. Our study provides the first empirical law connecting two societal forces: urban agglomeration and automation's impact on employment.

1. Introduction

Cities, which accommodate over half of the world's population [1], are modern society's hubs for economic productivity [2–4] and innovation [5–7]. As job migration is the leading factor in urbanization [1,8], policymakers are increasingly concerned about the impact of artificial intelligence and automation on employment in cities [9–11]. While researchers have investigated automation in national economies and individual employment, it remains unclear *a priori* how cities naturally respond to this threat. In a world struggling between localism and globalism, a question emerges: *how will different cities cope with automation?* Answering this question has implications on everything from urban migration to investment, and from social welfare policy to educational initiatives.

To construct a comparative picture of automation in cities, our first challenge is to get reliable estimates of how automation impacts workers. Existing estimates are wide ranging. Frey & Osborne [12] estimate that 47% of US employment is at 'high risk of computerization' in the foreseeable future, while an alternative OECD study concludes a more modest 9% of employment is at risk [13]. Note that these results do not tell us about the impact of automation in cities as they are presented at a national level. Differences in these predictions arise from discrepancies over two main skill dynamics: the substitution of routine skills, and complementarity of non-routine and communication skills [14–16]. Additionally, technology-driven efficiency may redefine the skill requirements of occupations and actually increase employment in low-skilled jobs [17,18].

Nevertheless, even if we take current estimates of the *absolute* risk of computerization of jobs with skepticism, these estimates can provide useful guidance about *relative* risk to different cities that is robust to errors in the estimates provided by Frey & Osborne [12] and Arntz *et al.* [13]. We can interpret the 'risk of computerization' estimates as an educated guess about which occupations will experience greater adjustment due to machine substitution of a large portion of their content. These adjustments represent a significant cost to an urban system from both technological unemployment and expensive worker retraining programmes.

A priori, it is not obvious whether large cities will experience more or less impact from automation. On one hand, an influx of occupational diversity explains the wealth creation, innovation and success of cities [19–22]. On the other hand, cities connect people with greater efficiency [22,23]. This enables a greater division of labour that increases overall productivity [24–26] through occupational specialization. However, the division of labour may facilitate automation as it identifies routine tasks and encourages worker modularity. If these modular jobs are at greater risk of computerization, then more workers may be impacted by automation in large cities. These observations pose a puzzle: *are the forces of diversity, specialization and the division of labour shaping a city's ability to accommodate automation?*

Here, we undertake a comparative examination of cities while measuring the relative impact of automation on employment. We also contextualize these measurements through a detailed analysis of the skill composition of different cities. Note that *impact* includes unemployment, but may also manifest itself through the changing skill demands of occupations as automation diminishes the need for individual types of skills [17,18]. In the light of imminent automation technology, we highlight a complicated relationship between labour diversity and specialization in cities, and discover that small cities are susceptible to the negative impact of automation.

2. Material and methods

2.1. Datasets

The US Bureau of Labor Statistics (BLS) data identify the employment distribution of about 700 different occupations across each of 380 US metropolitan statistical areas (MSAs) and combined statistical areas (CSAs) in 2014. (We refer to both CSAs and MSAs as 'cities'.) We consider MSAs in isolation only when they are not part of a CSA. CSAs have arisen as the best approximation for determining cities [5,6,27–31]. The resulting list of occupations considered in this study represents 99.99% of national employment according to the occupational employment statistics data produced annually by BLS. From these employment distributions, we calculate the probability of a worker in city m having job j according to

$$p_m(j) = \frac{f_m(j)}{\sum_{j \in \text{Jobs}_m} f_m(j)}, \qquad (2.1)$$

where Jobs_m denotes the set of job types in city m according to BLS data, and $f_m(j)$ denotes the number of workers in city m with job j.

For each occupation, the BLS O*NET dataset details the importance of 230 different workplace skills, such as Manual Dexterity, Finger Dexterity, Complex Problem Solving, Time Management and Negotiation. BLS obtains this information through several separate surveys which group the raw O*NET

skills into the following categories: Abilities, Education/Training/Experience, Interests Knowledge, Skills, Work Activities and Work Context. We normalize the raw survey responses to obtain a value between 0 (irrelevant to the occupation) and 1 (essential to the occupation) indicating the absolute importance of that skill to that occupation. We refer to these values of skill importance as *raw skill values*.

2.2. Measures for specialization and diversity

We assess the specialization or diversity of the employment distribution in city m by calculating the normalized Shannon entropy. Shannon entropy [32], an information-theoretic measure for the expected information in a distribution, can be normalized according to

$$H_{\text{job}}(m) = - \sum_{j \in \text{Jobs}_m} p_m(j) \times \frac{\log(p_m(j))}{\log(|\text{Jobs}_m|)}. \qquad (2.2)$$

This quantity measures the predictability of an employment distribution given the set of unique occupations in a city. The measure is maximized when the distribution is least predictable (i.e. the distribution is uniform). Therefore, the denominator of $\log(|\text{Jobs}_m|)$ normalizes the entropy score so that we can compare the distributions of jobs in cities with different sets of job categories (see electronic supplementary material, S2.1 for further discussion). The values for normalized Shannon entropy lie between 0 (specialization) and 1 (diversity). Normalized Shannon entropy has been used in a variety of fields, including virology [33], climatology [34] and city science [35].

For a given occupation, we normalize each raw skill value by the sum of the values to obtain the relative importance of each skill to that occupation (denoted $p_j(s)$). Similarly to above, we measure the normalized Shannon entropy of the relative skill distribution of job j according to

$$H_j = - \sum_{s \in \text{Skills}_j} p_j(s) \times \frac{\log(p_j(s))}{\log(|\text{Skills}_j|)}, \qquad (2.3)$$

where Skills_j denotes the set of O*NET skills with non-zero importance to job j. We employ normalized Shannon entropy here to facilitate a fair comparison of relative skill distributions between jobs which may have received the same raw O*NET value for a given skill, but have different numbers of non-zero raw O*NET skills.

We obtain a distribution of relative skill importance for a city according to

$$p_m(s) = \sum_{j \in \text{Jobs}_m} p_j(s) \cdot p_m(j), \qquad (2.4)$$

where $p_m(s)$ is the relative importance of skill s in city m. Again, we use normalized Shannon entropy to assess the skill specialization in a city according to
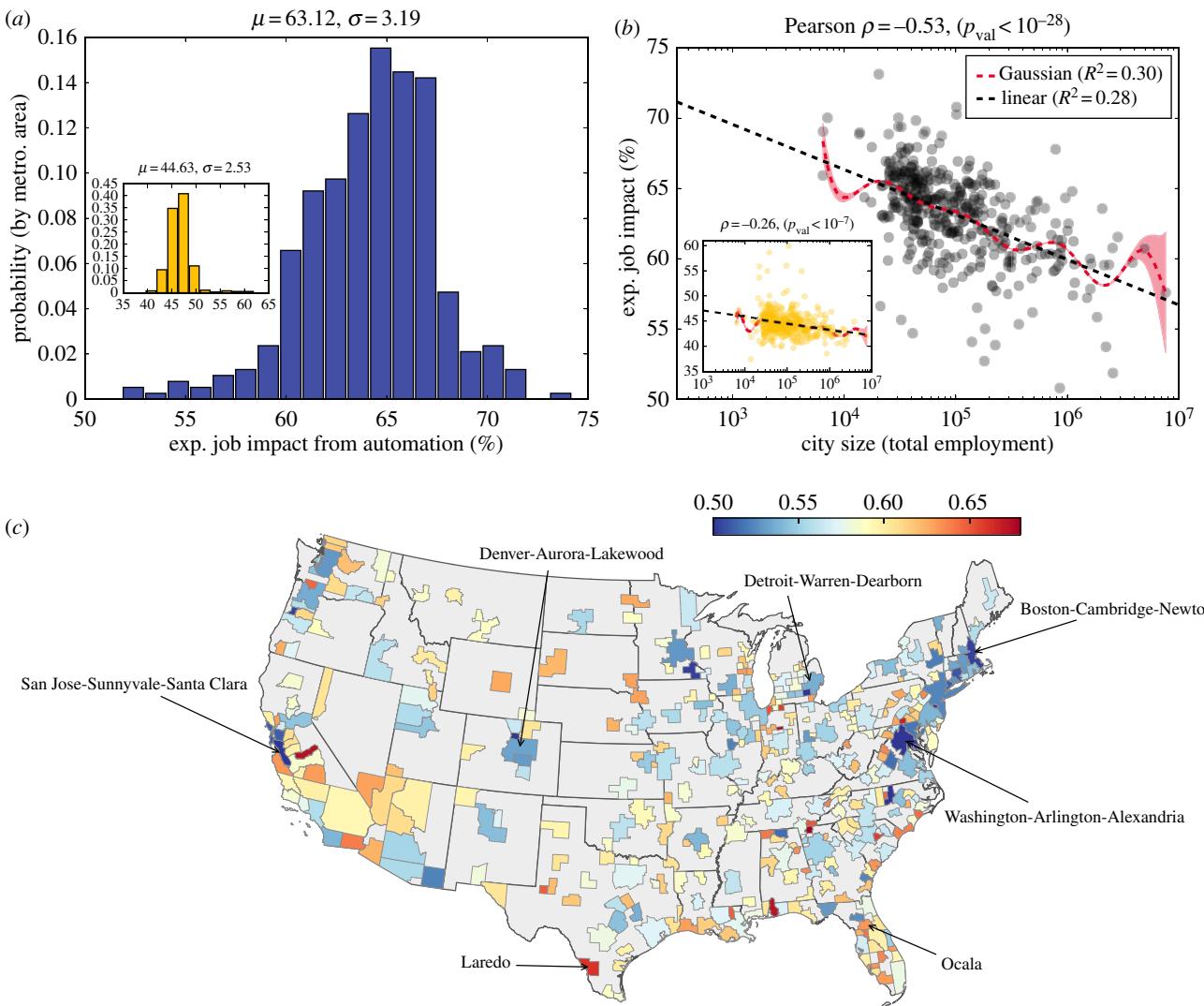
$$H_{\text{skill}}(m) = - \sum_{s \in \text{Skills}_m} p_m(s) \times \frac{\log(p_m(s))}{\log(|\text{Skills}_m|)}, \qquad (2.5)$$

where Skills_m represents the set of O*NET skills with non-zero importance in city m.

These aggregate skill distributions for a city may obfuscate the specialization of skills through the relative abundance of jobs in that city. For example, the city-level aggregation of skills may appear diverse, while the jobs within the city are actually specialized. The Theil entropy [36] of a city is a multi level information-theoretic measure defined by

$$T_m = \sum_{j \in \text{Jobs}_m} p_m(j) \cdot \frac{H_{\text{skill}}(m) - H_j}{H_{\text{skill}}(m)}. \qquad (2.6)$$

$T(m) = 1$ indicates that each job specializes in exactly one skill, and $T(m) = 0$ indicates that the specialization of skills among

Figure 1. The impact of automation in US cities. (a) The distribution of expected job impact (E_m) from automation across US cities using estimates from Frey & Osborne [12]. (Inset) The distribution using alternative estimates [13]. (b) Expected job impact decreases logarithmically with city size using estimates from Frey & Osborne [12]. We provide the line of best fit (slope $= -3.215$) with Pearson correlation to demonstrate significance (title). We also provide a Gaussian kernel regression model with its associated 95% confidence interval. (Inset) Decreased expected job impact with increased city size is again observed using alternative estimates [13] (best fit line has slope -1.24, Pearson $\rho = -0.26$, $p_{val} < 10^{-7}$). (c) A map of US metropolitan statistical areas coloured according to expected job impact from automation.

jobs is equal to the specialization of skills on the city-level aggregation. We do not observe any jobs relying on exactly one skill, and so we expect the Theil entropy of any given city to be well below 1. We present $1 - T_m$ throughout the study for easy comparison to Shannon entropy. Note that both normalized Shannon entropy and Theil entropy are unit-less measures due to the normalizations employed; we therefore do not focus on their range of values across cities, but instead we focus on the relationship between labour specialization/diversity and other urban indicators.

3. Results

3.1. The expected job impact of automation in cities
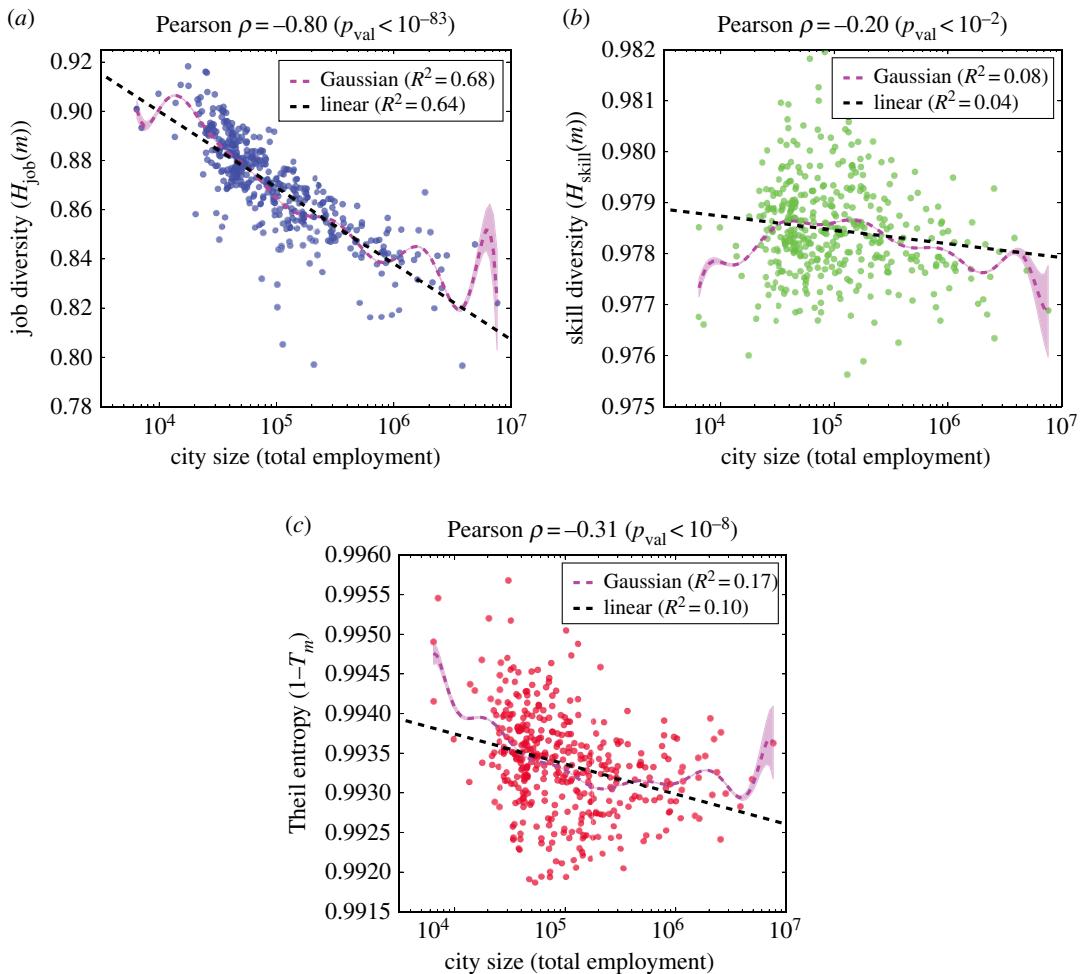
We estimate automation's expected impact on jobs in cities according to

$$E_m = \sum_{j \in \text{Jobs}} p_{\text{auto}}(j) \cdot \text{share}_m(j), \quad (3.1)$$

where Jobs denotes the set of occupations, $\text{share}_m(j)$ denotes

the employment share (as a percentage) in city m with occupation j according to the US BLS and $p_{\text{auto}}(j)$ denotes the probability of computerization for occupation j as estimated by [12] (see electronic supplementary material, S3 for more details). We can interpret E_m as the expected percentage of total employment in city m subject to computerization. Each city should expect between one-half and three-quarters of their current employment to be affected in the foreseeable future due to improvements in automation (see figure 1a; also note that this estimate differs from that in [12], which focused on national statistics). While this calculation omits potential job creation or job redefinition which typically accompany innovation [37,38], it highlights the differential impact of automation across cities and smooths potential noise in the predicted automation of individual jobs. Expected job impact may represent employment loss or changes in the type of work performed by those workers (e.g. see [11,17,18]), which, in turn, may not produce changes in net employment.

What differentiates cities' resilience to automation? Figure 1b demonstrates that expected job impact decreases

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) and (c), we provide the line of best fit for reference, and we provide a Gaussian kernel regression model with its associated 95% confidence interval.
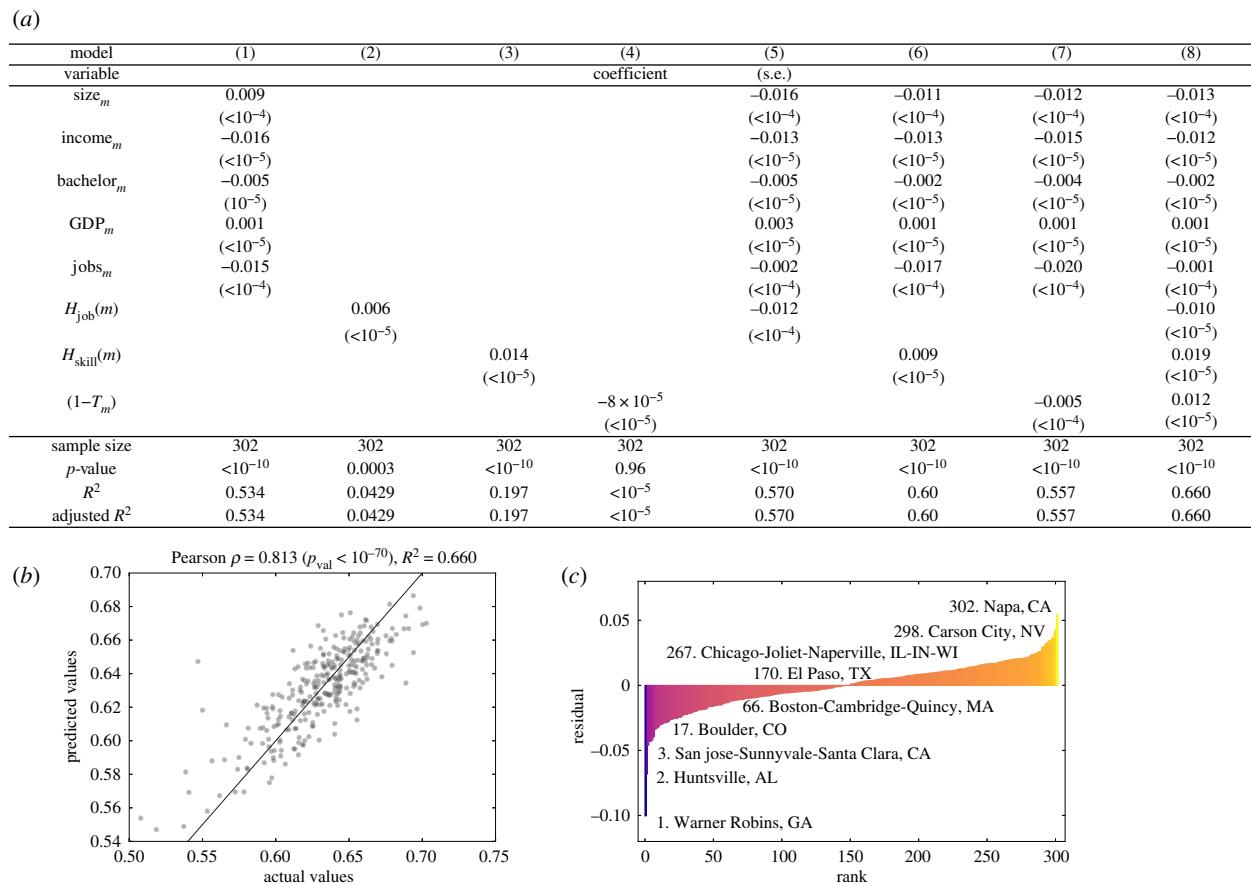
according to $E_m \propto -3.2 \times \log_{10}(\text{city size})$, which suggests that larger cities are more resilient to the negative effects of automation. This relationship is significant with a Pearson correlation $\rho = -0.53$ ($p_{val} < 10^{-28}$), and shows that labourers in smaller cities 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 [13] (Pearson $\rho = -0.26$ ($p_{val} < 10^{-7}$) and $E_m \propto -1.24 \times \log_{10}(\text{city size})$. (See figure 1b inset; electronic supplementary material, S3.2.) Despite the conservative nature of these alternative probabilities, we again observe increased resilience with city size. Furthermore, we demonstrate in electronic supplementary material, S3.1 that the observed negative trend relating city size to expected job impact from automation is robust to errors in the probabilities of computerization (i.e. p_{auto}) produced by Frey & Osborne [12] and robust to random removal of occupations from the analysis.

3.2. Labour specialization in large cities

We explore the mechanisms underpinning resilience to automation by examining the most distinctive characteristics of urban economies: diversification and specialization. In particular, how does labour diversity, or specialization, mediate the relationship between city size and the expected job impact from automation? As automation typically targets workplace skills [13], we consider the O*NET skill dataset, which relates occupations to their constituent workplace tasks and skills, in addition to employment data. For large cities, specialization (i.e. decreased Shannon entropy) appears in the employment distributions across occupations (figure 2a) and, separately, in the aggregate distributions of skills (figure 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 the whole. Figure 2c demonstrates an increasing proportion of specialized jobs in large cities (i.e. $1 - T_m$ decreases). See Material and methods for calculations of entropy measures.

In figure 3, we examine eight regression models attempting to model the differential impact of automation across cities. In model 1, we first examine a baseline model using only generic urban variables, including city size (denoted by size_m), median household income (income_m), the per cent of population with a bachelor's degree (bachelor_m), per capita GDP (GDP_m) and the number of unique job titles (jobs_m). This generic model captures 53% of the variance in expected impact from automation across US cities. Models 2, 3 and 4 use the information-theoretic measures in three separate linear regression models to reveal that skill

(a)

variable	(1)	(2)	(3)	(4) coefficient	(5) (s.e.)	(6)	(7)	(8)
$size_m$	0.009 ($<10^{-4}$)				−0.016 ($<10^{-4}$)	−0.011 ($<10^{-4}$)	−0.012 ($<10^{-4}$)	−0.013 ($<10^{-4}$)
$income_m$	−0.016 ($<10^{-5}$)				−0.013 ($<10^{-5}$)	−0.013 ($<10^{-5}$)	−0.015 ($<10^{-5}$)	−0.012 ($<10^{-5}$)
$bachelor_m$	−0.005 (10^{-5})				−0.005 ($<10^{-5}$)	−0.002 ($<10^{-5}$)	−0.004 ($<10^{-5}$)	−0.002 ($<10^{-5}$)
GDP_m	0.001 ($<10^{-5}$)				0.003 ($<10^{-5}$)	0.001 ($<10^{-5}$)	0.001 ($<10^{-5}$)	0.001 ($<10^{-5}$)
$jobs_m$	−0.015 ($<10^{-4}$)				−0.002 ($<10^{-4}$)	−0.017 ($<10^{-4}$)	−0.020 ($<10^{-4}$)	−0.001 ($<10^{-4}$)
$H_{job}(m)$		0.006 ($<10^{-5}$)			−0.012 ($<10^{-4}$)			−0.010 ($<10^{-5}$)
$H_{skill}(m)$			0.014 ($<10^{-5}$)			0.009 ($<10^{-5}$)		0.019 ($<10^{-5}$)
$(1-T_m)$				-8×10^{-5} ($<10^{-5}$)			−0.005 ($<10^{-4}$)	0.012 ($<10^{-5}$)
sample size	302	302	302	302	302	302	302	302
p-value	$<10^{-10}$	0.0003	$<10^{-10}$	0.96	$<10^{-10}$	$<10^{-10}$	$<10^{-10}$	$<10^{-10}$
R^2	0.534	0.0429	0.197	$<10^{-5}$	0.570	0.60	0.557	0.660
adjusted R^2	0.534	0.0429	0.197	$<10^{-5}$	0.570	0.60	0.557	0.660

(b) Pearson $\rho = 0.813$ ($p_{val} < 10^{-70}$), $R^2 = 0.660$

(c) 302. Napa, CA; 298. Carson City, NV; 267. Chicago-Joliet-Naperville, IL-IN-WI; 170. El Paso, TX; 66. Boston-Cambridge-Quincy, MA; 17. Boulder, CO; 3. San jose-Sunnyvale-Santa Clara, CA; 2. Huntsville, AL; 1. Warner Robins, GA

Figure 3. Labour specialization can model expected job impact (E_m) in cities. (a) A multiple linear regression analysis for predicting E_m that considers generic urban indicators including $\log_{10}$ city total employment ($size_m$), median annual household income ($income_m$), percentage of population with a bachelor's degree ($bachelor_m$), $\log_{10}$ GDP *per capita* (GDP_m) and the number of unique occupations ($jobs_m$). All variables have been standardized. (b) The actual E_m values for each city plotted against the predicted values using model 8 from (a), which captures 66% of the variance in expected job impact from automation across US cities (see electronic supplementary material, S4 for additional analysis). (c) The distribution of residuals between the actual and predicted values from model 8, and the rank of some example cities.

specialization (i.e. $H_{skill}(m)$) is the most predictive of expected job impact in cities ($R^2 = 0.20$) in the absence of other urban variables.

In models 5, 6 and 7, we demonstrate that the inclusion of each of the specialization measures produces models accounting for additional variance in expected impact over the use of generic variables alone. In particular, the inclusion of skill specialization (i.e. $H_{skill}(m)$) yields a model accounting for 60% of the variance in job impact (see model 6). Finally, we include all variables in a single model (see model 8) which produces the most predictive model accounting for 66% of the variance across cities (figure 3a,b). 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 electronic supplementary material, S4).

Each model that we tested yielded statistically significant coefficient estimates (note that variables were standardized before regression) and the inclusion of our labour specialization metrics yielded models with improved predictive power. Furthermore, we performed a formal mediation analysis that is presented in electronic supplementary material, §3.5. However, these observations should not be taken as conclusive evidence that labour specialization, or diversity, is causally related to the expected impact from automation in a city. This is due, in part, to the colinearity between variables used in model 8. For example, the city size coefficient ($size_m$)
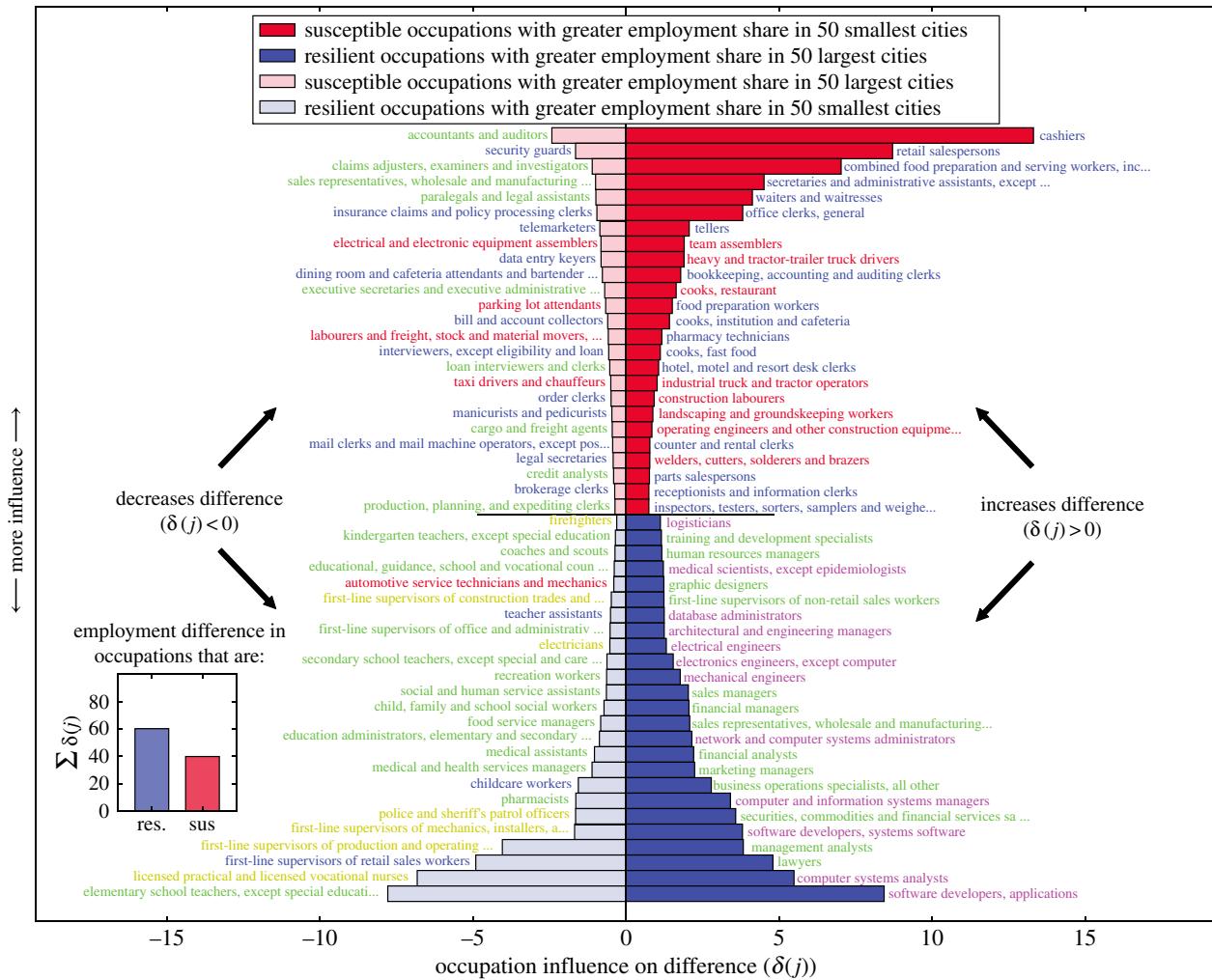
changes sign across the models in our analysis because of the strong relationship between city size and labour specialization, which we demonstrate in figure 2.

The residuals between the actual and modelled values according to model 8 highlight notably resilient cities (given the model), such as Boulder, CO, and Warner Robins, GA, and notably susceptible cities, such as Napa, CA, and Carson City, NV (figure 3c). 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.

3.3. How occupations and workplace skills change with city size

How do different types of occupations change with city size [39], and how do these changes contribute to the differential impact of automation across cities? While it is tempting to look only for the largest changes in employment share, more subtle differences for very automatable, or very not automatable, occupations can also produce big changes in expected job impact. We capture this confounding effect by decomposing the difference in expected job impact of cities

50 smallest cities (0.65) versus 50 largest cities (0.60)

Legend:
- susceptible occupations with greater employment share in 50 smallest cities
- resilient occupations with greater employment share in 50 largest cities
- susceptible occupations with greater employment share in 50 largest cities
- resilient occupations with greater employment share in 50 smallest cities

decreases difference ($\delta(j) < 0$)

increases difference ($\delta(j) > 0$)

more influence →

employment difference in occupations that are:

$\sum \delta(j)$: res. / sus.

occupation influence on difference ($\delta(j)$)

Left side occupations (decreases difference):
accountants and auditors, security guards, claims adjusters, examiners and investigators, sales representatives, wholesale and manufacturing ..., paralegals and legal assistants, insurance claims and policy processing clerks, telemarketers, electrical and electronic equipment assemblers, data entry keyers, dining room and cafeteria attendants and bartender ..., executive secretaries and executive administrative ..., parking lot attendants, bill and account collectors, labourers and freight, stock and material movers, ..., interviewers, except eligibility and loan, loan interviewers and clerks, taxi drivers and chauffeurs, order clerks, manicurists and pedicurists, cargo and freight agents, mail clerks and mail machine operators, except pos..., legal secretaries, credit analysts, brokerage clerks, production, planning, and expediting clerks, firefighters, kindergarten teachers, except special education, coaches and scouts, educational, guidance, school and vocational coun ..., automotive service technicians and mechanics, first-line supervisors of construction trades and ..., teacher assistants, first-line supervisors of office and administrativ ..., electricians, secondary school teachers, except special and care ..., recreation workers, social and human service assistants, child, family and school social workers, food service managers, education administrators, elementary and secondary ..., medical assistants, medical and health services managers, childcare workers, pharmacists, police and sheriff's patrol officers, first-line supervisors of mechanics, installers, a..., first-line supervisors of production and operating ..., first-line supervisors of retail sales workers, licensed practical and licensed vocational nurses, elementary school teachers, except special educati...

Right side occupations (increases difference):
cashiers, retail salespersons, combined food preparation and serving workers, inc..., secretaries and administrative assistants, except ..., waiters and waitresses, office clerks, general, tellers, team assemblers, heavy and tractor-trailer truck drivers, bookkeeping, accounting and auditing clerks, cooks, restaurant, food preparation workers, cooks, institution and cafeteria, pharmacy technicians, cooks, fast food, hotel, motel and resort desk clerks, industrial truck and tractor operators, construction labourers, landscaping and groundskeeping workers, operating engineers and other construction equipme..., counter and rental clerks, welders, cutters, solderers and brazers, parts salespersons, receptionists and information clerks, inspectors, testers, sorters, samplers and weighe..., logisticians, training and development specialists, human resources managers, medical scientists, except epidemiologists, graphic designers, first-line supervisors of non-retail sales workers, database administrators, architectural and engineering managers, electrical engineers, electronics engineers, except computer, mechanical engineers, sales managers, financial managers, sales representatives, wholesale and manufacturing..., network and computer systems administrators, financial analysts, marketing managers, business operations specialists, all other, computer and information systems managers, securities, commodities and financial services sa ..., software developers, systems software, management analysts, lawyers, computer systems analysts, software developers, applications

Figure 4. An occupation shift explaining the difference in expected job impact for the 50 largest cities (impact: 0.60) compared to the 50 smallest cities (impact: 0.65) using equation (3.2). Each horizontal bar represents $\delta_{\text{(small cities, large cities)}}(j)$. The occupation title is provided next to the corresponding bar and coloured according to its job cluster. Red bars represent occupations with higher risk of computerization compared to the expected job impact in large cities. Blue bars represent occupations with lower risk of computerization compared to the expected job impact in large cities. Dark colours represent occupations that increase the difference, while pale colours represent occupations that decrease the difference in expected job impact. Bars in each of the quadrants are vertically ordered according to $\left|\delta_{\text{(small cities, large cities)}}(j)\right|$. The inset in the bottom left of the plot summarizes the overall influence of resilient occupations compared to occupations that are at risk of computerization.

m and n according to

$$E_m - E_n = \sum_{j \in \text{Jobs}} p_{\text{auto}}(j) \times (\text{share}_m(j) - \text{share}_n(j))$$

$$= \sum_{j \in \text{Jobs}} (p_{\text{auto}}(j) - E_n) \times (\text{share}_m(j) - \text{share}_n(j)), \quad (3.2)$$
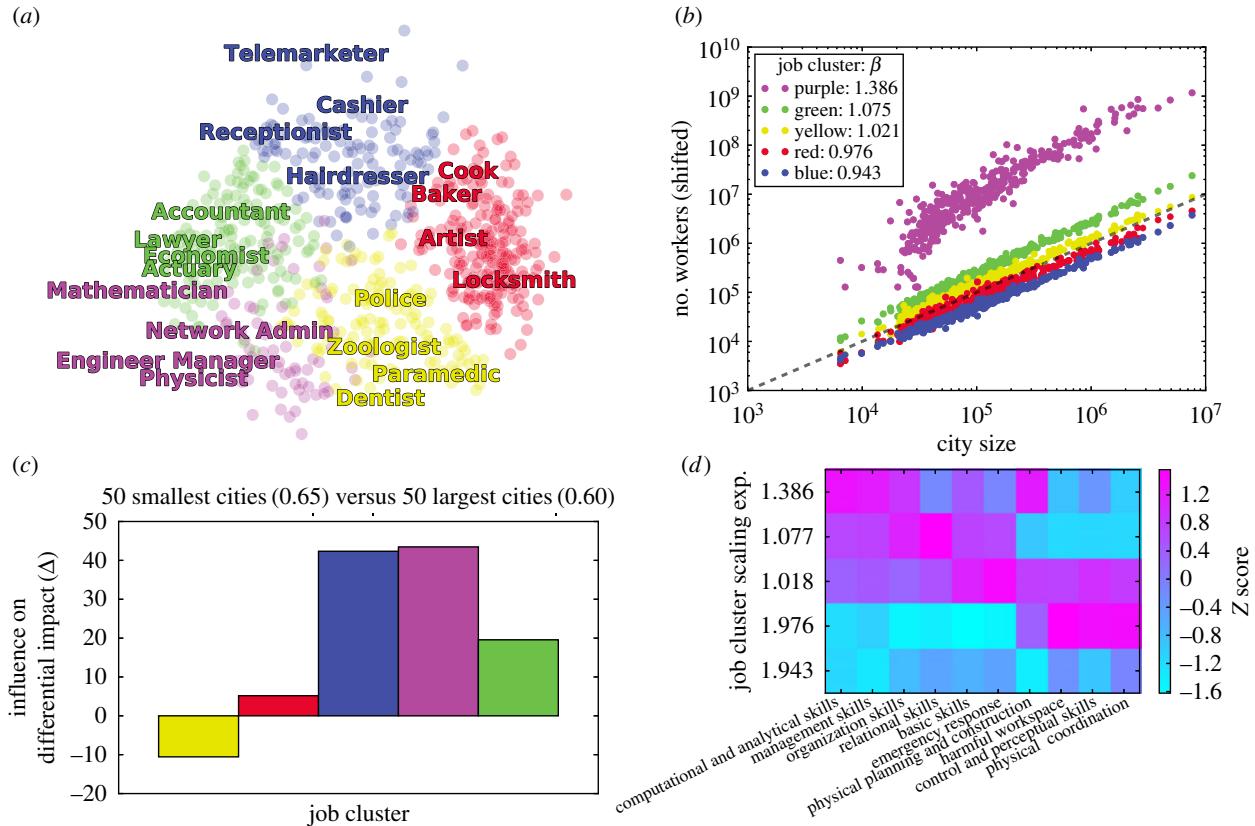
where we have profited from $\sum E_n \times (\text{share}_m(j) - \text{share}_n(j)) = 0$. We consider the percentage of the difference explained by occupation j according to

$$\delta_{m,n}(j) = 100 \times \frac{(p_{\text{auto}}(j) - E_n) \times (\text{share}_m(j) - \text{share}_n(j))}{E_m - E_n}. \quad (3.3)$$

Occupation j can increase or decrease the overall difference in expected job impact depending on the sign of the corresponding term in equation (3.2), or, equivalently, the sign of $\delta_{m,n}(j)$. In turn, this sign depends on the relative automatability of the occupation and the relative employment share. More details for this calculation and an example analysis comparing

individual cities are provided in electronic supplementary material, S3.4.

In figure 4, we employ an 'occupation shift' to visualize the contributions of each occupation to the difference in expected job impact in large and small cities. After adding the employment distributions for the 50 largest cities and 50 smallest cities together, respectively, we calculate $\delta(j)$ for each occupation. Each occupation is assigned a quadrant and colour based on the sign of $\delta(j)$ and the relative automatability of occupation j. This visualization identifies both occupations that increase the differential impact (i.e. occupations on the right) and occupations that decrease the differential impact (i.e. occupations on the left). For example, increased employment for Cashiers, which is relatively susceptible to automation, in small cities contributes the most to the overall difference in expected job impact. Likewise, differences in employment for Software Developers, a relatively resilient occupation, also increases the overall difference. On the other hand, increased employment for Elementary School Teachers, which is another relatively resilient occupation, in small cities decreases the difference. On

Figure 5. Technical occupations grow superlinearly with city size. (*a*) We project jobs onto a two-dimensional plane using principal component analysis. A few representative jobs from each cluster are highlighted (colour). (*b*) We plot the employment (*y*-axis) in a given job cluster (colour) versus the total employment in a city (*x*-axis), and vertically shift points according to the linear fit in log scale. The black dashed line has a slope of 1 for reference. (*c*) The influence of each job cluster on the difference in expected job impact of the 50 largest cities ($E_{\text{large cities}} = 0.60$) compared to the 50 smallest cities ($E_{\text{small cities}} = 0.65$) according to equation (3.4). (*d*) After summing the importance of each skill type to each job cluster, we calculate *z* scores for a skill type according to the distribution of importance across job clusters.

aggregate, differences in employment for occupations that are relatively resilient to automation contribute the most to the differential impact of automation in large and small cities (see figure 4 inset).

To explore the role of resilient occupations further, we focus on how employment for different occupation types changes with city size. We use the 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 figure 4 occupation labels and figure 5*a*; the complete list of occupations is provided in electronic supplementary material, S6.3) and examine the scaling relationship between job clusters and city size according to (number of workers)$\propto$(city size)$^{\beta}$ in figure 5*b*. Note that the exponent, β, entirely describes the growth rate of these job clusters relative to city size. The job cluster comprising 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 [6] and is in good agreement with similar studies on job growth [17]. 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 electronic supplementary material, S6.3.2). Managerial jobs also grow superlinearly, but to a weaker extent ($\beta = 1.08$). The job cluster exhibiting the slowest growth ($\beta = 0.94$) comprises entertainment and service jobs. We check the
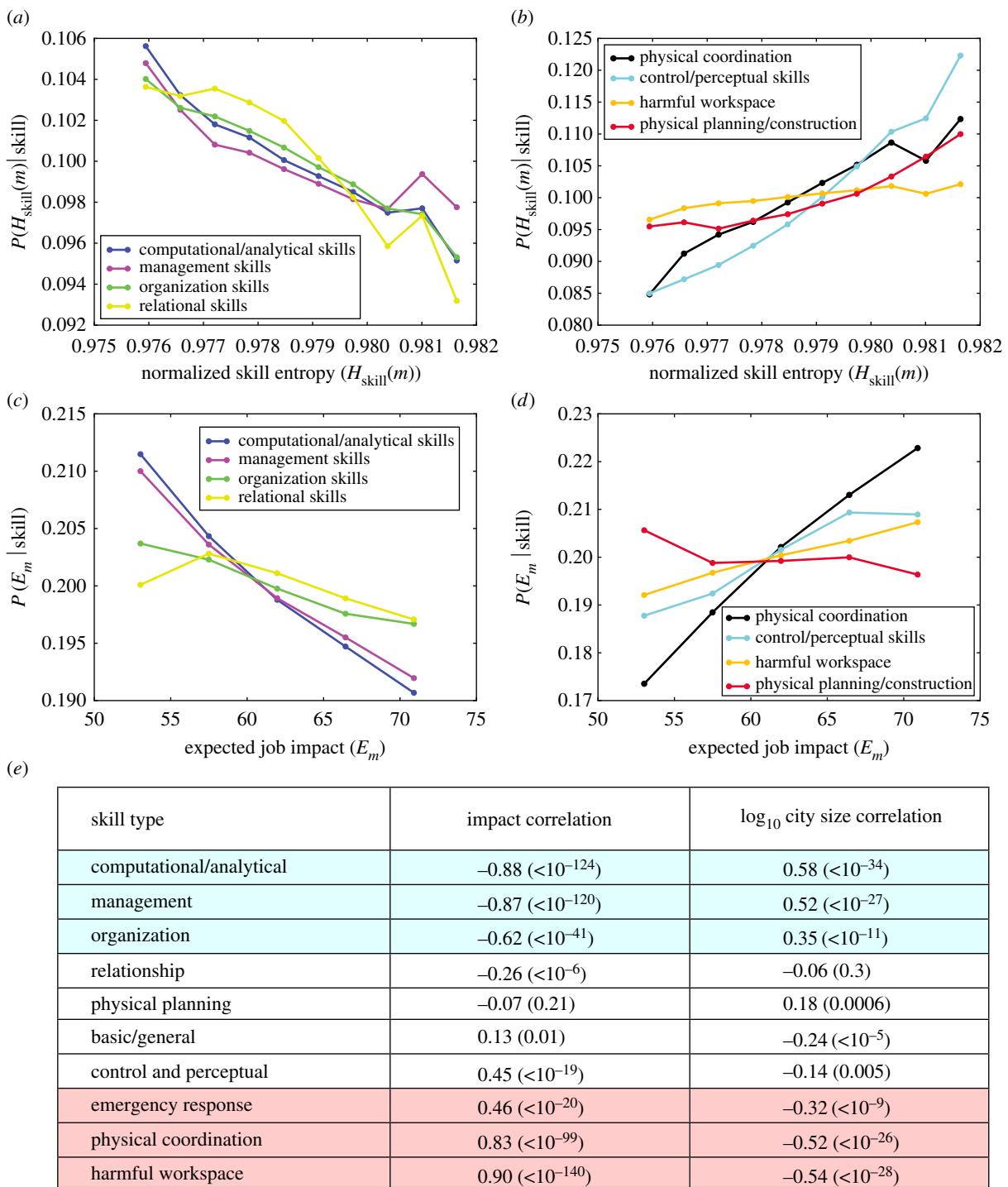
robustness of these scaling relationships using methods from [40] (see electronic supplementary material, S6.3.3).

In figure 5*c*, we quantify each job cluster's contribution to the differential impact of automation across large and small cities according to

$$\Delta_{\text{small cities, large cities}}(\text{job cluster})$$
$$= \sum_{j \in \text{job cluster}} \delta_{\text{small cities, large cities}}(j). \quad (3.4)$$

The low automatability and high difference in employment of highly specialized job cluster (represented by purple) in large and small cities indeed explains a significant amount of the difference in expected job impact. However, we also find that the more susceptible occupations represented by the blue job cluster in figure 5 accounts for a similar proportion of the difference. Interestingly, the differences in occupations from the yellow job cluster serve to decrease the differential impact of automation between large and small cities. This conclusion is supported by the analysis of individual occupations presented in figure 4.

We confirm that the fastest growing job cluster is indeed comprised of 'technical' jobs based on 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 10 skill types based on the co-occurrence of skills across jobs (see electronic supplementary

Figure 6. Workplace skills explain occupational specialization and job impact in cities. (a,b) Skill types in (a) indicate specialized cities, while skill types in (b) indicate occupational diversity. (c,d) Skill types in (c) indicate resilient cities, while skill types in (d) indicate increased job impact from automation. (e) The Pearson correlation of skill type abundance to the expected job impact and to $\log_{10}$ city size with p-values in parentheses. See electronic supplementary material, S6.4 for a similar table for raw O*NET skills.

material, S6.5 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 (figure 5d). We confirm our findings using alternative definitions for aggregate workplace tasks and skills (see electronic supplementary material, S5).

The skills that are relied on by fast-growing technical jobs suggest mechanisms for resilience and growth in cities.

Existing work [41] 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 labour 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 (figure 6a,c), while Physical Coordination and Control/Perceptual skills indicate both decreased specialization and decreased resilience in cities (figure 6b,d). We confirm our results using alternative groups of

workplace tasks [42] provided by O*NET (see electronic supplementary material, S5.1) and again by examining the routineness of workplace tasks [14] (see electronic supplementary material, S5.2). Figure 6e reflects the same conclusion by comparing the relationship of each skill type to city size (right column) and expected job impact (middle column; see electronic supplementary material, S6.4 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.

3.4. Limitations

Many of the limitations inherent to occupation-level predictions [12,13] apply to our study as well. Specifically, our measure for the expected impact of automation in cities may represent technological unemployment, but also represents the skill recomposition of occupations in response to new technology. This means the expected impact of automation in cities may not relate to changes in net employment in cities. The actual effects of automation on net employment levels depend on several systemic variables including the availability of cheap labour [43,44], future regulations around technology (e.g. taxing the use of robotics) and market demand with increased worker efficiency [17,18].

Nevertheless, we expect the impact we are measuring to correspond to costly real-world changes in labour that high-impact cities must overcome. For example, cities with high expected impact from automation will need to invest in worker retraining programmes. These programmes minimize technological unemployment by adapting the existing skills of workers to match the evolving skill demands with changing technology, but these programmes are costly. Urban policymakers may also mitigate net employment loss by investing in new industries, but successful investment of this kind requires costly research and capital to attract those companies to a city.

4. Discussion

Cities are modern society's hubs for economic productivity and innovation. However, the impact of automation on employment in cities threatens to alter urbanization, which is largely driven by employment opportunity. Fortunately, urbanization itself appears to contain a mitigating solution. 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 labour in large cities and show its importance as a piece of the automation and urbanization puzzle.

In particular, large cities have more unique occupations and industries [7], but distribute employment less uniformly across those occupations. This juxtaposition of both diversity and specialization in large cities is reconcilable through the division of labour theory [24]. Under the division of labour argument, large firms have better ability to support specialized workers along with the management required to coordinate them [45]. To this end, we find that the average number of workers per firm increases logarithmically with city size (see electronic supplementary material, figure S1A). At the same time, workers possessing specialized skills seek specific employment opportunities which maximize their financial return [46,47]. The demand for specific specialized jobs increases occupational specialization while also increasing the number of unique job types and industries in a city [8].

What do large cities specialize in and why? The division of labour encourages worker modularity, which has the potential to impact whole groups of workers who are competing with automation technology. Therefore, specialization alone is not enough to explain the resilience to automation impact that we observe across cities. For example, Detroit, which is famous for its specialization in automotive manufacturing, has experienced economic downturn [48], while the San Francisco Bay area, the epicentre of the information technology industry, continues to flourish despite the dot-com bubble (perhaps due to its support of a 'creative class' of workers [49]). Our analysis highlights specific occupations, such as Mathematician and Chemist, as well as specific types of skills, such as Computational/Analytical skill, that explain the increased resilience of large cities. These occupations and skills may inform policymakers in small cities as they identify new industries and design worker retraining programmes to mitigate the negative effects of automation on employment.

By quantifying relative impact, we provide an upper bound for technological unemployment in cities. Changing labour demands produce systemic effects, which make it difficult to precisely predict employment loss [15]. 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, and, as a result, national bank teller employment increased [17,18]. However, these bank 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 technological unemployment in a city 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 retraining, or skilled migration). Nevertheless, we observe a strong trend relating city size to automation impact that is robust to errors in the automatability of individual occupations and occupational subsampling. For example, the estimates of occupational automation, which we employ in our analysis, would need to be severely flawed (errors over 50%) for the negative dependency on city size to disappear. Recognizing that small cities will experience larger adjustments to automation calls on policymakers to pay special attention to the pronounced risks we have identified.

Despite being seemingly unrelated societal forces, we uncover a positive interplay between urbanization and automation. Larger cities not only tend to be more innovative [5,6], but also harbour 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. These findings open the door for more controlled investigations with input from policymakers.

References

1. Lee J. 2015 *World Migration Report 2015. Migrants and cities: new partnerships to manage mobility*. Geneva, Switzerland. IOM: International Organization for Migration.

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

3. Ash C, Jasny BR, Roberts L, Stone R, Sugden AM. 2008 Reimagining cities. *Science* **319**, 739–739. (doi:10.1126/science.319.5864.739)

4. Montgomery MR. 2008 The urban transformation of the developing world. *Science* **319**, 761–764. (doi:10.1126/science.1153012)

5. Bettencourt LMA, Lobo J, Strumsky D, West GB. 2010 Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* **5**, e13541. (doi:10.1371/journal.pone.0013541)

6. Bettencourt LM, Lobo J, Helbing D, Kühnert C, West GB. 2007 Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306. (doi:10.1073/pnas.0610172104)

7. Bettencourt LMA, Lobo J, Strumsky D, Samaniego H, West GB. 2016 Scaling and universality in urban economic diversification. *J. R. Soc. Interface* **13**, 20150937. (doi:10.1098/rsif.2015.0937)

8. Rozenblat C, Pumain D. 2007 Firm linkages, innovation and the evolution of urban systems. In *Cities in globalization: practices, policies, and theories* (eds PJ Taylor, B Derudder, P Saey, F Witlox), pp. 130–156. London, UK: Routledge.

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

10. Olsen M, Hemous D. 2014 The rise of the machines: automation, horizontal innovation and income inequality. In *2014 Meeting Papers*, 162 (Society for Economic Dynamics).

11. Acemoglu D, Restrepo P. 2017 Robots and jobs: evidence from US labor markets. NBER Working Paper No. 23285. Issued in March 2017. NBER Program(s): Economic Fluctuations and Growth, Labor Studies. Cambridge, MA: National Bureau of Economic Research. (doi:10.3386/w23285)

12. Frey CB, Osborne MA. 2016 The future of employment: how susceptible are jobs to computerisation? *Technol. Forecast. Soc. Change* **114**, 254–280. (doi:10.1016/j.techfore.2016.08.019)

13. Arntz M, Gregory T, Zierahn U. 2016 The risk of automation for jobs in OECD countries: a

comparative analysis. OECD Social, Employment and Migration Working Papers no. 189, Paris.

14. Autor DH, Levy F, Murnane RJ. 2001 The skill content of recent technological change: an empirical exploration. Technical Report, National Bureau of Economic Research.

15. Autor DH. 2015 Why are there still so many jobs? The history and future of workplace automation. *J. Econ. Perspect.* **29**, 3–30.

16. Brynjolfsson E, McAfee A. 2014 *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. New York, NY: W.W. Norton & Company.

17. Bessen JE. 2015 How computer automation affects occupations: technology, jobs, and skills. Boston Univ. School of Law, Law and Economics Research Paper.

18. Bessen J. 2015 *Learning by doing: the real connection between innovation, wages, and wealth*. New Haven, CT: Yale University Press.

19. Glaeser E. 2011 *Triumph of the city: how our greatest invention makes us richer, smarter, greener, healthier, and happier*. Harmondsworth, UK: Penguin.

20. Quigley JM. 1998 Urban diversity and economic growth. *J. Econ. Perspect.* **12**, 127–138. (doi:10.1257/jep.12.2.127)

21. Henderson JV. 1991 Urban development: theory, fact, and illusion. OUP Catalogue.

22. Pan W, Ghoshal G, Krumme C, Cebrian M, Pentland A. 2013 Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* **4**, 1961. (doi:10.1038/ncomms2961)

23. Sim A, Yaliraki SN, Barahona M, Stumpf MP. 2015 Great cities look small. *J. R. Soc. Interface* **12**, 20150315. (doi:10.1098/rsif.2015.0315)

24. Smith A. 1976 An inquiry into the nature and causes of the wealth of nations (eds RH campbell, AS skinner, WB todd). Oxford, UK: Clarendon Press.

25. Bettencourt LM, Samaniego H, Youn H. 2014 Professional diversity and the productivity of cities. *Sci. Rep.* **4**, 5393. (doi:10.1038/srep05393)

26. Sveikauskas L. 1975 The productivity of cities. *Q. J. Econ.* **89**, 393–413. (doi:10.2307/1885259)

27. Arcaute E, Molinero C, Hatna E, Murcio R, Vargas-Ruiz C, Paolo Masucci A, Batty M. 2016 Cities and regions in britain through hierarchical percolation. *Open Sci.* **3**, 150691. (doi:10.1098/rsos.150691)

28. Rozenfelda HD, Rybskia D, Andrade Jr JS, Battyc M, EugeneStanley H, Makse HA. 2008 Laws of population growth. *Proc. Natl Acad. Sci. USA*

105, 18 702–18 707. (doi:10.1073/pnas.0807435105)

29. Rozenfeld HD, Rybski D, Gabaix X, Makse HA. 2011 The area and population of cities: new insights from a different perspective on cities. *Am. Econ. Rev.* **101**, 2205–2225. (doi:10.1257/aer.101.5.2205)

30. Oliveira EA, Andrade Jr JS, Makse HA. 2014 Large cities are less green. *Sci. Rep.* **4**, 4235. (doi:10.1038/srep04235)

31. Operti FG, Oliveira EA, Carmona HA, Machado JC, Andrade JS. 2018 The light pollution as a surrogate for urban population of the US cities. *Phys. A: Stat. Mech. Appl.* **492**, 1088–1096. (doi:10.1016/j.physa.2017.11.039)

32. Kumar U, Kumar V, Kapur JN. 1986 Normalized measures of entropy. *Int. J. Gen. Syst.* **12**, 55–69. (doi:10.1080/03081078608934927)

33. Cabot B, Martell M, Esteban JI, Sauleda S, Otero T, Esteban R, Guàrdia J, Gómez J. 2000 Nucleotide and amino acid complexity of hepatitis C virus quasispecies in serum and liver. *J. Virol.* **74**, 805–811. (doi:10.1128/JVI.74.2.805-811.2000)

34. Wijesekera HW, Dillon TM. 1997 Shannon entropy as an indicator of age for turbulent overturns in the oceanic thermocline. *J. Geophys. Res.: Oceans* **102**, 3279–3291. (doi:10.1029/96JC03605)

35. Eagle N, Macy M, Claxton R. 2010 Network diversity and economic development. *Science* **328**, 1029–1031. (doi:10.1126/science.1186605)

36. Reardon SF, Firebaugh G. 2002 Measures of multigroup segregation. *Sociol. Methodol.* **32**, 33–67. (doi:10.1111/1467-9531.00110)

37. Mazzucato M. 2013 Financing innovation: creative destruction vs. destructive creation. *Ind. Corp. Change* **22**, 851–867. (doi:10.1093/icc/dtt025)

38. Archibugi D, Filippetti A, Frenz M. 2013 Economic crisis and innovation: is destruction prevailing over accumulation? *Res. Policy* **42**, 303–314. (doi:10.1016/j.respol.2012.07.002)

39. Pumain D. 2004 Scaling laws and urban systems. Santa Fe Working Paper. Santa Fe, NM: Santa Fe Institute.

40. Leitão JC, Miotto JM, Gerlach M, Altmann EG. 2016 Is this scaling nonlinear? *R. Soc. open sci.* **3**, 150649. (doi:10.1098/rsos.150649)

41. MacCrory F, Westerman G, Alhammadi Y, Brynjolfsson E. 2014 Racing with and against

the machine: changes in occupational skill composition in an era of rapid technological advance. In *Int. Conf. on Information Systems* (*ICIS*), *Auckland, New Zealand, 14–17 December.*

42. Kok S, Weel B t. 2014 Cities, tasks, and skills. *J. Reg. Sci.* **54**, 856–892. (doi:10.1111/jors.12125)

43. Acemoglu D. 2003 Labor- and capital-augmenting technical change. *J. Eur. Econ. Assoc.* **1**, 1–37. (doi:10.1162/154247603322256756)

44. Habakkuk HJ. 1962 *American and British technology in the nineteenth century: the search for labour saving inventions.* Cambridge, UK: Cambridge University Press.

45. Coase RH. 1937 The nature of the firm. *Economica* **4**, 386–405. (doi:10.1111/j.1468-0335.1937.tb00002.x)

46. Bloom DE, Canning D, Fink G. 2008 Urbanization and the wealth of nations. *Science* **319**, 772–775. (doi:10.1126/science.1153057)

47. Schich M, Song C, Ahn Y-Y, Mirsky A, Martino M, Barabási A-L, Helbing D. 2014 A network framework of cultural history. *Science* **345**, 558–562. (doi:10.1126/science.1240064)

48. Klier T. 2009 From tail fins to hybrids: how detroit lost its dominance of the US auto market. *Econ. Perspect.* **33**, 2–17.

49. Florida R. 2004 *The rise of the creative class.* New York, NY: Basic books.