NEW FRONTIERS: THE ORIGINS AND CONTENT OF NEW WORK, 1940–2018*

David Autor
Caroline Chin
Anna Salomons
Bryan Seegmiller

Abstract

We answer three core questions about the hypothesized role of newly emerging job categories ('new work') in counterbalancing the erosive effect of task-displacing automation on labor demand: what is the substantive content of new work; where does it come from; and what effect does it have on labor demand? We construct a novel database spanning eight decades of new job titles linked both to US Census microdata and to patent-based measures of occupations' exposure to labor-augmenting and labor-automating innovations. The majority of current employment is in new job specialties introduced since 1940, but the locus of new work creation has shifted from middle-paid production and clerical occupations over 1940–1980 to high-paid professional and, secondarily, low-paid services since 1980. New work emerges in response to technological innovations that complement the outputs of occupations and demand shocks that raise occupational demand. Innovations that automate tasks or reduce occupational demand slow new work emergence. Although the flow of augmentation and automation innovations is positively correlated across occupations, the former boosts occupational labor demand while the latter depresses it. The demand-eroding effects of automation innovations have intensified in the last four decades while the demand-increasing effects of augmentation innovations have not.

Keywords: Technological Change, New Tasks, Augmentation, Automation, Demand Shifts
JEL: E24, J11, J23, J24

*Corresponding author: dautor@mit.edu , Massachusetts Institute of Technology, 50 Memorial Drive, E52-438, Cambridge, MA 02142
Word count: 16,895
We thank Daron Acemoglu, David Deming, Brad DeLong, Matt Gentzkow, Maarten Goos, Gordon Hanson, Lawrence Katz, Lynda Laughlin, Magnus Lodefalk, Peter Lambert, Marin Soljačić, Sebastian Steffen, John Van Reenen, three anonymous referees, the expert staff of the U.S. Census Bureau, and innumerable seminar and conference participants for insights and critiques that have vastly improved the paper; and Fiona Chen, Grace Chuan, Rebecca Jackson, Zhe Fredric Kong, Jonathan Rojas, Edwin Song, Ishaana Talesara, Liang Sunny Tan, Jose Velarde, Yuting Brenda Wu, Rocky Xie, and Whitney Zhang for expert research assistance. We thank Dimitris Papanikolaou for sharing a database of raw patent texts; and Sammy Mark, Fredrick Soo, and Mary Wakarindy from Digital Divide Data for hand-keying historical Census Alphabetical Index records. We thank Greg Bronevetsky and Guy Ben-Ishai of Google for facilitating industrial-level usage of the Google Ngram Viewer to analyze the emergence of new job titles. We gratefully acknowledge support from The Carnegie Corporation, Google.org, Instituut Gak, the MIT Work of the Future Task Force, Schmidt Futures, the Smith Richardson Foundation, and the Washington Center for Equitable Growth.

I Introduction

A vast economic literature analyzes how rapidly evolving digital technologies—information and communications technologies, robotics, artificial intelligence—affect employment, skill demands, and earnings levels. Focusing on the substitution of machines for workers in tasks where automation has rising comparative advantage, this research anticipates and interprets the decline of middle-skill employment in high income countries (aka, job polarization) and attendant impacts on wage structure, documents the concentrated impact of industrial robotics on labor demand in heavy manufacturing industries and in manufacturing-intensive communities, and explores how artificial intelligence may change the structure of occupations.

This work is however comparatively silent—with key exceptions, discussed below—on the flip side of the ledger: the augmentation of human labor and the generation of new work activities that demand this labor. Economic literature on the impact of technological change on employment has primarily treated the set of human job tasks as finite and static, implying that as automation proceeds, labor is shunted into an ever-narrowing scope of activities, as in Susskind (2020). Casual observation and historical evidence suggest the opposite: as employment in labor-intensive sectors such as agriculture, textiles, and mining has eroded, the scope and variety of labor-demanding activities has arguably expanded, e.g., in medicine, software, electronics, healthcare, finance, entertainment, recreation, personal care, and other domains. Recent work by Acemoglu and Restrepo advances the theoretical frontier on this topic (Acemoglu and Restrepo, 2018b; Acemoglu and Restrepo, 2019), but a major empirical challenge remains: there is almost no direct, consistent measurement of either the emergence of new work tasks within occupations and industries, or of the technological and economic forces that are hypothesized to give rise to them.

This paper systematically studies the nature, sources, and consequences of the emergence of new work in the United States between 1940 and 2018. We seek to consistently measure the substantive content of new work over eight decades, to explore the forces that explain when and where new work emerges, and to assess whether, as hypothesized in recent literature, new work exerts a countervailing force to the employment-eroding effects of task-displacing automation. Our empirical analysis is motivated by a two-sector general equilibrium task model that, building on Acemoglu and Autor (2011), Acemoglu and Restrepo (2018a), and Acemoglu and Restrepo (2018b), draws economic linkages between task automation, new task creation, innovation incentives, and the locus and attendant skill demands of new work. Because our main contributions are empirical, the formal model is developed in the Theory Appendix (Section J), with intuition provided in the body of the paper.

Our analysis is grounded in three constructs that we operationalize using purpose-built data. The first is *new work* itself, by which we mean the introduction of new job tasks or job categories requiring specialized human expertise. Following path-breaking work by Lin (2011), we construct a database of new job tasks introduced over eight decades, sourced from nearly a century of internal reference volumes used by U.S. Census Bureau employees to classify the free-text job descriptions of Census respondents into occupation and industry categories in each decade. This *Census Alphabetical Index of Occupations and Industries* (CAI hereafter) is updated during the processing of each decade's Census to reflect new write-in titles ('micro-titles') detected by Census coders.[1] Following Lin (2011), we track the emergence of new micro-titles in each decade by comparing successive editions of the CAI. Consider, for example, the micro-titles of "Technician, fingernail" (added in CAI 2000) and "Solar photovoltaic electrician" (added in CAI 2018), which highlight two salient attributes of new work. First, new work typically requires *expertise* acquired through study, apprenticeship, or practical experience (e.g., in cosmetology, in the electrical trades), though the extent of expertise clearly varies across categories. Second, new work often reflects the development of novel expertise within existing work activities (e.g., electrical trades skills specific to solar installations) or through an increase in the market scale of a niche activity (e.g., nail care) rather than a fundamentally new human endeavor.[2]

The second and third constructs that we operationalize are the flow of *augmentation* and *automation* innovations over eight decades. In our terminology, augmentation innovations are technologies that increase the capabilities, quality, variety, or utility of the *outputs* of occupations, potentially generating new demands for worker expertise and specialization. Conversely, automation innovations are technologies that substitute for the labor inputs of occupations, potentially replacing workers performing these tasks. We construct both augmentation and automation measures using natural language processing tools (NLP) to map the full text of all U.S. utility patents issued between 1920 and 2018 to the domain of occupations. Following methods introduced by Kogan et al. (2021), we represent both patent documents and occupational descriptions as weighted averages of word embeddings, which are geometric representations of word meanings. We characterize the semantic closeness between patent texts and occupational descriptions by measuring their proximity in embedding space. Each patent may be classified as an automation innovation, an augmentation innovation, both, or neither.

To capture *augmentation* innovations, we harness the full set of micro-titles in the CAI associated with each macro-occupation in each decade (both pre-existing and new titles) to provide a text corpus describing the occupation's *outputs*. For example, in 1999, the U.S. Patent and Trademark Office (USPTO) granted patent US5924427A for a "Method of strengthening and repairing fingernails". Our procedure links this patent to the Census macro-occupation of "Miscellaneous personal appearance workers", which encompasses the micro-title of "Technician, fingernail". Similarly, our algorithm links the 2014 patent US7605498B2 "Systems for highly efficient solar power conversion"

[1]While Census tabulations and public use data sources report several hundred distinct occupation and industry codes in each Census year (which we call 'macro-titles'), these titles reflect concatenations of approximately 30,000 occupational and 20,000 industry-level 'micro-titles' enumerated in the Census Alphabetical Index of Occupations and Industries in each decade between 1930 and 2018 (US Census Bureau).

[2]There are exceptions: in the 1900 Census, Orville and Wilbur Wright each reported their occupation as "Merchant, bicycle"; in the subsequent U.S. Census, they both reported "Inventor, aeroplane" (United States Census, 1900; United States Census, 1910). Four decades later, the title of Airplane Designer becomes prevalent enough to gain its own entry in the 1950 CAI.

to the macro-occupation "Electrical and electronics engineers" which includes the micro-title "Solar photovoltaic electrician".

To capture *automation* innovations, we follow Webb (2020), Dechezleprêtre et al. (2021), Kogan et al. (2021), and Mann and Püttmann (2021) in identifying similarities between the content of patents and the tasks that workers perform in specific occupations—i.e., their work inputs. We measure these task inputs using the 1939 *Dictionary of Occupational Titles* (DOT, hereafter) for the 1940–1980 period, and the 1977 DOT for the 1980–2018 period (U. S. Department of Labor, Employment and Training Administration, 1939; U. S. Department of Labor, Employment and Training Administration, 1977), again calculating semantic similarities with the full corpus of utility patents issued between 1920 and 2018. For example, in 1979, the USPTO granted patent US4141082A for a "Wash-and-wear coat". Our algorithm links this patent to the macro-occupation of "Laundry and dry cleaning workers". Similarly, our algorithm links the 1976 patent US3938435A, "Automatic mail processing apparatus", to the macro-occupation of "Mail and paper handlers". It bears emphasis that we apply fully parallel procedures for classifying augmentation and automation innovations, with the sole difference being the corpora of text used to characterize occupational outputs versus inputs. Accordingly, the distinct classifications we generate for augmentation and automation innovations stem entirely from differences in semantic content between the corpora of task inputs and occupational outputs rather than procedural artifacts.

The empirical analysis proceeds in three steps. We first characterize the substantive content of new work. We quantify its (approximate) contribution to overall employment growth between 1940 and 2018, and document how its occupational composition and educational requirements have evolved across the two four-decade intervals of the sample. We estimate that the majority of current employment is found in new job specialties ('new work') introduced after 1940. But the locus of new work creation has substantially shifted over the eight decades of our sample. Between 1940 and 1980, a substantial share of new work accrued in middle-paid production and clerical occupations. In the more recent four decades, new work has primarily emerged in high-paid professional occupations and, secondarily, in low-paid service occupations.

The second step analyzes where new work comes from. We posit that new job tasks (i.e., new work) derive from two primary sources. A first source is *augmentation* innovations, meaning the introduction of new processes and products (e.g., solar voltaic cells), new services (e.g., fingernail hardening), and entirely new products or industries (e.g., dry-cleaning, commercial air travel), that create novel demands for expertise and specific competencies that correspond to new work tasks and are reified in new job titles in our empirical analysis. Conversely, if automation innovations primarily enable machines to subsume existing tasks, as per Acemoglu and Restrepo (2018a) and Acemoglu and Restrepo (2018b), these innovations should *not* spur the introduction of new labor-using tasks. (While quite plausibly, automation innovations give rise to new *machine* tasks, we are not able to measure such tasks.)

Consistent with this reasoning, we find that augmentation and automation innovations have distinct, asymmetric relationships to the creation of new work. Augmentation innovations strongly predict the locus of new task creation, as measured by the emergence of new job titles, across occupations and over time. By contrast, automation innovations do not predict where new work emerges. Since, as we show below, augmentation and automation innovation flows are strongly *positively* correlated at the occupational level, this is a stringent test.

The second driver of new work that we consider is fluctuations in market size that increase or depress the value of occupational outputs. We document that the flow of new work responds elastically to shifts in demand for occupational output. Adverse demand shocks, which we identify using the widely studied China trade shock (Autor, Dorn, and Hanson, 2016), slow the emergence of new work tasks in exposed occupations, even conditional on exposure to augmentation innovations and contemporaneous changes in employment. Conversely, positive occupational demand shocks, which we identify using shifts in demographic structure following DellaVigna and Pollet (2007), accelerate the emergence of new work in exposed occupations. These demographic shifts help account for the emergence of new job types in lower-paid personal services that have been relatively immune to labor-augmenting innovations but are nevertheless a key locus of new work creation for non-college workers over the last four decades.

Our final set of results turns the focus from new work—meaning new occupational titles—to shifts in overall occupational labor demand. While it is easy to verify that new tasks differentially emerge in growing occupations—or, equivalently, that employment differentially expands in occupations in which new tasks are emerging—this correlation does not directly test the role of innovation in spurring or eroding employment. To provide that test, we regress occupational employment and wagebill changes over two four-decade periods on measures of the flow of augmentation and automation innovations to which these occupations are exposed. We document that employment and wagebills grow in occupations exposed to augmentation innovations and contract in occupations exposed to automation innovations. The pattern of results further suggests that the demand-eroding effects of automation innovations have intensified in the last four decades while the demand-increasing effects of augmentation innovations have not—though we stress that our innovation exposure measures lack sufficient cardinality to draw definitive conclusions.

Since both innovation and employment are plausibly jointly determined at the occupational level, it is critical to ask whether these findings capture causal relationships. We test for causality by developing an instrumental variables strategy that identifies shifts in the flow of augmentation and automation innovations stemming from breakthrough innovations occurring two decades earlier. Using these breakthrough innovations as instruments for subsequent downstream patent flows, we establish that augmentation innovations *cause* the emergence of new work and the growth of occupational employment, while conversely, automation innovations do not spur new work but do erode occupational employment.

Our work contributes to three branches of literature on technology, skills, and employment. A first studies the interplay between supply, demand, technologies, and institutions in shaping the long-run evolution of skill demands, occupational structure, and wage inequality.[3] A foundational assumption of this literature is that technological change has non-neutral impacts on the skill composition of labor demand. We contribute to this literature by linking changes in the structure of occupational demands to the shifting locus of innovation over eight decades, showing that: new work is a quantitatively large contributor to aggregate employment change; that it emerges where innovative activity is focused; and that the locus of new work generation has shifted across recent decades, leading overall changes in occupational structure.

We secondarily contribute to a contemporary literature that explores how automation technologies substitute for existing work, as measured by occupational structure or job tasks.[4] Our work is closely related to Webb (2020), Dechezleprêtre et al. (2021), Mann and Püttmann (2021), and most directly Kogan et al. (2021), who employ NLP tools to identify innovations recorded in patents that potentially substitute for the tasks performed by workers, as well as papers by Brynjolfsson and Mitchell (2017), Brynjolfsson, Mitchell, and Rock (2018), Eloundou et al. (2023), Felten, Raj, and Seamans (2018), Felten, Raj, and Seamans (2019), and Felten, Raj, and Seamans (2023) that predict which occupational tasks can be performed by artificial intelligence. Distinct from this literature, we develop a method to identify innovations that generate *new* work tasks by *augmenting* occupational outputs.

Most directly, our paper contributes to research on the micro- and macroeconomic origins of new work, including Goldin and Katz (1998), Lin (2011), Acemoglu and Restrepo (2018b), Acemoglu and Restrepo (2019), Atack, Margo, and Rhode (2019), Frey (2019), Atalay, Phongthiengtham, et al. (2020), Atalay and Sarada (2020), and Deming and Noray (2020), and Kim (2022). We extend the pioneering work in Lin (2011), while expanding its scope to provide direct, representative, and time-consistent measurement of new task creation across eight decades.[5] Distinct from any prior work that we are aware of, we identify innovations that complement occupational outputs, which we hypothesize (and empirically confirm) spur new task creation. This technique enables us to document that automation and augmentation have measurable and distinct impacts on both new work creation and occupational employment.

The paper proceeds as follows. Section II details our methods for identifying new work, describes

[3]This vast literature includes Goldin and Margo 1992; Katz and Murphy 1992; DiNardo, Fortin, and Lemieux 1996; Acemoglu 1998; Autor, Katz, and Krueger 1998; Katz and Autor 1999; Krusell et al. 2000; Card and Lemieux 2001; Goldin and Katz 2008; Autor, Goldin, and Katz 2020; Haanwinckel 2020; Vogel 2023.

[4]Relevant works include Autor, Levy, and Murnane 2003, Autor, Katz, and Kearney 2006, Goos and Manning 2007, Goos, Manning, and Salomons 2009, Acemoglu and Autor 2011, Autor and Dorn 2013, Goos, Manning, and Salomons 2014, Michaels, Natraj, and Van Reenen 2014, Akerman, Gaarder, and Mogstad 2015, Bárány and Siegel 2018, Cortes et al. 2020, Acemoglu and Restrepo 2022, and Böhm, Gaudecker, and Schran 2022.

[5]In related work, Acemoglu and Restrepo (2019) develop a set of ingenious proxies for the appearance of new work based on changes in labor share and in the mix of 3-digit occupations (what we call 'macro-occupations') within industries. Atalay, Phongthiengtham, et al. (2020) and Atalay and Sarada (2020) and Deming and Noray (2020) measure the appearance of new work by analyzing the text of job advertisements. Kim (2022) finds that greater import competition spurs a relative increase in employment in managerial occupations (which are intensive in new work).

how the locus of new work has evolved over 1940–2018, and outlines how we identify augmentation and automation innovations embodied in patents that link to specific occupations. Section III briefly sketches our theoretical model, which is formalized in the appendix. Section IV uses both OLS and instrumental variables methods to test the relationship between innovation, demand shifts, and new work creation. The final empirical section of the paper, Section V, assesses whether augmentation and automation innovations have measurable, countervailing effects on occupational employment and wagebill growth. Section VI concludes.

II Data and Measurement

Our analysis links data on the emergence of new titles, the flow of augmentation and automation innovations, and the evolution of employment and earnings within industry-occupation cells over eight decades, 1940–2018. To characterize employment and earnings by occupation and industry, we use IPUMS Census samples for 1940 through 2000 and Census ACS samples for 2014–2018 (Ruggles, Flood, et al., 2022). Our databases of new titles, augmentation flows, and automation flows are purpose-built for this study and we document them here.

II.A Measuring new work

We leverage Census Bureau historical coding volumes for occupations and industries for the years 1930 through 2018 (industry data starts in 1940) that are released each decade by the Census Bureau as the *Census Alphabetical Index of Occupations and Industries* (CAI). Each decade's Index contains approximately 35,000 occupation and 15,000 industry 'micro' titles, each classified to a more aggregated ('macro') Census occupation or industry code. These indexes serve as reference documents for Census coders who classify individual Census write-ins for job title and industry of employment, which are always reported as free text fields. This process has been performed consistently for occupations since 1900 and is illustrated for occupations in the American Community Survey (ACS) in Figure A.VIII. Unfortunately, the Census Bureau does not systematically remove potentially extinct titles from the CAI (since they may occasionally reoccur), meaning that the CAI is not suitable for measuring work obsolescence.

When Census coders encounter multiple instances of a write-in occupational title that cannot be ascribed to an existing micro-title, they bring it to the attention of Census bureau managers, who perform an internal review to determine whether the title is sufficiently distinct and prevalent to be added to the Index. For example, the micro occupation title "Artificial intelligence specialist" was added to the CAI in 2000 and classified to the broader Census ('macro') occupational title "Computer scientists and systems analysts", which appears in published Census tabulations and public use data sets. The two stage process for adding a new title—detection by coders, review by managers—clarifies why, for example, "Mental-health counselor" was added in 1970, "Artificial intelligence specialist" in 2000, and "Sommelier" in 2010. Although workers performed each of these jobs in earlier decades, these specializations were too rare to warrant inclusion beyond a generic counselor, computer scientist, or restaurant server title. To provide concrete illustration, Appendix D2 enumerates examples of new titles introduced between 1940 and 2018 within two office occupations (Office machine operators, Stenographers, typists, and secretaries), one professional occupation (Dentists), and one blue-collar occupation (Automobile mechanics). New title additions typically reflect emerging expertise demands for new technologies, changes in educational requirements, and in the case of dentists, the addition of new clientele (Pediatric orthodontist) and the provision of entirely new types of services (e.g., Maxillofacial surgeon).

The Census Bureau does not highlight or separately list titles that are newly-added to the CAI, and it frequently renames outdated job titles, adds new phrasings of existing titles, and removes gendered forms (e.g., Key boy). Hence, consistently tracking the flow of new CAI titles requires a mixture of automated and manual steps. We compare title lists across successive CAI decades using fuzzy matching and extensive manual revision to discard false positives stemming from rewording, reformatting of the index, and title additions that do not reflect discernible modifications to preexisting counterparts. We do not, for example, count "Software applications developer" as a new micro-occupation since "Software developer" was already present when this title appeared. Our overarching aim is to retain new titles that reflect plausibly new work activities, meaning that they add a particular task specialization, work method or tool, or professional or educational requirement (Appendix D1 provides details.)

Table I provides examples of the diversity of micro-titles added to the CAI in each decade from 1940 through 2018 (where, for example, new titles in the 1950 CAI reflect those detected by Census staff between 1940 and 1950.) The left-hand column reports titles that are plausibly associated with new or evolving technologies: "Airplane designers" in 1950, "Engineers of computer applications" in 1970, "Circuit layout designers" in 1990, and "Technicians of wind turbines" in 2010. Many new titles do not have immediate technological origins, however. As shown in the right-hand column of Table I, these titles appear to reflect changing tastes, income levels, and demographics, including "Tattooers" in 1950, "Hypnotherapists" in 1980, "Conference planners" in 1990, and "Drama therapists" in 2018. These examples motivate looking beyond exclusively technological forces, as we do below, in analyzing the sources of new work creation.

How representative of new work is the CAI-based measure? Absent a comparable source against which to benchmark it, we provide a novel test of external validity by confirming that the CAI captures new occupational titles as they enter common usage in the English language. Using the Google Ngram viewer (Michel et al., 2011), we calculate the frequency in digitized English language books of both new and preexisting occupational titles in each year between 1940 to 2018.[6] The eight panels of Figure A.I report the usage frequency of all titles added to the CAI in each decade relative to those present at the start of the decade. Frequencies are normalized as the ratio of instances of the new title (of length n) in a year (excluding unmatched titles) relative to the total number of n-grams in that year. In each cohort of new titles, the median new title is added to the

[6]We thank Peter Lambert of LSE for suggesting this analysis.

CAI during approximately the decade before or during which it attains peak usage in the English language, suggesting that the CAI provides a somewhat conservative enumeration of new work.[7] Figure A.V further corroborates that this pattern holds across four broad occupational categories that encompass the totality of employment: blue-collar occupations; professional and information occupations; personal service occupations; and business service occupations. (Appendix C details these occupational groups.) These findings substantiate that our database built from successive CAI editions provides a representative, time-consistent measure of the emergence of new job titles in the U.S. over the eight decades of the sample.

II.B The evolution of new work, 1940–2018

What do these data tell us about the evolution of new work? A first finding is that new work is quantitatively important. Our estimates imply that the majority of contemporary work as of 2018—roughly 60%—is found in new job titles added since 1940. Figure I reports the contribution of new work to overall employment between 1940 and 2018 in twelve exhaustive, mutually exclusive broad occupational categories. These are ordered from lowest to highest-paying, with farming and mining occupations on the left-hand side of the scale and managerial workers on the right-hand side. We further distinguish between 2018 employment found in occupational titles that existed in 1940 versus 2018 employment in occupational titles that were added thereafter. While the majority of contemporary work is 'new' since 1940, the share of employment in new work differs substantially by occupational category. Among professionals—the occupation that added the most workers during these eight decades—this share is 74 percent. For the production occupation—which added the second smallest number of workers since 1940 (after farming)—this share is 46 percent. Because the Census Bureau does not record or report the count of respondents within micro-titles, the estimates in Figure I should be understood as approximate. We follow Lin (2011) in imputing employment counts in new titles by multiplying the *new title share* in each macro-Census occupation—equal to the ratio of new micro-titles to total micro-titles within a Census occupation—by total employment in the occupation.[8]

A second result is that the occupational distribution of new work has changed markedly over the past eight decades, and in a manner that presages the changing overall shape of employment growth and skill demands. This is shown in Figure II, which plots the contrast between the flow of new work and the stock of preexisting work during 1940–1980 and 1980–2018 for both non-college

[7] Titles that become newly prominent in one era do not remain at peak prominence in perpetuity, however, typically declining by 20% to 50% in the three to five decades after entering the CAI.

[8] Lin (2011) validates this approach using a special version of the April 1971 Current Population Survey where a subsample of workers are assigned DOT titles. We further explore this imputation in Appendix H2 using Census Complete Count data for 1940, which contains both macro-titles and the free text write-in micro-titles supplied by Census respondents. The count of workers in new titles is strongly increasing in the new title share—though the slope is below one—and this relationship is more precise when using ordinal share ranks rather than cardinal shares.

workers (high school degree or below) and college workers (BA or above).[9] Between 1940 and 1980, the flow of new work largely replicated the stock of existing work. The subsequent four decades display a marked contrast: there is a sharp decline after 1980 in the flow of new work relative to the stock of existing work in traditional middle-skill, middle-pay occupations, including construction, transportation, production, and clerical and administrative support. This pattern is consistent with the widely documented phenomenon of 'occupational polarization' unfolding in these decades (Autor, Katz, and Kearney, 2006; Goos and Manning, 2007; Autor and Dorn, 2013; Goos, Manning, and Salomons, 2014; Michaels, Natraj, and Van Reenen, 2014; Autor, 2019).[10]

A third finding is that the pattern of polarization is distinct for college and non-college workers. Among college-educated workers, the entirety of the decline in new middle-skill occupations is accounted for by a corresponding rise in employment in new professional occupations. Among non-college workers, however, most of the decline in new middle-skill occupations is offset by a sharp increase in new low-paid personal and health service occupations, accompanied by a modest increase in new professional occupations. Thus, the emergence of new work over the last four decades has led the overall polarization of occupational structure. By implication, employment polarization does not merely reflect an erosion of employment in existing middle-skill work but also a change in the locus of new work creation.

II.C Measuring augmentation and automation innovations

Critical to analyzing the relationship between innovation and new work is identifying innovations that potentially complement worker outputs (augmentation) and substitute for worker inputs (automation). We draw on three data sources for this task: the corpus of all U.S. utility patents issued between 1920–2018; the CAI (discussed above); and the Dictionary of Occupational Titles (DOT) discussed below. The patent data provide a lexicon of potential augmentation and automation innovations for the period of interest. Patent data present the best substantively detailed, time-consistent measure of the flow of innovative activity in the U.S. economy, despite well-known limitations—including not capturing the totality of technological innovation (Moser, 2012) and over-representing innovations that are difficult to protect as trade secrets.[11]

Identifying patents that may substitute for workers' occupational inputs (automation innova-

[9] Analogous to the procedure for Figure I, we estimate the average educational attainment of workers employed in new titles by taking the average of educational attainment of all workers in their macro occupation-by-industry cell and weighting across all macro-titles to aggregate to broader occupation categories (where using respondents' macro-occupation and macro-industry to assign characteristics increases the specificity of the imputation).

[10] Figure A.VI documents that between the first and second halves of out sample, there is a sharp absolute decline in the flow of new work in traditional middle-skill, middle-pay occupations.

[11] Patent texts were obtained by Kelly et al. (2021) from the USPTO patent search website for patents issued from 1976–2015, and from Google Patents prior to 1976. We extend the Kelly et al. (2021) sample of patents issued from 2015 to 2018 by scraping the patent text from the Google Patents website. We also scrape Google Patents for the issue date and primary three-digit Cooperative Patent Classification (CPC) code for each patent in our sample. Prior to 1976, each patent document comprised a single block of text. Subsequently, patent texts were divided into abstract, description, and claims sections. We use the entire text of each patent.

tions) requires a text corpus that characterizes the tasks that workers perform in their jobs. For this purpose, we follow Webb (2020) and Kogan et al. (2021) in employing the U.S. Bureau of Labor Statistics' *Dictionary of Occupational Titles* (DOT hereafter) (U. S. Department of Labor, Employment and Training Administration, 1939; U. S. Department of Labor, Employment and Training Administration, 1977), which describes in detail the primary duties of workers in each occupational title. To avoid capturing occupational outputs, we use only task descriptions from the DOT, purging any occupational titles contained in these descriptions. Consistent with Webb (2020) and Kogan et al. (2021), we expect that patents that overlap with workers' occupational tasks reflect innovations that may potentially replace workers in these tasks.

Identifying patents that may complement workers' occupational outputs (augmentation innovations) requires a text corpus that characterizes the outputs of each occupation. We harness the CAI, which supplies tens of thousands of occupation and industry micro-titles in each decade's CAI. We hypothesize that patents that link to the CAI corpus reflect innovations that increase the capabilities, quality, variety, or utility of the outputs of occupations, potentially generating new demands for worker expertise and specialization. This step precisely parallels our method above for identifying automation innovations, but we are not aware of any prior work that systematically characterizes innovations that are complementary to labor. The empirical success of this method (demonstrated below) is a key contribution of the paper.

To link occupations to relevant patents, we follow Kogan et al. (2021), who measure textual similarities between patent texts and job descriptions using word embeddings, which are geometric representations of word meanings. Distinct from conventional measures of text similarity (e.g., the commonly used bag of words approach found in Gentzkow, Kelly, and Taddy 2019; Atalay, Phongthiengtham, et al. 2020, and Kim 2022), word embeddings locate closely related words nearby to one another in embedding space. This is valuable since patent texts and occupational descriptions often use dissimilar terms for related concepts. As an example, the verbs *research* and *quantify* are not synonymous, nor do they have common etymology, but each is among the three closest relatives of the other in English language embedding space.[12]

Figure III provides a schematic overview of this procedure, which we detail in Appendix E. Briefly, our procedure cleans each patent, CAI, and DOT text *document* by removing prepositions and other stop words and retaining nouns and verbs; represents each document as a Term Frequency-Inverse Document Frequency (TF-IDF) weighted average of the word embeddings for each of the words in the document; calculates the matrix of cosine similarity scores among the document vectors for each patent issued in a decade and the CAI and DOT document vector for each occupation (DOT) and occupation-by-industry (CAI) cell; retains the top 15% most similar patent-occupation (DOT) and patent-industry-occupation (CAI) matches in a decade; and sums the weighted count of top 15% matched patents for each occupation-decade (or occupation-industry-

[12]Calculated using http://vectors.nlpl.eu/explore/embeddings/en/associates/# on the English Wikipedia corpus with queries `research_VERB` and `quantify_VERB` on 2022/06/24.

decade), with weights equal to each patent's cohort-specific relative citation frequency. Because it is infeasible to construct a fully balanced panel of detailed occupations over the eight decades of 1940–2018 without sacrificing substantial resolution, we instead create two balanced panels that cover the first and second halves of our sample. Appendix C provides details, and Table A.VI reports descriptive statistics on these augmentation and automation exposure measures.[13] We emphasize that the links we establish between patents and occupations have a many-to-many structure: a given patent may have an augmentation link to one occupation and an automation link to another (or even to the same occupation that it is augmentation-linked). Most occupations are linked to dozens—and in some cases hundreds or thousands—of patents in each decade, so no single patent plays an outsized role in determining the augmentation and automation exposure of any given occupation. It is the totality of augmentation and automation links drawn for each occupation that provides identification.

II.D Occupational exposure to innovation

Figure IV previews the substantive content of the automation and augmentation exposure measures by plotting the bivariate relationship between percentiles of each at the level of consistent three-digit ('macro') occupations over 1940–1980 (166 occupations) and 1980–2018 (306 occupations), where the occupation-level augmentation and automation exposure are averaged within each of the two four-decade periods. Immediately evident from the figure is that occupations that are more exposed to augmentation are on average also more exposed to automation. The employment-weighted cross-occupation correlation between augmentation and automation exposure is 0.74 over 1980–2018 and 0.63 over 1940–1980. This positive correlation is expected as many technologies contain both automation and non-automation components. It is also consistent with our conceptual framework, sketched below, which highlights that demand forces that raise the value of occupational output create incentives for the introduction of both augmentation and automation innovations. (This may in part also reflect the limitations of our classification procedure for cleanly distinguishing between augmentation and automation innovations.)

While the positive occupational correlation between augmentation and automation exposure is prominent in Figure IV, it is the off-diagonal components of this figure that ultimately provide identifying variation for the analysis that follows. Between 1980–2018, the macro-occupations of "Radiologic technologists and technicians", ""cabinetmakers and bench carpenters", and "Machinists" all had high rates of automation relative to augmentation exposure. The same applies to compositors and typesetters, elevator operators, and telegraph operators in the prior four decades of 1940–1980. Our conceptual framework predicts that employment in these occupations would

[13] When estimating occupation-level models, we establish patent linkages to CAI occupation titles alone; and when estimating occupation-by-industry models, we additionally leverage CAI industry titles to establish patent matches. Unlike the CAI, the DOT contains only occupation-level textual information, so automation exposure measures are always defined at the occupation level.

tend to erode. Conversely, augmentation outpaced automation between 1980 and 2018 in the occupations of industrial engineers, operations and systems researchers and analysts, and (to a lesser degree) managers and specialists in marketing, advertising, and PR, and business and promotion agents. In the prior four decades, shipping and receiving clerks, buyers and department heads, civil and aeronautical engineers, and, to a lesser extent, bookkeepers were all more exposed to augmentation than automation. We would expect these occupations to expand in these subperiods. Conversely, the on-diagonal examples capture occupations with a relatively 'balanced' degree of exposure to both automation and augmentation, either because they are highly subject to both forces (e.g., assemblers of electrical equipment over 1980–2018 and office machine operators over 1940–1980) or because they are relatively insulated from both (e.g., clergy and religious workers in both periods).

Inspection of patents that are augmenting for one occupation but automating for another reveals further logical relationships. For example, patents issued between 2010–2018 that are *augmentation*-linked to the macro-occupation "Computer systems analysts & computer scientists" are most often *automation*-linked to occupations in technicians or clerical & administrative support—occupations which are particularly susceptible to software-based automation during this period of rapid digital growth. Table A.I enumerates examples of individual patents that are augmenting for "Computer systems analysts & computer scientists" yet automating for other occupations. Example patents include "Methods and apparatus for storing confidential information", which is automation-linked to "Billing clerks & related financial records processing"; "Direct connectivity system for healthcare administrative transactions", which is automation-linked to "Health record technologists & technicians"; and "System and method for securing data", which is automation-linked to "Office machine operators, computer and peripheral equipment operators, and other telecom operators". We expect that technologies developed in these patents can be harnessed by computer scientists to automate tasks in linked occupations.

Figure A.IV provides insight into augmentation and automation potential at the *technology* level, distinct from the occupation-level as above. This figure plots the augmentation *shares* of all linked patents within each three-digit CPC technology class on a zero to one scale, where a share of zero indicates all class patents are automation innovations and a share of one indicates that all class patents are augmentation innovations. These share data are organized into nine broad technology classes and are reported separately for 1940–1980 and 1980–2018. The broad class of Instruments and Information technologies contains on average the most augmenting detailed 3-digit CPC classes, followed by Electricity and Electronics, and Consumer Goods and Entertainment. Conversely, Engineering, Construction, and Mining is the most automating broad class, followed by Chemistry and Metallurgy, Health and Agriculture, and Manufacturing Process. There is also substantial variation in augmentation shares within broad classes and over time: innovations in Transportation and in Lightning, Heating, and Nuclear have become less augmenting in the most recent four decades relative to the prior four, whereas the opposite is true for Instruments and Information, and Electricity and Electronics. As detailed in Section IV.B below, these secular shifts in the locus of patenting across broad technology classes that have occurred over the last eight decades affect the intensity of augmentation and automation innovations and shape the set of occupations exposed to these forces.

III Theoretical Motivation

To formalize the intuitions that guide the empirical analysis, we provide a model in Appendix J that considers how three forces shape the endogenous creation of new job tasks, the elimination of old tasks, and the demand for labor at the level of occupations. These forces are: augmentation, which generates new labor-using job tasks; automation, which reallocates existing tasks from labor to capital; and shifts in consumer demand, which affect task automation and new task creation by changing innovation incentives. Our framework generalizes the single-sector setting in Acemoglu and Restrepo (2018b) to two sectors with different skill-intensities, serving two purposes. First, it allows us to consider the implications of augmentation and automation for occupational demand (where sectors represent occupations) rather than exclusively for aggregate labor demand. Second, the interaction of these sectors is central for considering the impact of demand shifts on new work creation.

We posit that new job tasks (i.e., new work) derive from two primary sources. A first is augmentation innovations, meaning new processes and products (e.g., solar voltaic cells), new services (e.g., fingernail hardening), and entirely new products or industries (e.g., dry-cleaning, commercial air travel) that create new demands for expert knowledge and competencies, which are reflected in new job titles in our empirical analysis. Second, even absent specific technological advances, our conceptual framework implies that by raising the value of occupational outputs, positive demand shocks—stemming for example from demographic shifts—create incentives for entrepreneurs to introduce both augmentation and automation innovations.[14] Formally, sectoral demand shocks in the model raise the value of sectoral output, spurring entrepreneurs to introduce both new augmentation and automation innovations in that sector. Further, by generating sector-specific capital scarcity, these shocks create an additional incentive for entrepreneurs to prioritize new augmentation innovations, since these are labor-using and capital-saving (as in Acemoglu and Restrepo 2018b).

[14]For example, the Bureau of Labor Statistics reports that Solar energy installation managers (O*NET code 47-1011.03) is a recently emerging occupation whose primary task is to "Direct work crews installing residential or commercial solar photovoltaic or thermal systems" in the rapidly expanding renewable energy sector (U.S. Department of Labor: Employment and Training Administration, 2022). This title is in turn a specialization of a broader occupational category, First-line supervisors of construction trades and extraction workers (O*NET code 47-1011.00), that is also commonplace in this sector. This title exemplifies the intuition that sectoral demand shifts generate both *more* work—here greater demand for First-line supervisors—and *new work*, more specialized sectoral expertise, here Solar energy installation managers. For negative demand shifts—stemming from for example import competition— we expect the opposite effect, such that new work emergence decelerates. We test for the link between demand shifts and new work emergence, holding constant the direct effect of augmentation innovations, in Section IV.C.

These mechanisms give rise to two empirical implications:

1. Augmentation innovations spur the emergence of new work whereas automation innovations do not. This asymmetry has both a conceptual and empirical foundation. Conceptually, automation should primarily displace workers from existing tasks rather than directly creating new work tasks. Empirically, our new work measure enumerates new job titles as they are captured by the Census Bureau, but it does not detect titles that go out of common usage since these are not systematically purged by the Census Bureau. We test for these asymmetric effects in Section IV.

2. Positive demand shocks spur the introduction of new occupational titles and, further, generate a positive cross-occupation correlation between the flow of augmentation and automation innovations (visible in Figure IV).

Our final set of predictions concerns the relationship between innovation and occupational labor demand. Augmentation innovations unambiguously raise occupational labor demand in our conceptual framework since they both create new labor-using tasks that raise (relative) demand for labor (a substitution effect) *and* increase the value of occupational services, yielding a complementary scale effect. Conversely, substitution and scale effects are partly offsetting in the case of automation innovations: automation of labor-using tasks reduces employment via the substitution effect while the scale effect (stemming from higher productivity) pushes in the opposite direction (as in Acemoglu and Restrepo 2018b). The structure of consumer preferences in our model guarantees that the substitution effect dominates: automation innovations spur a relative contraction in labor demand in exposed occupations, opposite to the case of augmentation innovations. Section V tests whether, despite their positive cross-occupation correlation, augmentation and automation innovations have measurable, countervailing effects on occupational employment and wagebills (i.e., the product of employment and wages).

This framework implies that the positive effects of augmentation flows and demand shifts on the emergence of new work are causal, as are the countervailing effects of augmentation and automation flows on occupational labor demand. We employ two instrumental variables strategies to test causality. To assess the causal effects of demand shocks on new title emergence, we exploit exogenous occupational labor demand shifts stemming from trade shocks and demographic changes. To estimate the causal effect of augmentation and automation innovations on occupational labor demand, we isolate unanticipated flows of occupational augmentation and automation innovations stemming from novel upstream innovations occurring in prior decades.

IV The Effects of Innovation and Demand Shifts on New Work Creation

This section empirically characterizes the forces that explain where and when new job tasks emerge by relating the emergence of new occupational tasks to the exposure of occupations to: (1) augmentation innovations; (2) automation innovations; and (3) positive and negative demand shifts. Our focus here is on forces that affect the creation of *new work*, measured by the emergence of new titles. We take up the effects of new work creation on occupational labor demand and employment in Section V.

IV.A Do augmentation innovations spur new work?

In our conceptual framework, economic forces that complement occupational outputs lead to the demand for new specialties and expertise reflected in new tasks. One of those forces is augmentation. The first hypothesis that we test is that new job tasks emerge differentially in occupations that are more exposed to augmentation innovations. Using the flow of new job titles as a measure of the emergence of new work, we estimate models of the following form:

$$\ln E\left[\text{Newtitles}_{j,t}\right] = \beta_1 \text{AugX}_{j,t} + \beta_2 \text{AutX}_{j,t} + \beta_3 \frac{E_{j,t-10}}{\sum_j E_{j,t-10}} + \delta_t \left(+\delta_{J,t}\right), \tag{1}$$

where j indexes consistent Census occupations, and t indexes decadal intervals.[15] The dependent variable is a measure of the flow of new work titles emerging in a consistent Census occupation in a decade, and the independent variables of interest are $\text{AugX}_{j,t}$ and $\text{AutX}_{j,t}$, measuring occupational exposure to augmentation and automation innovations respectively, as revealed by textual links between patents and the CAI (augmentation) and DOT (automation). Both the dependent and key independent variables in year t are measured as cumulative flows over the preceding decade: new titles observed in 1950 are those that emerge between 1940 and 1949; augmentation and automation exposure in 1950 are equal to the log count of linked augmentation and automation patents awarded over 1940–1949. Accordingly, β_1 and β_2 can be read as elasticities.

Decade fixed effects absorb temporal variation in the total number of new titles. We control for the employment share of each occupation j at the start of the decade ($E_{j,t-10}/\sum_j E_{j,t-10}$) to remove any mechanical association between new title counts and relative occupational employment size.[16] In some specifications, we further add main effects for the twelve consistently-defined broad occupational groups, indexed by J, and their interaction with decade fixed effects. All regressions are weighted by start-of-decade occupational employment shares. (All core results for new title flows and occupational employment changes, found in Tables II, V, and VI, are qualitatively comparable and statistically robust when estimated without weights.) Based on our conceptual framework, we expect $\beta_1 > 0$: more augmentation-exposed occupations will add more new titles.

The first panel of Table II reports estimates of equation (1) for the full eight decades of the sample, where all coefficients have been multiplied by 100 for readability. Because the outcome

[15] Because the CAI was largely not updated between 2000 and 2010, and was then substantially updated in 2018, the last time interval of our sample is 2000–2018 rather than 2010–2018.

[16] Excluding this control leads to larger point estimates for both the augmentation measure and automation measure but does not qualitatively change the results.

measure in this table, the count of new titles in a macro-occupation in a decade, contains zeros, we fit a negative binomial regression for count data.[17] In subsequent instrumental variables (IV) estimates of these models, we use a combination of logarithmic and linear dependent variable models since IV models for count data are less well developed. As will be apparent, our results are robust to these alternative ways of handling zeros. The first column of Table II reports a specification that excludes major occupation-by-decade main effects. The point estimate of 17.81 (se = 3.52) implies that each 10 percent increment to augmentation exposure predicts an additional 1.8 percent faster rate of new title emergence over the course of a decade. The second column adds 12 occupation dummies and their interactions with decade effects, and hence is identified by variation in new title emergence rates across detailed occupations within broad occupational categories within decades. The point estimate of 21.46 (se = 3.74) is slightly larger than in the prior column while precision is unaffected.

Our conceptual model implies distinct relationships between augmentation versus automation innovations and the flow of new work: augmentation innovations generate new labor-using tasks, reflected in new titles; automation innovations do *not* generate such tasks. The next columns of Table II test this prediction. Column 3 removes the augmentation exposure variable (AugX) and replaces it with the automation exposure variable (AutX) using the specification in column 1. Entered independently, the flow of automation innovations *also* predicts the flow of new titles, with a point estimate on AutX of 12.75, (se = 3.93). This pattern is expected given that automation and augmentation exposures are strongly positively correlated across occupations (see Figure IV). When both innovation measures are simultaneously included in columns 4 and 5, augmentation patents continue to robustly predict the flow of new titles ($\hat{\beta}_1 = 16.85\,(3.96)$ in column 4, and $\hat{\beta}_1 = 21.02\,(3.54)$ in column 5) whereas automation patents have a small and statistically insignificant relationship with new title flows ($\hat{\beta}_2 = 1.89\,(4.52)$ in column 4, and $\hat{\beta}_2 = 2.35\,(4.07)$ in column 5).

Recognizing that count data models may be somewhat functional form dependent, we present a number of alternative estimators in Table A.II, including: Poisson estimates in panel A; OLS log count estimates (the intensive margin, excluding zeros) in panel B; and a linear probability model for the occurrence of any new title (the extensive margin) in panel C. These models support the findings in Table II: augmentation exposure strongly predicts the emergence of new titles on both the intensive and extensive margins; this relationship is generally strengthened by the inclusion of broad occupation by time effects (drawing identification from within broad occupation by decade comparisons); and, conditional on augmentation exposure, automation exposure has no robust predictive relationship with new title emergence and, moreover, is always insignificant and close to zero when broad occupation by decade effects are controlled for. In Appendix I we further show that our findings are robust to instead using Lin (2011)'s measure of new titles for each of the three

[17] The working paper version of this paper (Autor, Chin, et al., 2022) estimated count models using the Inverse Hyperbolic Sine (IHS) transformation. Because subsequent work has underscored the sensitivity of the IHS transformation to the treatment of zeros (Chen and Roth, 2022; Mullahy and Norton, 2022), we use more traditional count models in the final paper. This leaves our results substantively unchanged.

overlapping sample decades (1980–2000). The comparability of findings is reassuring because, for the 1980s and 1990s, Lin (2011) identifies new titles from the DOT rather than the CAI, and obtains newly added titles directly from a Conversion Table of Code and Title Changes available for these two decades only.

Leveraging the fact that we have distinct occupational panels for the two halves of our sample, we report estimates for 1940–1980 and 1980–2018 respectively in panels B and C of Table II. These provide a consistent picture. In all cases, augmentation patents significantly predict faster arrival rates of new occupational job titles. Automation patents are either weakly positive or weakly negative predictors of new title emergence when augmentation patents are simultaneously included (with one exception in column 4 of panel C). The precision of the augmentation estimates rises (with almost no change in magnitude) when broad occupation category by decade dummies are included (columns 2 and 5) to contrast new title flows across detailed occupations within broad categories in each decade. Thus, despite marked differences in the locus of new work emergence across these two four-decade periods (documented in Figure II), augmentation innovations strongly predict where new titles emerge in both eras. A comparison of the point estimates for 1980–2018 versus 1940–1980 does however reveal a substantially shallower, though still robustly significant, elasticity of new title creation with respect to augmentation innovations in the latter half of the sample ($\hat{\beta}_1^{40-80} = 26.11\,(4.45)$ vs. $\hat{\beta}_1^{80-18} = 13.86\,(2.09)$) in the final columns of panels B and C. This pattern hints that the relationship between innovation and new task creation may have slowed in the recent several decades relative to the first four post-WWII decades. This deceleration of the new work creation component of innovation will be a recurrent theme of our analysis, which we synthesize in Section V.B.[18]

Figure V depicts the variation that drives these estimates with a set of bin scatters plotting the (log of) the flow of augmentation innovations within each of the twelve broad occupational categories against the (log of) the flow of new occupational job titles, separately for 1940–1980 and 1980–2018. The predictive relationship between augmentation innovations and the arrival of new titles is a pervasive feature of the data. In eleven of twelve occupations (all but Transportation), and in both halves of the sample, there is a clear, upward-sloping relationship between augmentation innovations and new title flows. (There are not enough detailed occupations with non-zero new titles to estimate this relationship for Health Service occupations in the 1940–1980 period.)

[18] A potential concern with comparing estimated elasticities across time periods is that the CAI's sensitivity at detecting new titles might fluctuate across decades. This should not bias elasticity estimates, however, if such fluctuations affect only the rate of new title capture but not its distribution across occupations. Formally, let the number of new occupational titles captured in decade t be proportional to the true (latent) number of new titles: $Y_{j,t} = \phi Y^*_{j,t}$ where $\phi > 0$. If the structural equation relating to new title flows to augmentation flows is $\ln Y^*_{j,t} = \alpha + \gamma \ln \mathrm{Aug}X_{j,t}$ (as estimated above), then $\frac{\partial \ln Y_t}{\partial \ln \mathrm{AugX}_t} = \frac{\partial \ln Y_t}{\partial \ln Y^*_t} \times \frac{\partial \ln Y^*_t}{\partial \ln \mathrm{AugX}_t} = \gamma$.

IV.B Testing causal relationships: Augmentation, automation, and new work

The new title estimates reflect conditional correlations that may not have a causal interpretation. Our conceptual model, for example, implies that demand shifts favoring a given sector generate both new augmentation and new automation patents, where the former catalyzes the introduction of new titles. This raises the concern that the estimates are tainted by simultaneity bias stemming from unmeasured demand shifts. (Though, if variation in innovation rates and new task creation is primarily driven by demand shifts, we would expect *both* augmentation and automation innovations to predict new title emergence—which is not the case.) To more definitively distinguish causality from correlation requires isolating exogenous flows of augmentation and automation innovations. We sketch a strategy for doing so here.

Our instrumental variables strategy exploits the occurrence of *breakthrough* innovations (Kelly et al., 2021), which represent discontinuous advances in the innovation environment that materialize at a specific point in time. Breakthrough patents are both novel and impactful: novel in that they are conceptually distinct from their predecessors, relying less on prior art; impactful in that they influence future scientific advances, catalyzing a host of follow-on innovations. This chain of reasoning implies that breakthroughs, recorded in patents, can potentially be used to identify exogenous flows of follow-on innovations. The identification assumption is that the precise timing of breakthroughs is not anticipated (conditional on preexisting trends), while the arrival of follow-on innovations is causally affected by the timing of antecedent breakthroughs.

Our implementation of this idea harnesses Kelly et al. (2021)'s breakthrough innovation measure, which operationalizes the novelty and impactfulness of patents using natural language processing tools. Kelly et al. (2021) compare the similarity of each U.S. utility patent granted relative to those that precede it (backward-similarity) to measure its novelty, and compare it to those that follow it (forward-similarity) to measure impact. A breakthrough patent is one that has a high ratio of forward- to backward-similarity, where each is measured over the five years before (backward) or ten years after (forward) the patent date. Our analysis uses the top 10 percent of breakthrough patents granted in each decade, but our results are robust to a range of breakthrough thresholds and forward/backward similarity window lengths.

The two panels of Figure VI illustrate the substantive difference between breakthrough patents and the full population of patents (inclusive of breakthroughs). The figure reports stacked area plots of the distribution of breakthrough and non-breakthrough patents granted in each year between 1900 and 2002 across ten broad, exhaustive, and mutually exclusive technology classes.[19] Panel A shows that the locus of breakthroughs has changed dramatically across decades. In the first two decades of the twentieth century, breakthroughs are concentrated in Engineering, Transportation, and Manufacturing processes. Over the next four decades, between 1920 and 1960, the

[19]These classes are Transportation; Manufacturing process; Lighting, heating and nuclear; Instruments and information; Health; Engineering, construction, and mining; Electricity and electronics; Consumer goods and entertainment; Chemistry and metallurgy; and Agriculture and food.

locus of breakthrough innovations shifts increasingly to Chemistry and metallurgy. After 1960, however, two technology categories grow to dominate breakthrough innovations: Instruments and information, foremost, followed by Electricity and electronics. Panel B reports analogous data for overall patenting. Unlike the pattern for breakthrough patents in panel A, the evolution of *overall* patenting across domains in panel B is relatively muted and slow-moving. Nevertheless, consistent with the expectation that breakthrough patents shift the tide of innovation across domains, the class distribution of the full set of patents appears to echo that of breakthroughs with a lag of approximately two decades.

Panel A of Figure VII confirms this pattern by plotting the predictive relationship between breakthroughs within each of 119 3-digit detailed Cooperative Patent Classification (CPC) technology classes between 1920 through 2000 and subsequent patenting within that class two decades later.[20] This figure documents that prior breakthroughs are a strong predictor of subsequent patent flows, even conditional on prior non-breakthrough patent flows. The nine binscatters in the figure—one for each broad technology class, each plotted with a distinct marker—reveal a consistent upward-sloping relationship between detailed class-level patent flows and breakthrough patents occurring in those classes two decades earlier, with an overall slope of $\hat{\beta}_{t-20} = 0.42$ (0.03).[21]

Might this pattern reflect existing trends whereby breakthroughs occur in already active patenting classes? Panel B of Figure VII replaces the 20-year lagged breakthrough predictor with the 10-year *lead* of breakthroughs in each class, controlling for the corresponding (log) count of non-breakthrough patents as well as decade fixed effects. The estimate of $\hat{\beta}_{t+10} = -.02$ (0.02) is small, imprecise, and opposite in sign to the predictive relationship in panel A. This is consistent with the hypothesis that breakthroughs spur *subsequent* downstream innovations in the technology classes in which they originate, whereas flows of contemporary non-breakthrough innovations do not spur *subsequent* breakthroughs in the same classes.

We exploit this mechanism to isolate exogenous variation in the augmentation and automation patent flows to which each occupation is exposed. This variation provides instrumental variables for the endogenous flows of augmentation and automation innovations used above:

1. Following the textual matching procedure detailed in Section II.C, we form automation and augmentation links for each occupation (and later, for occupation-industry cells) using patents lagged by two decades and occupational description texts lagged by zero to three decades, depending on the data source.[22]

[20]The predictive relationship is estimated by regressing log patent flows by technology class and decade on log breakthrough patent flows lagged by two decades, controlling for log non-breakthrough patent flow counts (also lagged by two decades) and decade fixed effects. We aggregate several small CPC classes to arrive at 119 classes.

[21]The broad classes of Health and Agriculture are combined because the former broad class contains only two detailed 3-digit classes.

[22]Recall that the automation links for 1940–1980 are based on the DOT 1939 volume, while automation links for 1980–2018 are based on the DOT 1977 volume. Implicitly, these automation links are lagged by zero to three decades. Augmentation textual links, by contrast, are based on CAI volumes that are lagged by one decade.

2. For each set of occupation-patent matches from the prior step, we sum the matched patents from two decades earlier that are among the top 10 percent of breakthrough patents for the decade. This is done twice, once for augmentation matches, $N^{Aug,10}_{j,t-20}$ and once for automation matches, $N^{Aut,10}_{j,t-20}$.

3. Occupations are then classified as augmentation breakthrough-exposed if they are above the median of the matched augmentation breakthrough count measure, and similarly for automation breakthrough exposure: $B^{Aug,10}_{j,t-20} = \mathbb{1}\left[N^{Aug,10}_{j,t-20} > \tilde{N}^{Aug,10}_{j,t-20}\right]$, $B^{Aut,10}_{j,t-20} = \mathbb{1}\left[N^{Aut,10}_{j,t-20} > \tilde{N}^{Aut,10}_{j,t-20}\right]$, where tilde denotes the decadal employment-weighted cross-sectional median of a variable within the regression sample. Because manufacturing industries are intrinsically patent-intensive, we compute medians separately for groups of occupations that have above versus below 50% start-of-period employment in manufacturing.

4. To similarly account for non-breakthrough patent exposure, we perform a parallel procedure that identifies occupations that are highly exposed to *all* lagged matched patents in each decade: $B^{Aug,100}_{j,t-20} = \mathbb{1}\left[N^{Aug,100}_{j,t-20} > \tilde{N}^{Aug,100}_{j,t-20}\right]$, $B^{Aut,100}_{j,t-20} = \mathbb{1}\left[N^{Aut,100}_{j,t-20} > \tilde{N}^{Aut,100}_{j,t-20}\right]$.

5. The instrumental variables models estimate a variant of equation (1) for the effect of automation and augmentation exposure on the flow of occupational new titles, where $B^{Aug,10}_{j,t-20}$ and $B^{Aut,10}_{j,t-20}$ serve as instruments for the endogenous variables $\text{AugX}_{j,t}$ and $\text{AutX}_{j,t}$. In this specification, $B^{Aug,100}_{j,t-20}$ and $B^{Aut,100}_{j,t-20}$ serve as controls for overall prior augmentation and automation exposure. As highlighted above, we use 2SLS models for this exercise, distinct from the negative-binomial count models reported in Table II. (Recall that OLS variants of the new title models presented earlier in Table A.II strongly corroborate the new title models using count data models.)

Table III reports first-stage estimates. Columns 1 and 2 tabulate models with only one endogenous variable (AugX in column 1, AutX in column 2) and the corresponding breakthrough instruments. The first-stage coefficient in column 1 of 2.26 (se = 0.41) indicates that, accounting for overall occupational patent exposure, occupations that are above the median of augmentation breakthrough exposure are exposed to approximately 2.3% more augmentation patents two decades later. Column 2 reports a similar relationship for automation exposure, with a first stage coefficient 2.40 (se = 0.32). The third specification in the table, which spans two columns, reports the first stages for augmentation and automation patents estimated jointly. The first stage coefficients for augmentation and automation in this joint specification are closely comparable to the standalone estimates in column 2 while the the cross-instrument predictive relationships are surprisingly modest. In

Absent a consistent occupation scheme for bridging CAI volumes prior to 1930—which would require considerable coarsening of the data—we cannot apply a longer lag. As documented in panel D of Table A.II, temporal variation across CAI volumes is not ultimately central to the predictive relationship between patent flows and new title emergence: we obtain similar results when linking patents to the *initial* CAI volume for each four-decade subperiod (1940-1980, 1980-2018)—similar to the DOT. We also obtain similar results when removing new titles from each decade's CAI before obtaining augmentation-linked patents, as shown in panel E of Table A.II, and Table A.XII.

particular, $B^{Aut,10}_{j,t-20}$ has essentially no predictive power for subsequent augmentation patent flows $\text{AutX}_{j,t}$ while, similarly, $B^{Aug,10}_{j,t-20}$ is not predictive of subsequent augmentation patent flows $\text{AutX}_{j,t}$. The instruments are able to isolate independent variation in augmentation and automation patent flows.

Panels B and C of the table report separate first-stage estimates for the two halves of the sample, 1940–1980 and 1980–2018. While there is strong identification of both endogenous variables in the pooled full-sample specification, the instruments have a weaker first stage in the first half of the sample, 1940–1980, with F-stats of 8.20 and 11.46 on augmentation and automation, respectively. We treat the second-stage estimates for this period with due caution. In the latter half of the sample, both instruments play a role in identification of each endogenous variable, though the within-category coefficients remain larger and more precise than the cross-category coefficients.

Table IV reports 2SLS estimates of the effects of augmentation and automation patents on the emergence of new work between 1940 and 2018. Instrumenting augmentation patent flows using breakthrough patents from two decades earlier, column 1 obtains a coefficient of $\hat{\beta}^{Aug}_{40-18} = 24.42\,(7.46)$, implying that a 10% increase in augmentation patenting exposure increases new title emergence by about 2.4% over the course of a decade. When broad occupation by decade effects are added in column 2, this point estimate falls modestly to $\hat{\beta}^{Aug}_{40-18} = 21.30\,(5.81)$ and precision increases. Column 3 removes the augmentation exposure variable and adds the automation exposure measure in its place. Distinct from the OLS estimates, automation patents do not by themselves positively predict the flow of new titles—the point estimate is small and imprecise ($\hat{\beta}^{Aut}_{40-18} = -3.01$ (11.82)). When instrumented augmentation and automation patent flows are both included in columns 4 and 5, the estimated effect of augmentation patents on new title emergence is comparable in magnitude and precision to the estimates in earlier columns, while the coefficient on automation patents is imprecise, with an inconsistent sign in both columns. These 2SLS results are overall substantively similar to their non-instrumented counterparts in Table II. (Unreported placebo regressions that instead use IV dummies for exposure to the *bottom* 10% of augmenting or automating innovations—the least innovative patents—have weaker first-stage predictive power and no clear second-stage relationship.)

Following the format of Table II, panels B and C of Table IV report separate estimates for the two four-decade halves of the sample. Focusing for brevity on the final column of results (column 5), the estimated causal effect of augmentation innovations on new title emergence is positive and statistically significant in both 1940–1980 and 1980–2018. Consistent with earlier findings, the point estimates for automation innovations are generally imprecise and not consistently signed. Also echoing earlier results, the point estimates for the effect of augmentation innovations on new title flows are considerably larger in the earlier 1940–1980 period ($\hat{\beta}^{Aug}_{40-80} = 24.82\,(11.78)$) than in the later 1980–2014 period ($\hat{\beta}^{Aug}_{80-18} = 15.36\,(5.93)$). We caveat however that the estimates for the earlier period have a weaker (though still conventionally robust) first stage and hence should be viewed somewhat cautiously.

While the negative binomial analysis in Table II simultaneously accounts for intensive margin (number of new titles) and extensive margin (*any* new titles) responses to augmentation and automation, the 2SLS models compel us to analyze the two separately. Table A.IX explores the extensive margin impacts of augmentation and automation patent flows for the 1940–2018 period, with corresponding first stage estimates in Table A.VIII. In these linear probability models, the dependent variable is a dummy equal to one if an occupation has a positive count of new titles in the relevant decade (and zero otherwise). The first-stage estimates again reveal strong identification in this slightly larger sample (where occupations with zero new titles are retained). The second-stage estimates for the effect of augmentation patents on new title flows are positive but relatively imprecise, with the exception of column (2) ($\hat{\beta}^{Aug}_{40-18} = 1.99\,(0.97)$). Similarly, the estimated slopes on automation innovations are imprecise and vary in magnitude across columns. It bears note that the vast majority of the variation in new title emergence occurs on the intensive margin: 83% of occupation-year observations have at least one new title, and when weighting by employment shares (as done in the regressions), this number rises to 94%. It is therefore perhaps unsurprising that the dummy IV linear probability model does not identify this effect precisely. Nevertheless, the OLS and 2SLS estimates for the extensive margin are broadly consistent: OLS extensive margin estimates reported earlier in panel C of Table A.II find quantitatively comparable—but far more precise—patterns than their 2SLS counterparts. (As with the intensive margin results, 2SLS estimates of the impact of augmentation are slightly larger than corresponding OLS models.)

In summary, 2SLS estimates corroborate earlier non-instrumented results: new augmentation innovations catalyze the emergence of new work across occupations and over time whereas new automation innovations have no similar effect, suggesting that these measures capture economically distinct components of innovation. One skeptical interpretation, however, is that automation patents are simply a noisy measure of augmentation patents, which causes the automation measure to be significant on its own but non-robust to the inclusion of augmentation. Militating against this interpretation is that automation patents have significant independent and opposite-signed (relative to augmentation patents) predictive power for employment and wagebill growth in both OLS and 2SLS models, as shown below in Section V. We view the distinct relationships between augmentation, automation, and new title flows as (provisional) confirmation of our hypotheses.

The 2SLS estimates also echo another finding from the OLS models above: the new-work generating effects of augmentation innovations appear to have slackened over time—indeed, the point estimates are less than half the size in 1980–2018 than 1940–1980. This time pattern could potentially reflect a change in the sensitivity of the Census Bureau's procedure for capturing new titles in the CAI. But there are two reasons to suspect otherwise. First, the identification of $\hat{\beta}$ stems from cross-occupation comparisons in the flow of augmentation patents and new titles—and there is no expectation that a change in CAI data collection procedures would weaken these cross-occupation correlations. Second, our evidence below for employment and wagebills—where measurement is highly stable across decades—shows a similar evolution in the latter four decades of the sample.

IV.C Demand shocks and new work creation

While technological innovations are one force contributing to new work creation, our conceptual framework implies that occupational demand shifts also shape the flow of new work: positive demand shifts spur innovations that generate new occupational specialties, and negative demand shifts slow the rate of new work emergence. We test these predictions, first by exploiting the well-documented China trade shock (Autor, Dorn, and Hanson, 2016) to identify the impact of adverse demand shocks on the flow of new titles; and second, by exploiting demographic shifts, following DellaVigna and Pollet (2007), to identify the impact of *positive* demand shocks on the flow of new titles. In both cases, we identify occupations' differential exposure to positive and negative demand shifts by leveraging their employment distributions across more- versus less-exposed industries. Although both inward and outward demand shifts are expected to affect the arrival rate of new work, we find it useful to apply countervailing tests since new work never flows in reverse and occupational titles are rarely removed from the Census Index. In addition, the two demand shock instruments cover different time intervals.

Trade shocks and demand contraction

Starting in the early 1990s, import competition from China generated a sizable negative demand shock to many labor-intensive domestic U.S. manufacturing industries (Bernard, Jensen, and Schott, 2006; Autor, Dorn, and Hanson, 2013; Autor, Dorn, Hanson, and Song, 2014; Pierce and Schott, 2016; Acemoglu, Autor, et al., 2016; Autor, Dorn, and Hanson, 2021). These shocks directly affected product demand across industries and, since occupational employment shares differ across industries, indirectly affected labor demand across occupations. We use this variation to construct a measure of occupational exposure to the China trade shock, leveraging shifts in imports from China among a set of developed countries other than the United States.[23] This trade exposure measure captures a plausibly exogenous shock to domestic occupational demand stemming from rising Chinese productivity and falling China-facing trade barriers between 1991 and 2014. Occupational exposure to import competition varies substantially both between and within production and non-production occupations over 1990–2014, as highlighted in Figure A.II.

Demographics shocks and demand expansion

As a source of *positive* demand shifts, we follow DellaVigna and Pollet (2007) in exploiting changes in the demographic structure of the U.S. population to predict movements in industry-level de-

[23]Related work maps trade shocks to labor market outcomes in Brazil, Canada, India, Norway, Germany, Mexico, and elsewhere (Chiquiar, 2008; Topalova, 2010; Kovak, 2013; Dauth, Findeisen, and Suedekum, 2014; Balsvik, Jensen, and Salvanes, 2015; Branstetter et al., 2019; Devlin, Kovak, and Morrow, 2021). Our approach follows Autor, Dorn, and Hanson (2013) and subsequent papers, with the difference that we project the trade shock variation to the occupation level. Appendix F1 details the construction of this demand shift measure.

mands, which in turn affect occupation-level demands.[24] For this exercise, we obtain predicted consumption across product categories for household members of different ages, and combine these age-specific coefficients with population data to construct occupational exposure as a function of changes in predicted consumption by industry over 1980–2018, which are in turn based on changes in population age structure. This demand measure accounts for inter-industry input-output linkages, and hence corresponds to final demands. Appendix F2 provides details on the measure's construction.

Demand shifts and new work emergence: Estimates

We leverage these two sources of variation to estimate the effects of occupational demand shocks on the emergence of new occupational titles using the following count data model:

$$\ln E\left[\text{Newtitles}_{j,t}\right] = \beta_1^k \times \text{DemandX}_{j,t}^k + \beta_2 \text{AugX}_{j,t} + \beta_3 \sum_k \gamma_{jk,t} + X_{j,t}' \beta_4 + \delta_t. \tag{2}$$

where $\text{DemandX}_{j,t}^k = \sum_K \gamma_{jk,t} \hat{D}_{k,t}$ is the shift-share demand measure, equal to the sum of predicted log changes in industry employment $\hat{D}_{k,t}$ in industries k multiplied by the start-of-decade share of employment in occupation j contained in industry k, denoted by $\gamma_{jk,t}$. The sum of occupational exposure weights, $\sum_k \gamma_{jk,t}$, is included to account for the incomplete shares term in shift-share instrumental variables models (Borusyak, Hull, and Jaravel, 2022).[25] Additional controls include occupational augmentation exposure, $\text{AugX}_{j,t}$, and a set of decade dummies, δ_t. To purge any potential mechanical association between occupational size and the rate of new title emergence, $X_{j,t}$ controls for occupations' initial (start of period) employment share and, in one specification, for their contemporaneous employment share change. The coefficient β_1^k corresponds to the estimated effect of demand shift $\text{DemandX}_{j,t}^k$ for $k \in \{C, D\}$ on new title flows, where C denotes China trade exposure and D denotes exposure to demographic demand. As with earlier models for new title counts, we estimate this equation with a negative binomial regression.

Our conceptual model predicts that $\beta_1^C < 0$ and $\beta_1^D > 0$: inward occupational demand shifts slow new work creation and outward occupational demand shifts accelerate it. Estimates of equation (2) in Table V support these predictions. Panel A reports the effect of inward occupational demand shocks (stemming from trade shocks) on the emergence of new occupational titles between 1990–2000 and 2000–2018. The first specification controls for decade main effects, occupational employment shares, and the sum of exposure shares. The column 1 estimate finds an imprecise negative effect of demand contractions on new title emergence ($\hat{\beta}_1^C = -9.83\,(6.09)$). Column 2 adds dummies for broad occupational groups, so identification stems from contrasts across detailed oc-

[24] Our work is related to Mazzolari and Ragusa (2013), Leonardi (2015), and Comin, Danieli, and Mestieri (2020), who analyze the causal effects of age, education, and income on employment via consumption.

[25] Shares do not sum to one because: non-manufacturing industries are not directly exposed to trade shocks; and public sector services are not enumerated in Consumer Expenditure Survey data and hence do not contribute to the demographic demand shock measure.

cupations within the twelve broad occupational categories depicted in Figure A.II—from variation within rather than between rows of the figure. The estimated effect of import exposure increases in absolute magnitude and precision to $\hat{\beta}_1^C = -15.43\,(5.23)$.

Column 3 adds the occupation-level augmentation exposure measure from above to the column 1 model, thus allowing both demand shocks and augmentation innovations to affect new title emergence. The augmentation variable obtains a positive coefficient of $\hat{\beta}_2 = 7.94\,(4.60)$. Column 4 again includes broad occupation fixed effects: this further increases the magnitude and precision of the import exposure variable ($\hat{\beta}_1^C = -17.49\,(5.13)$) as well as the augmentation exposure term ($\hat{\beta}_2 = 9.38\,(3.00)$). Column 5 additionally controls for the contemporaneous change in occupational employment shares—which is quite conservative since employment shares are an intermediate outcome that is directly affected by the China trade shock. Nevertheless, the coefficients of interest are hardly affected.

Might these estimates reflect longstanding trends in new work creation in trade-exposed occupations that predate the China trade shock? Panel B of Table V confronts this concern with a placebo test. The dependent variable in this panel is the flow of new occupational titles from the twenty years *prior* to the China trade shock, 1970–1980 and 1980–1990, while the independent variable is the post-1990 change in import exposure, as per panel A. These placebo estimates reveal that recent Chinese import competition has little to no predictive power for the emergence of new titles from 1970 through 1990 in subsequently trade-exposed occupations. Estimates are imprecisely *positive* in the first three columns and are negative and close to zero in the final two columns, which contain the full complement of controls. These results increase our confidence that the estimates in panel A reflect the causal impact of demand contractions on the flow of new work.

Panel C of Table V presents a parallel exercise exploiting demographic demand shifts in place of trade shocks, here extending the sample window backward to 1980. Consistent with expectations, *positive* demand shifts speed the emergence of new occupational titles. In column 1, the point estimate of $\hat{\beta}_1^D = 18.38\,(5.49)$ on the demographic demand index implies that a 10% increase in occupational demand shock exposure increases the rate of occupational new title emergence by approximately 1.8%. Subsequent columns of panel C explore robustness by applying the same specifications used for the trade-based demand shock. Across all columns, occupations with higher predicted demand growth stemming from demographic change exhibit faster new title emergence. As in panels A and B, the coefficient on augmentation exposure is positive and precisely estimated. Moreover, its inclusion has a relatively small effect on the coefficient on the demand shift variable. Figure VIII shows why this is the case: the set of occupations most exposed to demographic demand shifts is substantively different from those most exposed to augmentation innovations. Using the partial predicted effects from a version of Table V, we find that augmentation exposure and demand exposure are essentially uncorrelated over 1980–2018. Figure VIII indicates that demographic demand shifts can help account for the emergence of new titles in personal service and care jobs such as housekeepers, waiters, food preparation workers, and licensed practical nurses (see also Mazzolari and Ragusa 2013; Leonardi 2015; Comin, Danieli, and Mestieri 2020). Conversely, new title emergence in high-tech jobs, such as electrical engineers and computer systems analysts, is primarily predicted by augmentation exposure.

The evidence in Table V highlights that new work has multiple origins: the flow of augmentation innovations and the operation of demand shifts both shape where new specialized, labor-using tasks emerge. These results also offer a further substantive implication. Much evidence documents that rising import competition has depressed employment in trade-exposed industries and associated occupations during the last three decades (see evidence summarized in Autor, Dorn, and Hanson (2021)). The Table V results reveal that this competitive pressure not only numerically reduces employment but also depresses the emergence of new categories of specialized work: it yields not only fewer jobs but also less *new* work. This distinction is relevant because, as we show in Table A.XV, new work appears to be better-remunerated than existing work within the same occupation, possibly because it is more specialized.

V Augmentation, Automation, and Occupational Labor Demand

What does our evidence on the determinants of new work imply for occupational employment and labor demand more generally? The findings so far do not unambiguously answer this question since employment could potentially contract in occupations where new titles emerge or expand in occupations where tasks are automated. Our conceptual model predicts however, that because new task creation generates new demand for human expertise, it will expand employment and wagebills in augmentation-exposed occupations; and, conversely, because task automation is labor-displacing, it will erode employment and wagebills in automation-exposed occupations (see Proposition 2 in Appendix J). We explore those implications here.

Figure IX motivates this analysis by documenting the striking association between new title emergence and occupational employment growth across both halves of our sample, netting out the correlation between start-of-period title count and occupational employment growth (which is generally negative since larger occupations grow more slowly). The locus of new title introduction and employment growth has shifted substantially across the two four-decade intervals of our sample, yet new title flows strongly predict the set of occupations that gain and lose ground in both periods. Between 1940 and 1980, occupations rapidly adding both employment and new titles include clerical occupations such as stenographers, typists, secretaries, office machine operators, and bookkeepers; and blue-collar production occupations such as foremen, mechanics and repairmen, and operative workers. Simultaneously, occupations including elevator operators, lumbermen, railroad switchmen, dressmakers, and mining occupations grew slowly and added few titles. In the subsequent four decades, 1980–2018, professional and personal services dominate the occupations rapidly adding employment and new titles: computer software developers; dental laboratory and medical appliance technicians; vocational and educational counselors; registered nurses; and hairdressers and cosmetologists. Conversely, blue-collar production occupations, such as machinists, machine feeders and offbearers, and white-collar clerical occupations, including bank tellers, typists, and secretaries and stenographers, added relatively few occupational titles and exhibited substantial relative employment declines.

These patterns echo the evidence from Figure II, showing that the locus of new work creation shifted from the middle-paid center of the occupational distribution before 1980, to the high-paid and low-paid tails of this distribution after 1980. Figure A.III facilitates a more direct comparison of new title growth rates across periods by applying a consistent set of occupation codes over 1940–2018, albeit at the cost of substantially lower occupational resolution. What stands out from this figure is the shifting fortunes of routine task-intensive occupations—both blue-collar occupations such as operative and kindred workers, metal workers, and mechanics; as well as white-collar occupations such as shipping and receiving clerks; stenographers, typists, and secretaries; and library attendants and assistants. These occupations were well above the median of new title growth between 1940 and 1980, and then fell substantially in rank (i.e., below the 45-degree line) during the following four decades, coinciding with the decline of routine task-intensive employment (Autor and Dorn, 2013).

These correlations between employment growth and new title flows should not be interpreted as causal. Indeed, the prior results in Table V underscore that the co-occurrence of new occupational tasks and rising occupational employment could reflect the operation of demand shifts. To understand the effects of augmentation and automation on labor demand, we shift focus from new titles to employment and wagebills. Using both OLS and 2SLS models, we analyze the correlational and causal relationships between augmentation and automation innovations and changes in occupational employment and wagebills.

V.A Employment and wagebills: Main results

To assess the relationship between augmentation exposure, automation exposure, employment and wagebills, we estimate models of the form:

$$100 \times \Delta \ln(Y_{ij,t}) = \beta_1 \text{AugX}_{ij,t} + \beta_2 \text{AutX}_{j,t} + \gamma_{i,t} + \delta_{J,t} + \varepsilon_{ij,t}, \tag{3}$$

where the dependent variable is the log change in full-time equivalent employment (annual hours divided by 35 hours $\times$ 50 weeks) or wagebill in consistent three-digit industry i by occupation j cells, multiplied by 100 so changes roughly correspond to percentage points. While the new title models above were limited to occupation-by-decade variation in the outcome variable, this estimation is able to exploit richer occupation-by-industry-by-decade variation in employment and wagebills. The key independent variables are $\text{AugX}_{ij,t}$, quantifying exposure to augmentation in industry-by-occupation cells, and $\text{AutX}_{j,t}$, quantifying exposure to automation in occupation cells. Paralleling the structure of the dependent variable, the augmentation measure is extended to vary at the level of occupation-industry by appending industry-specific titles from the Census Alphabetic Index onto the occupational titles used above (see Section II.C). Because the analysis in this section focuses on employment changes over two four-decade intervals (1940–1980 and 1980–2018), we take the (log of) the sum of citation-weighted patent matches for $\text{AugX}_{ij,t}$ and $\text{AutX}_{j,t}$ within each four-decade interval. The inclusion of a full set of industry-by-decade dummy variables, $\gamma_{i,t}$ (116 industries for 1940–1980, 206 industries for 1980–2018), means that the coefficients of interest are identified by changes in within-industry occupational employment, holding constant overall industry employment shifts. Additional specifications include broad occupation fixed effects further interacted with time period dummies, $\delta_{J,t}$. Standard errors are clustered on industry-by-occupation cells.

Table VI presents OLS estimates in columns 1 and 2, and 2SLS estimates in columns 3 and 4. The first panel of Table VI fits a stacked long-difference version of equation (3) over eight decades where each observation is the log change in employment in consistent occupation-industry cells in a four-decade interval (1940–1980 or 1980–2018). Column 1 finds that occupations that are more exposed to augmentation exhibit faster employment growth and conversely, those more exposed to automation exhibit slower employment growth. Indeed, augmentation and automation have precisely-estimated but opposite-signed predictive relationships with occupational employment. In column 1, 10 percent more augmentation exposure predicts 0.15 percent more employment growth ($\hat{\beta}_1 = 1.51\,(0.21)$), and 10 percent more automation exposure predicts 0.23 percent less employment growth ($\hat{\beta}_2 = -2.27\,(0.27)$). Column 2 adds controls for broad occupation by time-effects, demonstrating that these predictive relationships stem both from contrasts within and between broad occupational categories.

To address endogeneity concerns discussed earlier, the second pair of columns in Table VI present 2SLS estimates that use breakthrough patent flows from the prior two decades as instruments for observed augmentation and automation patent flows.[26] First stage estimates for these models are reported in Table A.III. The 2SLS results confirm the implications of the OLS estimates but with greater magnitude: a 10% increase in augmentation exposure spurs 0.36% additional employment growth ($\hat{\beta} = 3.60$ (0.96)) over the course of a decade in the 2SLS model relative to 0.14% in the corresponding OLS column; a 10% increase in automation exposure spurs a 0.42% reduction in employment growth ($\hat{\beta} = -4.21$ (0.93)) over the course of a decade in the 2SLS model relative to 0.10% in the OLS estimate. The larger point estimates for both the automation and augmentation employment elasticities in these 2SLS estimates suggest that attenuation due to measurement error may impart a larger bias to the OLS estimates than does simultaneity. We caveat, however, that the 2SLS estimates have broader confidence intervals than their OLS counterpart, such that OLS point estimates generally fall within the 95% confidence bands of their 2SLS counterparts.

[26] As with earlier 2SLS models, we normalize the breakthrough exposure instruments with sector-specific (manufacturing vs. non-manufacturing) median breakthrough thresholds. Let $N^{A,10}_{ji,t-20}$ equal breakthrough patent exposure count for occupation j in industry i in decade $t-20$, where $A \in \{\text{Aug}, \text{Aut}\}$. Like the endogenous variable, the breakthrough matches $N^{A,10}_{ji,t-20}$ are also aggregated over a 4-decade period, except the aggregation window is lagged by 20 years as with our previous 2SLS analysis. Occupation by industry cells are classified as breakthrough-exposed according to $B^{A,10}_{ji,t-20} = \mathbb{1}\left[N^{A,10}_{ji,t-20} > \tilde{N}^{A,10}_{j,i\in\text{manuf},t-20}\right] \times [i \in \text{manuf}] + \mathbb{1}\left[N^{A,10}_{j,i,t-20} > \tilde{N}^{A,10}_{j,i\notin\text{manuf},t-20}\right] \times [i \notin \text{manuf}]$. Here, the tilde denotes the employment-weighted cross-sectional median of the variable defined over industry-occupation cells, sub-setting according to whether or not industry i is or is not in the manufacturing sector. We analogously construct $B^{A,100}_{ji,t-20}$, denoting above-median overall (breakthrough + non-breakthrough) patent exposure.

The next set of results in panel B of Table VI focus on wagebill, equal to the product of employment and earnings. These wagebill estimates prove highly comparable to the corresponding employment results in panel A: the OLS coefficient for the effect of augmentation innovations on wagebill in column 2 of panel B is 1.41 (0.25) versus 1.36 (0.22) on employment in column 2 of panel A. For the 2SLS estimates in column 4, the point estimates are 4.04 (0.98) for wagebill (panel B) versus 3.63 (0.96) for employment (panel A). The corresponding OLS estimates for automation innovations are $-0.97\,(0.42)$ for wagebill and $-1.00\,(0.40)$ for employment, while their 2SLS counterparts are $-3.43\,(0.97)$ for wagebill and $-4.21\,(0.93)$ for employment. These patterns are consistent with the interpretation that employment and wagebills move one-for-one, suggesting a unitary elasticity of substitution across occupations (also assumed in our model, where occupations correspond to sectors). This literal interpretation should be qualified, however. Papers by Böhm, Gaudecker, and Schran (2022) and Autor and Dorn (2009) demonstrate that contracting occupations (here, those more exposed to automation) tend to retain both more experienced workers and workers with relatively high earnings given their experience, while the opposite occurs in expanding occupations (here, those more exposed to augmentation). These compositional shifts, which are akin to quantity rather than price changes in an earnings equation, cloud inference on the effect of augmentation and automation on wagebills *net* of these compositional shifts.

To (imperfectly) address this compositional issue, we estimate the effect of augmentation and automation innovations on composition-adjusted wages, as detailed in Appendix G. Briefly, we use cross-sectional Mincerian wage regressions in each Census year to predict the log hourly wage of each worker based exclusively on her schooling, race, and gender, each interacted with a quadratic in age. We take the cell-level average of these predictions to form the expected wage in each occupation-industry cell ($\widehat{W}_{ij,t}$), purged of occupation-industry premia. This procedure delivers two new wagebill change measures (alongside the observed wagebill change, $\Delta W_{ij,t}$): the expected wagebill change ($\Delta\widehat{W}_{ij,t}$); and the composition-adjusted wagebill change ($\Delta\widetilde{W}_{ij,t}$), equal to the observed log change in employment plus the log difference between the observed and composition-adjusted wage change.

Panel C of Table VI uses as the dependent variable the expected wage, $\Delta\widehat{W}_{ij,t}$. Consistent with Böhm, Gaudecker, and Schran (2022), *expected* wagebill changes are less positive for augmentation exposure and more negative for automation exposure than are *observed* wagebill changes—particularly in 2SLS models. This pattern suggests that adverse compositional shifts in augmentation-exposed occupations partly mask any positive wage relationship, whereas positive compositional shifts in automation-exposed occupations mask negative wage impacts. Panel D combines these results to estimate models for composition-adjusted wagebills. Accounting for compositional changes within occupation-industry cells, estimates find that augmentation innovations have a meaningfully larger positive effect on wagebills than on employment. Focusing on 2SLS estimates, column 4 of panel A finds that a 10% rise in augmentation exposure spurs a 0.36% rise in employment ($\hat{\beta}_1 = 3.63\,(0.96)$), and in column 4 of panel D, a 0.43% rise in the adjusted wagebill ($\hat{\beta}_1 = 4.36\,(1.00)$). By implication, the wage premium in the exposed industry-occupation cell rises by 0.07% ($0.43 - 0.36 = 0.07$). This difference between employment and wagebill impacts, which is pronounced in 2SLS models but nevertheless apparent in OLS models, hints that 'new work' may be more-skilled or better-remunerated than (simply) 'more work'. We return to this point below.[27]

As a further robustness test, we explore sectoral impacts. Much of the economic literature on the labor demand impact of technological innovations focuses on manufacturing, where capital investments are well-measured and technologies are often intensively used, e.g., in the case of industrial robotics (e.g., Berman, Bound, and Griliches 1994; Doms, Dunne, and Troske 1997; Lewis 2011; Humlum 2021; Acemoglu and Restrepo 2020; Curtis et al. 2021; Dechezleprêtre et al. 2021; Aghion et al. 2022; Hirvonen, Stenhammar, and Tuhkuri 2022). The augmentation and automation exposure measures developed here are not however confined to manufacturing since they are based on the semantic content of occupations and innovations. Leveraging this fact, Table A.V plots a corresponding set of OLS and 2SLS employment and adjusted-wagebill models for the 1940–2018 period, estimated separately for manufacturing and non-manufacturing. Within both the manufacturing and non-manufacturing sectors, along employment and wagebill margins, and using both OLS and 2SLS models, we find that augmentation innovations significantly raise occupational labor demand whereas automation innovations significantly depress it.

V.B Evidence of accelerating automation

We observed above that the elasticity of new work creation, as proxied by new job titles, with respect to augmentation innovation appears to have declined between 1940–1980 and 1980–2018 (Tables II and IV). We perform a parallel exercise for occupational labor demand in Table VII, which reports OLS and 2SLS models for the effects of augmentation and automation innovations on employment and composition-adjusted wagebills separately for 1940-1980 and 1980-2018. We note that there is no intrinsic statistical relationship between these two margins of response—new title creation versus occupational labor demand—so we view this test of parallelism as informative. The by-period estimates in Table VII yield three results. First, OLS and 2SLS estimates strongly corroborate the positive impact of augmentation innovations on employment and wagebills in both four-decade intervals under study. Given that the measures from these two time intervals are built from different patent cohorts, different occupational labor task descriptors (1939 DOT vs. 1977 DOT), and different occupational output descriptors (1950–1980 CAI vs. 1990–2018 CAI), the stability and robustness of the estimates across epochs buttresses confidence in the findings.

Second, akin to the evidence for new title flows, both OLS and 2SLS models suggest a fall in the elasticity of employment creation with respect to augmentation innovations (and a more negative elasticity with respect to automation) in the second relative to first half of the sample. Referring to the even-numbered columns of Table VII (which include broad occupation by year dummies), the OLS elasticity estimate for employment falls from $\hat{\beta}_1^{40-80} = 2.12$ (0.38) to $\hat{\beta}_1^{80-18} = 0.78$ (0.23).

[27] We do not, however, find that automation has larger negative effects on occupational wagebills than occupational employment in either OLS or 2SLS models.

The corresponding 2SLS elasticity estimate for employment falls from $\hat{\beta}_1^{40-80} = 5.24$ (2.42) to $\hat{\beta}_1^{80-18} = 2.79$ (0.93). OLS and 2SLS estimates for the augmentation elasticity of the wagebill also finds qualitatively similar declines in magnitudes. Moreover, coefficient estimates in odd-numbered columns which instead rely on between-broad occupation variation are generally larger and much more precise for our automation measure (and for augmentation in the case of IV estimates), but they also deliver the same pattern of decreased augmentation and increased automation intensity when comparing coefficients across time periods.

Finally, we find a rise in the elasticity of employment and wagebill erosion to automation during 1980–2018 relative to 1940–1980. Focusing on employment, the OLS estimate of this elasticity rises in absolute magnitude from $\hat{\beta}_2^{40-80} = -0.27$ (0.64) to $\hat{\beta}_2^{80-18} = -1.98$ (0.46) The corresponding 2SLS estimate grows from $\hat{\beta}_2^{40-80} = -2.79$ (2.95) to $\hat{\beta}_2^{80-18} = -5.07$ (0.92) This pattern is inescapable in Figure X, which plots the conditional relationship between augmentation, automation, and employment growth separately for these two four-decade intervals, using the somewhat more precise OLS estimates from columns 1 and 5 which utilize between- instead of within-broad occupation variation.

In summary, Tables VI and VII, as well as the supporting material in Tables A.V and A.XV, corroborate a central implication of the conceptual framework: despite their positive rate of co-occurrence across occupations, augmentation and automation have countervailing causal effects on sectoral labor demand. This is true for both employment and wagebills, for both the 1940–1980 and 1980–2018 periods, for both manufacturing and non-manufacturing, and for both OLS and 2SLS estimates, the latter of which leverage breakthroughs occurring two decades earlier as instruments for subsequent downstream innovations.

Our findings are consistent with Kogan et al. (2021) and Webb (2020), who find that occupations that are more exposed to automation (Kogan et al., 2021) or software (Webb, 2020) patents have seen falling relative employment in recent decades.[28] Our analysis add two central findings to this body of work: isolating a countervailing augmentation force, also stemming from innovation, that is strongly predictive of occupational employment growth; and providing causal evidence that such innovations boost labor demand while automation innovations erode it. In the terminology of Acemoglu and Restrepo (2019), we find that automation is labor-displacing and augmentation is labor-reinstating. Also consistent with their evidence, the constellation of results for both new task creation and occupational labor demand suggests that the labor-reinstating power of augmentation has attenuated in recent decades while the labor-replacing force of automation has accelerated.

[28] Mann and Püttmann (2021) present a contrasting finding: service sector employment grows relatively faster in commuting zones exposed to more automation patents. Given the differing unit of analysis—geographies versus occupations—these findings are not directly comparable.

V.C Is new work different from more work?

Our primary public use Census and ACS data files, augmented with new title counts, document the prevalence of new work within macro-occupations but do not allow us to observe which workers are directly employed in that work. These public-use data sources are accordingly ill-suited to estimating wage differentials in new work. Appendix H seeks to fill this gap by harnessing data from the 1940 Census Complete Count file (Ruggles, Fitch, et al., 2021) to test whether new work is more skilled or better paid than conventional work. The unique virtue of these no-longer-confidential Census Complete Count data is that they report *all* fields for each Census respondent, enabling direct observation of which workers are employed in new work, i.e., in occupational titles that first appeared in the 1940 CAI.

Panel A of Table A.XV explores whether there is observable positive selection of workers into new work. Using a classification procedure detailed in Appendix H1, we estimate that 2.6% of workers in 1940 were employed in new work titles that were first reported in that decade. Fitting a linear probability model to these data reveals that better-educated workers are significantly more likely to be employed in new work. This is true unconditionally, and also when accounting for a full set of macro-occupation main effects—thus contrasting workers within the same Census occupations—and when further accounting for workers' age, sex, race, state, and urban versus rural status. Net of these factors, workers with any post-secondary education (some college or greater) are around 25% more likely to be employed in new work than those with less than a high school degree (the modal educational category in 1940).

Panel B of Table A.XV explores whether there is a wage differential in new work net of selection on observables. Fitting an OLS cross-sectional wage regression to the 1940 data shows that, even after conditioning on a rich set of observables, including macro-occupation main effects, education, race, gender, age, state, and urban/rural status, workers employed in new work earn on average 6.0% more than observably similar workers in preexisting work. While this evidence does not establish that employment in new work raises wages, it indicates that either workers earn a premium for new work or that new work attracts workers with unobservably higher skill (even within the same macro-occupation and level of education). Either interpretation is consistent with the hypothesis that new work demands specialized skills or expertise relative to conventional work within the same occupations.

VI Conclusion

Much recent empirical work has documented the displacement of labor from existing job tasks by automation but is mostly silent on the countervailing force of labor reinstatement occurring through the creation of new job tasks or categories requiring specialized human expertise. Using newly constructed measures of the emergence of new work within occupations, as well as flows of innovations that may potentially augment or automate these occupations, we find that the locus of new work creation has shifted from middle-paid production and clerical occupations in the first four post-WWII decades, to high-paid professional and, secondarily, low-paid services since 1980. We show that new work emerges in response to both technological innovations that complement the outputs of occupations, and demand shocks that raise occupational demand. Conversely, innovations that automate existing job tasks do not yield new work, while adverse occupational demand shifts slow the rate of new work emergence. These flows of augmentation and automation innovations are positively correlated across occupations but have countervailing effects on labor demand: augmentation innovations boost occupational labor demand while automation innovations erode it. These effects are present in both four-decade epochs of our sample and are evident in both the manufacturing and non-manufacturing sectors. By harnessing shocks to the flow of augmentation and automation innovations spurred by breakthrough innovations occurring two decades prior, we establish that the effects of augmentation and automation innovations on new work emergence and occupational labor demand are causal.

The data, methods, and findings in this paper provide the foundation for further study of the role of innovation and incentives in automating existing work and catalyzing demands for novel human expertise and specialization. Some core questions that future research may seek to address include:

1. Is automation *accelerating* relative to augmentation, as many researchers and policymakers fear, or is it primarily the case that the locus of automation and augmentation has shifted across occupations and skill groups? On the latter possibility, our evidence is clear: automation has intensified in middle-skill occupations while augmentation has concentrated in professional, technical, and managerial occupations and, to a lesser extent in personal service occupations. On the former possibility, the results above suggest that in recent decades, the demand-eroding effects of automation innovations have intensified whereas the demand-increasing effects of augmentation innovations have slowed. While the augmentation and automation exposure measures developed here are too novel to warrant definitive conclusions about acceleration or deceleration over the course of many decades, the pattern of results across the margins of both employment and new work emergence suggest, as in Acemoglu and Restrepo (2019), that the last four decades have seen relatively more automation and less augmentation than the prior four.

2. Is 'new work' more labor-augmenting than 'more work'? The finding that augmentation innovations increase occupational wagebills by boosting *both* employment and wages suggests that 'new work' may be more valuable than 'more work'—plausibly because new work demands novel expertise and specialization that (initially) commands a scarcity premium. Identifying and quantifying such premia will require direct earnings observations of job tasks and earnings of workers engaged in new work, something that is largely infeasible in our Census public use microdata. If, as we suspect, new work provides additional opportunities for skill formation and earnings growth beyond 'more work', then policies that foster new work creation may be of particular interest.

3. While our evidence establishes that augmentation and automation move labor demand in countervailing directions, it does not answer why the locus of innovation has shifted over time, nor—more broadly—how elastic the locus and pace of augmentation and automation are to incentives. Theory has much to say on this topic (Acemoglu, 2010; Acemoglu and Restrepo, 2018b; Ray and Mookherjee, 2021), but direct measurement is a challenge. The augmentation and automation innovation measures developed here, and tools for identifying them, may prove valuable to this pursuit.

4. Will rapid advances in Artificial Intelligence shift the balance of innovation towards more rapid automation across an expanding set of occupational domains? And to the degree that AI is *not* exclusively automating, what novel specialties will it catalyze and what skill sets will it complement? Early evidence on this question points to a broad range of potential effects, but definitive findings are elusive so far (Brynjolfsson and Mitchell, 2017; Felten, Raj, and Seamans, 2019; Babina et al., 2020; Webb, 2020; Acemoglu, Hazell, Restrepo, et al., 2021; Autor, 2022; Brynjolfsson, Li, and Raymond, 2023; Eloundou et al., 2023; Noy and Zhang, 2023; Peng et al., 2023). AI innovations are however well-represented in patents (Webb, 2020), suggesting that the classification methodology developed here may offer insight into AI's potential for both augmentation and automation.

MIT AND NBER
MIT
UTRECHT UNIVERSITY
NORTHWESTERN UNIVERSITY KELLOGG SCHOOL

References

Acemoglu, Daron (1998). “Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality”. In: *The Quarterly Journal of Economics* 113.4, pp. 1055–1089.

— (2010). “When Does Labor Scarcity Encourage Innovation?” In: *Journal of Political Economy* 118.6, pp. 1037–1078.

Acemoglu, Daron and David H. Autor (2011). “Skills, Tasks and Technologies: Implications for Employment and Earnings”. In: *Handbook of Labor Economics* 4, pp. 1043–1171.

Acemoglu, Daron, David H. Autor, et al. (2016). “Import Competition and the Great US Employment Sag of the 2000s”. In: *Journal of Labor Economics* 34.S1, S141–S198.

Acemoglu, Daron, Joe Hazell, Pascual Restrepo, et al. (2021). “AI and Jobs: Evidence from Online Vacancies”. In: *Journal of Labor Economics*.

Acemoglu, Daron and Pascual Restrepo (2018a). “Artificial Intelligence, Automation and Work”. In: *The Economics of Artificial Intelligence*. Ed. by Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

— (2018b). “The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares and Employment”. In: *American Economic Review* 108.6, pp. 1488–1542.

— (2019). “Automation and New Tasks: How Technology Displaces and Reinstates Labor”. In: *Journal of Economic Perspectives* 33.2, pp. 3–30.

— (2020). “Robots and Jobs: Evidence from US Labor Markets”. In: *Journal of Political Economy* 128.6, pp. 2188–2244.

— (2022). “Tasks, automation, and the rise in us wage inequality”. In: *Econometrica* 90.5, pp. 1973–2016.

Aghion, Philippe et al. (June 2022). *Modern Manufacturing Capital, Labor Demand, and Product Market Dynamics: Evidence from France*. Tech. rep. LSE Working Paper, p. 100.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad (2015). “The Skill Complementarity of Broadband Internet”. In: *The Quarterly Journal of Economics* 130.4, pp. 1781–1824.

Atack, Jeremy, Robert A Margo, and Paul W Rhode (2019). “”Automation” of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study”. In: *Journal of Economic Perspectives* 33.2, pp. 51–70.

Atalay, Enghin, Phai Phongthiengtham, et al. (2020). “The Evolution of Work in the United States”. In: *American Economic Journal: Applied Economics* 12.2, pp. 1–34.

Atalay, Enghin and Sarada (2020). *Firm Technology Upgrading Through Emerging Work*. Tech. rep.

Autor, David H. (2019). “Work of the Past, Work of the Future”. In: *AEA Papers and Proceedings*. Vol. 109, pp. 1–32.

— (May 2022). “The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty”. In: *Brookings Institution: Global Forum on Democracy and Technology*.

Autor, David H., Caroline Chin, et al. (Aug. 2022). “New Frontiers: The Origins and Content of New Work, 1940–2018”. In: *NBER Working Paper No. 30389*.

Autor, David H. and David Dorn (2009). “This Job is ”Getting Old”: Measuring Changes in Job Opportunities using Occupational Age Structure”. In: *American Economic Review* 99.2, pp. 45–51.

— (2013). “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market”. In: *American Economic Review* 103.5, pp. 1553–97.

Autor, David H., David Dorn, and Gordon Hanson (2021). “On the Persistence of the China Shock”. In: *Brookings Papers on Economic Activity*, pp. 381–447.

Autor, David H., David Dorn, Gordon H Hanson, and Jae Song (2014). “Trade Adjustment: Worker-level Evidence”. In: *The Quarterly Journal of Economics* 129.4, pp. 1799–1860.

Autor, David H., David Dorn, and Gordon H. Hanson (2013). “The China Syndrome: Local Labor Market Effects of Import Competition in the United States”. In: *American Economic Review* 103.6, pp. 2121–68. DOI: 10.1257/aer.103.6.2121. URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.103.6.2121.

— (2016). “The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade”. In: *Annual Review of Economics* 8.1, pp. 205–240.

Autor, David H., Claudia Goldin, and Lawrence F. Katz (2020). “Extending the Race Between Education and Technology”. In: *AEA Papers and Proceedings* 110, pp. 347–51.

Autor, David H., Lawrence F. Katz, and Melissa S Kearney (2006). “The Polarization of the US Labor Market”. In: *American economic review* 96.2, pp. 189–194.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger (Nov. 1998). “Computing Inequality: Have Computers Changed The Labor Market?” In: *The Quarterly Journal of Economics* 113.4, pp. 1169–1213.

Autor, David H., Frank Levy, and Richard J Murnane (2003). “The Skill Content of Recent Technological Change: An Empirical Exploration”. In: *The Quarterly journal of economics* 118.4, pp. 1279–1333.

Babina, Tania et al. (2020). *Artificial Intelligence, Firm Growth, and Industry Concentration*. SSRN Working Paper 3651052.

Balsvik, Ragnhild, Sissel Jensen, and Kjell G Salvanes (2015). “Made in China, Sold in Norway: Local Labor Market Effects of an Import Shock”. In: *Journal of Public Economics* 127, pp. 137–144.

Bárány, Zsófia L and Christian Siegel (2018). “Job Polarization and Structural Change”. In: *American Economic Journal: Macroeconomics* 10.1, pp. 57–89.

Berman, Eli, John Bound, and Zvi Griliches (1994). “Changes in the Demand for Skilled Labor within US Manufacturing: Evidence from the Annual Survey of Manufactures”. In: *The quarterly journal of economics* 109.2, pp. 367–397.

Bernard, Andrew B, J Bradford Jensen, and Peter K Schott (2006). “Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of US Manufacturing Plants”. In: *Journal of international Economics* 68.1, pp. 219–237.

Böhm, Michael J., Hans-Martin von Gaudecker, and Felix Schran (2022). *Occupation Growth, Skill Prices, and Wage Inequality*. eng. CESifo Working Paper 7877. Munich.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel (2022). “Quasi-Experimental Shift-Share Research Designs”. In: *The Review of Economic Studies* 89.1, pp. 181–213.

Branstetter, Lee G et al. (2019). *The China Shock and Employment in Portuguese Firms*. National Bureau of Economic Research.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond (2023). *Generative AI at Work*. Tech. rep. National Bureau of Economic Research.

Brynjolfsson, Erik and Tom Mitchell (Dec. 2017). “What Can Machine Learning Do? Workforce Implications”. In: *Science* 358.6370, pp. 1530–1534. ISSN: 0036-8075. DOI: 10.1126/science.aap8062. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aap8062.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock (2018). “What Can Machines Learn, and What Does It Mean for Occupations and the Economy?” In: *AEA Papers and Proceedings*. Vol. 108, pp. 43–47.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro (2019). “Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock”. In: *Econometrica* 87.3, pp. 741–835.

Card, David and Thomas Lemieux (2001). “Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis”. In: *The Quarterly Journal of Economics* 116.2, pp. 705–746.

Chen, Jiafeng and Jonathan Roth (2022). “Log-like? ATEs Defined with Zero Outcomes are (Arbitrarily) Scale-Dependent”. In: *arXiv preprint arXiv:2212.06080*.

Chiquiar, Daniel (2008). “Globalization, Regional Wage Differentials and the Stolper–Samuelson Theorem: Evidence from Mexico”. In: *Journal of International Economics* 74.1, pp. 70–93.

Comin, Diego A, Ana Danieli, and Martı Mestieri (2020). *Income-driven Labor Market Polarization*. Tech. rep. National Bureau of Economic Research.

Cortes, Guido Matias et al. (2020). “The Dynamics of Disappearing Routine Jobs: A Flows Approach”. In: *Labour Economics* 65, p. 101823.

Curtis, E Mark et al. (2021). *Capital Investment and Labor Demand*. Tech. rep. National Bureau of Economic Research.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum (2014). “The Rise of the East and the Far East: German Labor Markets and Trade Integration”. In: *Journal of the European Economic Association* 12.6, pp. 1643–1675.

Dechezleprêtre, Antoine et al. (2021). “Induced Automation: Evidence from Firm-Level Patent Data”. In: *University of Zurich, Department of Economics, Working Paper* 384.

DellaVigna, Stefano and Joshua M. Pollet (Dec. 2007). “Demographics and Industry Returns”. In: *American Economic Review* 97.5, pp. 1667–1702. DOI: 10.1257/aer.97.5.1667. URL: https://www.aeaweb.org/articles?id=10.1257/aer.97.5.1667.

Deming, David J and Kadeem Noray (2020). "Earnings Dynamics, Changing Job Skills, and STEM Careers". In: *The Quarterly Journal of Economics* 135.4, pp. 1965–2005.

Devlin, Allison, Brian K Kovak, and Peter M Morrow (2021). "The Long-Run Labor Market Effects of the Canada-US Free Trade Agreement". In: *Carnegie Mellon University Manuscript*.

DiNardo, John, Nicole M Fortin, and Thomas Lemieux (1996). "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach". In: *Econometrica*, pp. 1001–1044.

Doms, Mark, Timothy Dunne, and Kenneth R Troske (1997). "Workers, Wages, and Technology". In: *The Quarterly Journal of Economics* 112.1, pp. 253–290.

Dorn, David (Sept. 2009). "Essays on Inequality, Spatial Interaction, and the Demand for Skills". PhD thesis. University of St. Gallen. URL: https://www.ddorn.net/data/Dorn_Thesis_Appendix.pdf (visited on 01/22/2019).

Eloundou, Tyna et al. (2023). "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models". In: *arXiv preprint arXiv:2303.10130*.

Felten, Edward, Manav Raj, and Robert Channing Seamans (2019). "The Effect of Artificial Intelligence on Human Labor: An Ability-Based Approach". In: *Academy of Management Proceedings*. Vol. 2019. 1. Academy of Management Briarcliff Manor, NY 10510, p. 15784.

Felten, Edward W, Manav Raj, and Robert Seamans (2018). "A Method to Link Advances in Artificial Intelligence to Occupational Abilities". In: *AEA Papers and Proceedings* 108, pp. 54–57.

— (2023). "Occupational Heterogeneity in Exposure to Generative AI". In: *Available at SSRN 4414065*.

Frey, Carl Benedikt (2019). *The Technology Trap*. Princeton University Press.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (Sept. 2019). "Text as Data". In: *Journal of Economic Literature* 57.3, pp. 535–74. DOI: 10.1257/jel.20181020. URL: https://www.aeaweb.org/articles?id=10.1257/jel.20181020.

Goldin, Claudia and Lawrence F. Katz (1998). "The Origins of Technology-Skill Complementarity". In: *The Quarterly Journal of Economics* 113.3, pp. 693–732.

— (2008). *The Race Between Education and Technology*. Harvard University Press.

Goldin, Claudia and Robert A Margo (1992). "The Great Compression: The Wage Structure in the United States at Mid-Century". In: *The Quarterly Journal of Economics* 107.1, pp. 1–34.

Goos, Maarten and Alan Manning (2007). "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain". In: *Review of Economics and Statistics* 89.1, pp. 118–133.

Goos, Maarten, Alan Manning, and Anna Salomons (May 2009). "Job Polarization in Europe". In: *American Economic Review* 99.2, pp. 58–63. DOI: 10.1257/aer.99.2.58.

— (2014). "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring". In: *American Economic Review* 104.8, pp. 2509–2526.

Haanwinckel, Daniel (2020). "Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution". In: *UCLA Working Paper*.

Hirvonen, Johannes, Aapo Stenhammar, and Joonas Tuhkuri (2022). "New Evidence on the Effect of Technology on Employment and Skill Demand". In: *Available at SSRN 4081625*.

Hsieh, Chang-Tai et al. (2019). "The Allocation of Talent and US Economic Growth". In: *Econometrica* 87.5, pp. 1439–1474.

Humlum, Anders (2021). *Robot Adoption and Labor Market Dynamics*. University of Chicago Working Paper.

Katz, Lawrence F. and David H. Autor (1999). "Changes in the Wage Structure and Earnings Inequality". In: *Handbook of Labor Economics*. Vol. 3. Elsevier, pp. 1463–1555.

Katz, Lawrence F. and Kevin M Murphy (1992). "Changes in Relative Wages, 1963–1987: Supply and Demand Factors". In: *The Quarterly Journal of Economics* 107.1, pp. 35–78.

Kelly, Bryan et al. (Sept. 2021). "Measuring Technological Innovation over the Long Run". In: *American Economic Review: Insights* 3.3, pp. 303–20. DOI: 10.1257/aeri.20190499. URL: https://www.aeaweb.org/articles?id=10.1257/aeri.20190499.

Kim, Gueyon (2022). *Trade-Induced Adoption of New Work*. IZA Discussion Paper, No. 15165.

Kogan, Leonid et al. (2019). *Technological Change and Occupations over the Long Run*. Working Paper 3585676. SSRN. DOI: 10.2139/ssrn. URL: https://ssrn.com/abstract=3585676.

Kogan, Leonid et al. (2021). *Technology-Skill Complementarity and Labor Displacement: Evidence from Linking Two Centuries of Patents with Occupations*. Tech. rep. National Bureau of Economic Research.

Kovak, Brian K (2013). "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?" In: *American Economic Review* 103.5, pp. 1960–76.

Krusell, Per et al. (2000). "Capital-skill Complementarity and Inequality: A Macroeconomic Analysis". In: *Econometrica* 68.5, pp. 1029–1053.

Leonardi, Marco (2015). "The Effect of Product Demand on Inequality: Evidence from the United States and the United Kingdom". In: *American Economic Journal: Applied Economics* 7.3, pp. 221–47.

Lewis, Ethan (2011). "Immigration, Skill Mix, and Capital Skill Complementarity". In: *The Quarterly Journal of Economics* 126.2, pp. 1029–1069.

Lewis, Mike et al. (2019). "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension". In: *CoRR* abs/1910.13461. arXiv: 1910.13461. URL: http://arxiv.org/abs/1910.13461.

Lin, Jeffrey (2011). "Technological Adaptation, Cities, and New Work". In: *Review of Economics and Statistics* 93.2, pp. 554–574. URL: https://EconPapers.repec.org/RePEc:tpr:restat:v:93:y:2011:i:2:p:554-574.

Loo, M.P.J. van der (2014). "The Stringdist Package for Approximate String Matching". In: *R Journal* 6.1, pp. 111–122.

Mann, Katja and Lukas Püttmann (2021). "Benign Effects of Automation: New Evidence from Patent Texts". In: *The Review of Economics and Statistics*, pp. 1–45.

Mazzolari, Francesca and Giuseppe Ragusa (2013). “Spillovers from High-Skill Consumption to Low-Skill Labor Markets”. In: *Review of Economics and Statistics* 95.1, pp. 74–86.

Michaels, Guy, Ashwini Natraj, and John Van Reenen (2014). “Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years”. In: *Review of Economics and Statistics* 96.1, pp. 60–77.

Michel, Jean-Baptiste et al. (2011). “Quantitative Analysis of Culture Using Millions of Digitized Books”. In: *Science* 331.6014, pp. 176–182.

Mikolov, Tomas et al. (2013). “Efficient Estimation of Word Representations in Vector Space”. In: *arXiv preprint arXiv:1301.3781*.

Moser, Petra (2012). “Innovation without Patents: Evidence from World’s Fairs”. In: *The Journal of Law and Economics* 55.1, pp. 43–74.

Mullahy, John and Edward C Norton (Dec. 2022). *Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes*. Tech. rep. NBER Working Paper No. 30735.

Noy, Shakked and Whitney Zhang (2023). “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence”. In: *Science* 381, pp. 187–192.

Peng, Sida et al. (2023). “The Impact of AI on Developer Productivity: Evidence from Github Copilot”. In: *arXiv preprint arXiv:2302.06590*.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (Oct. 2014). “GloVe: Global Vectors for Word Representation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543. DOI: 10.3115/v1/D14-1162. URL: https://www.aclweb.org/anthology/D14-1162.

Pierce, Justin R and Peter K Schott (2016). “The Surprisingly Swift Decline of US Manufacturing Employment”. In: *American Economic Review* 106.7, pp. 1632–62.

Ray, Debraj and Dilip Mookherjee (2021). “Growth, Automation, and the Long-Run Share of Labor”. In: *Review of Economic Dynamics*.

Ruggles, Steven, Catherine A. Fitch, et al. (2021). *IPUMS Ancestry Full Count Data: Version 3.0 [dataset]*. Minneapolis, MN. URL: https://doi.org/10.18128/D014.V3.0.

Ruggles, Steven, Sarah Flood, et al. (2022). *IPUMS USA: Version 12.0*. Minneapolis, MN.

Susskind, Daniel (2020). “A Model of Task Encroachment in the Labour Market”. In: *Oxford University Discussion Paper* 819.

Topalova, Petia (2010). “Factor Immobility and Regional Impacts of trade Liberalization: Evidence on Poverty from India”. In: *American Economic Journal: Applied Economics* 2.4, pp. 1–41.

U. S. Department of Labor, Employment and Training Administration (1939). *Dictionary of Occupational Titles (First Edition*. U.S. Government Printing Office.

— (1977). *Dictionary of Occupational Titles: Fourth Edition*. U.S. Government Printing Office.

U.S. Department of Labor: Employment and Training Administration (2022). *O*NET OnLine: Solar Energy Installation Managers, 47-1011.03*. URL: https://www.onetonline.org/link/summary/47-1011.03 (visited on 09/21/2023).

United States Census (1900). *Database with Images: Orville Wright in Entry for Milton Wright, 1900*. Tech. rep. URL: https://www.familysearch.org/ark:/61903/1:1:MM6W-31L.

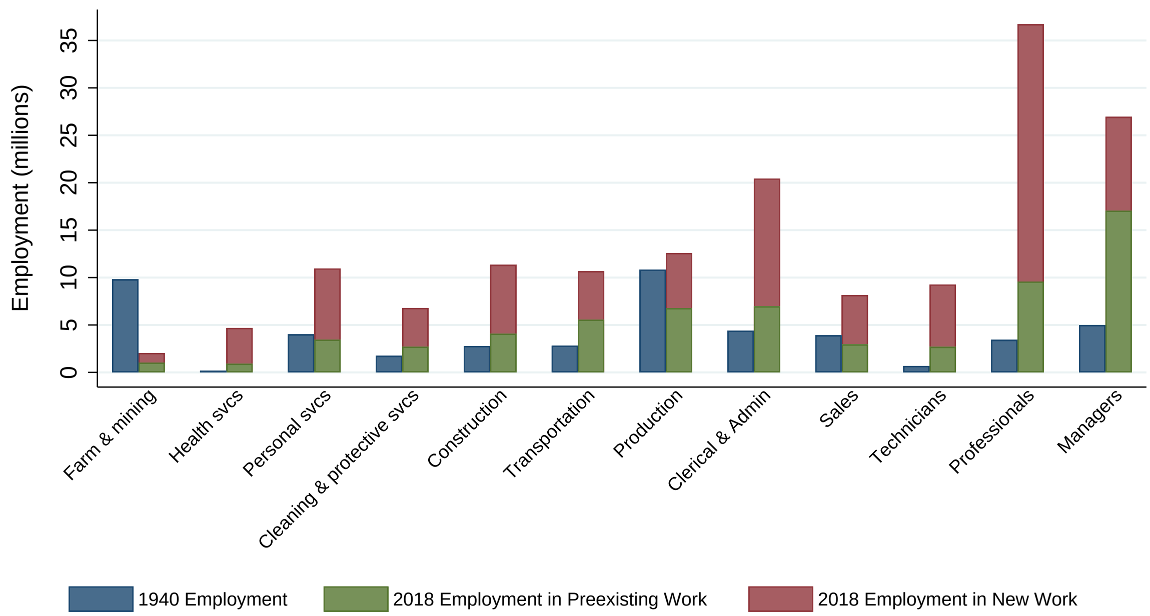
— (1910). *Database with Images: Orville Wright in Household of Milton Wright, Dayton Ward 5, Montgomery, Ohio, United States*. Tech. rep. URL: https://familysearch.org/ark:/61903/1:1:ML64-97P.

US Census Bureau (1915–2018). *Census of Population: Alphabetical Index of Industries and Occupations*. US Bureau of the Census.

Vogel, Jonathan (2023). *The Race between Education, Technology, and the Minimum Wage*. Tech. rep. NBER Working Paper No. 31028.

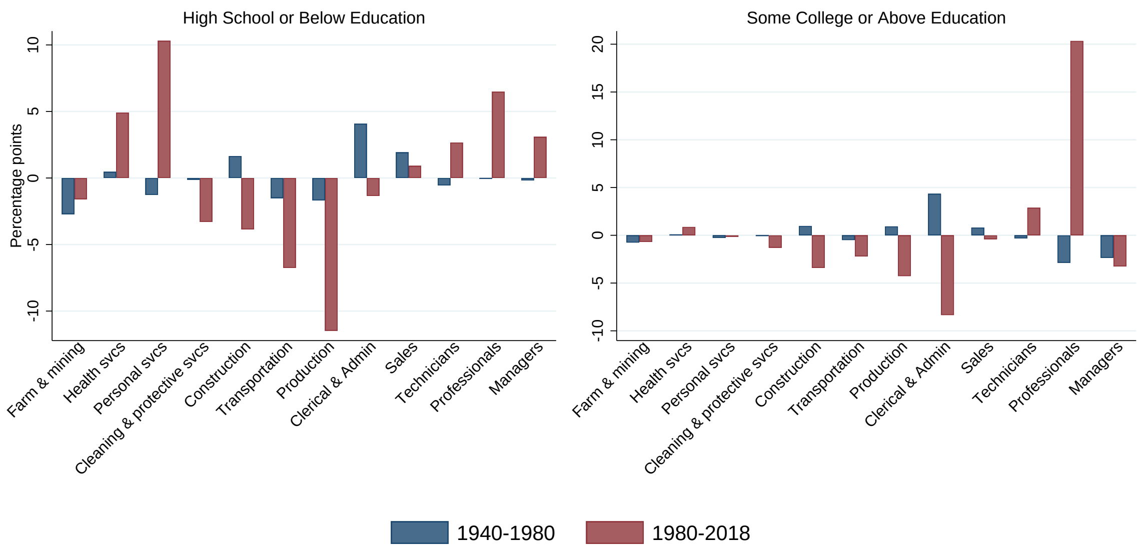
Webb, Michael (2020). *The Impact of Artificial Intelligence on the Labor Market*. Working Paper.

FIGURE I
Employment Counts by Broad Occupation in 1940 and 2018, Distinguishing Between Titles Present in 1940 Versus Those Added Subsequently



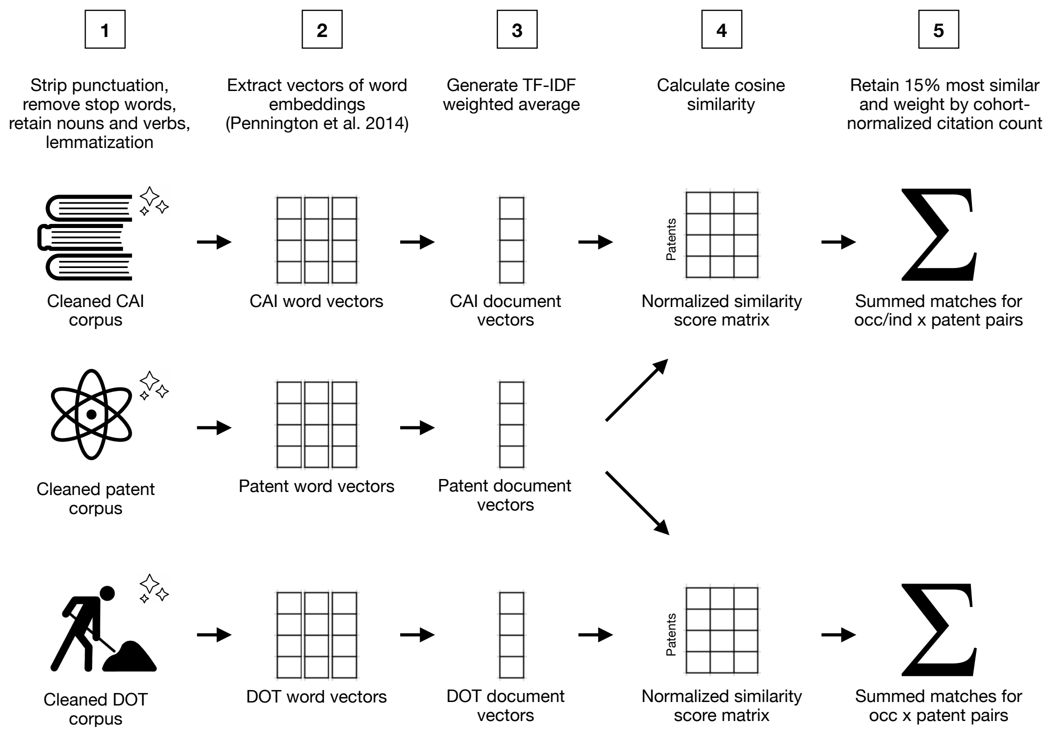
This figure reports the distribution of employment in 1940 and 2018 across broad occupational categories ordered from lowest to highest-paying. The first set of bars show 1940 employment. The second set of bars show 2018 employment, divided into estimated 2018 employment in occupational titles that existed in 1940 (bottom), and estimated 2018 employment in occupation titles that were added since 1940 (top). Employment in 2018 is estimated by constructing a cumulative new title share in each broad occupation—summing the number of new titles added over 1940–2018—and dividing this by the total number titles in the 2018 index adjusted for (the small number of) titles that were removed. See Appendix D3 for additional details.

FIGURE II
Difference Between the Occupational Distribution of the Flow of New Work Versus the Stock of Preexisting Work by Education Group, 1940–1980 and 1980–2018



This figure reports the difference in the share of employment in new work vs. the share of employment in existing work separately by education group and period. Each panel reports this difference across twelve broad occupational categories ordered from lowest to highest-paying. The first set of bars represent the average difference in employment shares over 1940–1980, while the second set of bars represent the difference over 1980–2018.

FIGURE III
Linking Patents to Occupations and Industries

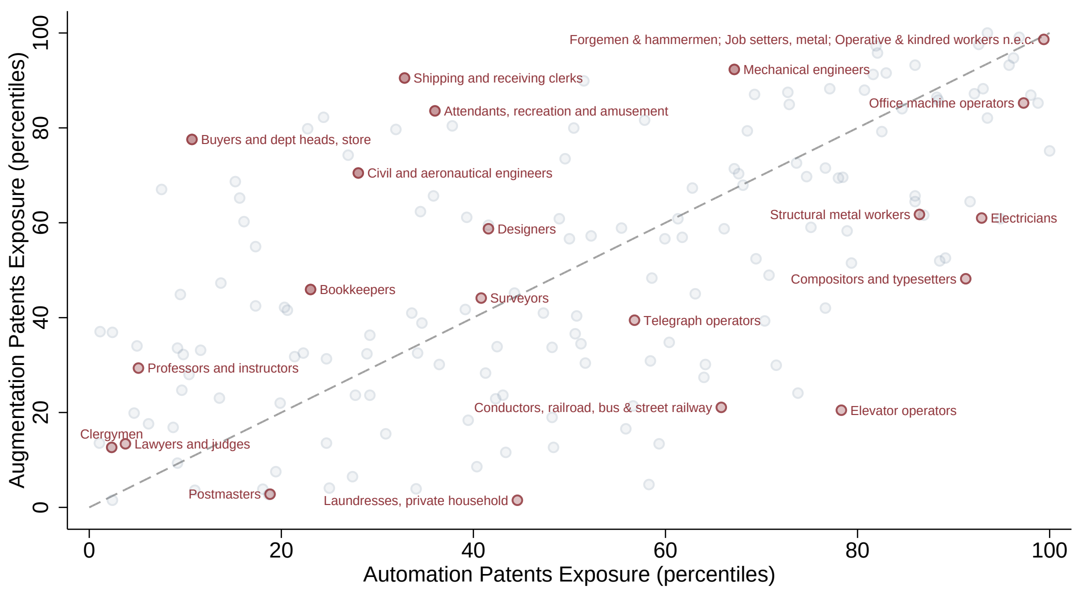


This figure summarizes our five-step procedure for linking patents to occupations. The middle row describes the method for cleaning and processing patent data in preparation for creating the linkages. The top row depicts the method for creating augmentation-based patent matches using occupation and industry titles from the CAI. The bottom row depicts the method for creating automation-based patent matches using DOT task descriptions.

FIGURE IV

The Relationship between Exposure to Automation and Augmentation Patents at the Occupation Level

Panel A. 1940-1980 Average

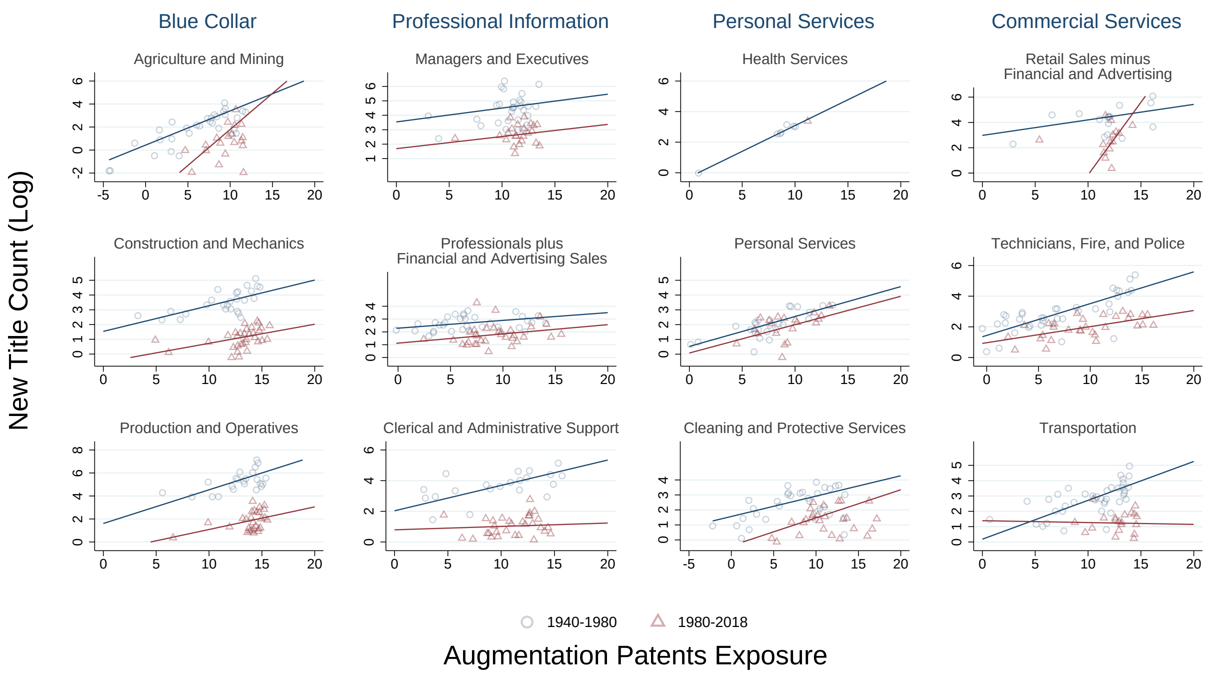


Panel B. 1980-2018 Average

This figure presents a scatter plot of the relationship between occupational exposure to automation and augmentation patents for 1940–1980 (panel A) and 1980–2018 (panel B). Each point corresponds to the average percentile of automation ($x-$axis) and augmentation ($y-$axis) exposure of one consistently defined three-digit Census occupation, where the average is taken over 1940–1980 ($N = 166$ occupations per year) in panel A and over 1980–2018 ($N = 306$ occupations per year) in panel B. The 45 degree line in each panel is plotted with dashes.

FIGURE V

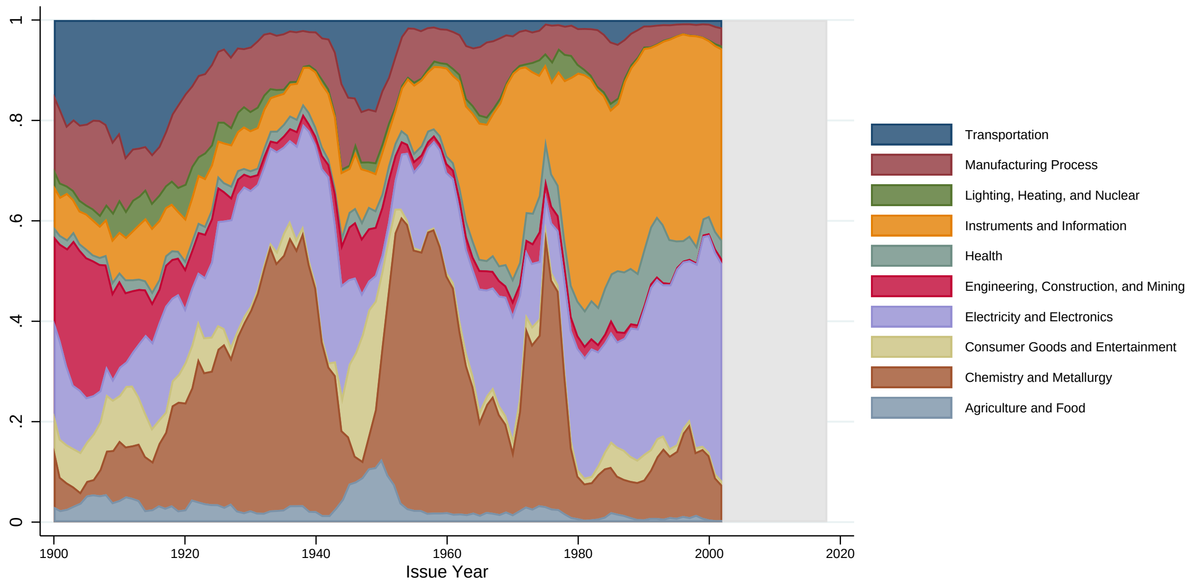
The Relationship between Exposure to Augmentation Patents and Occupational New Title Emergence



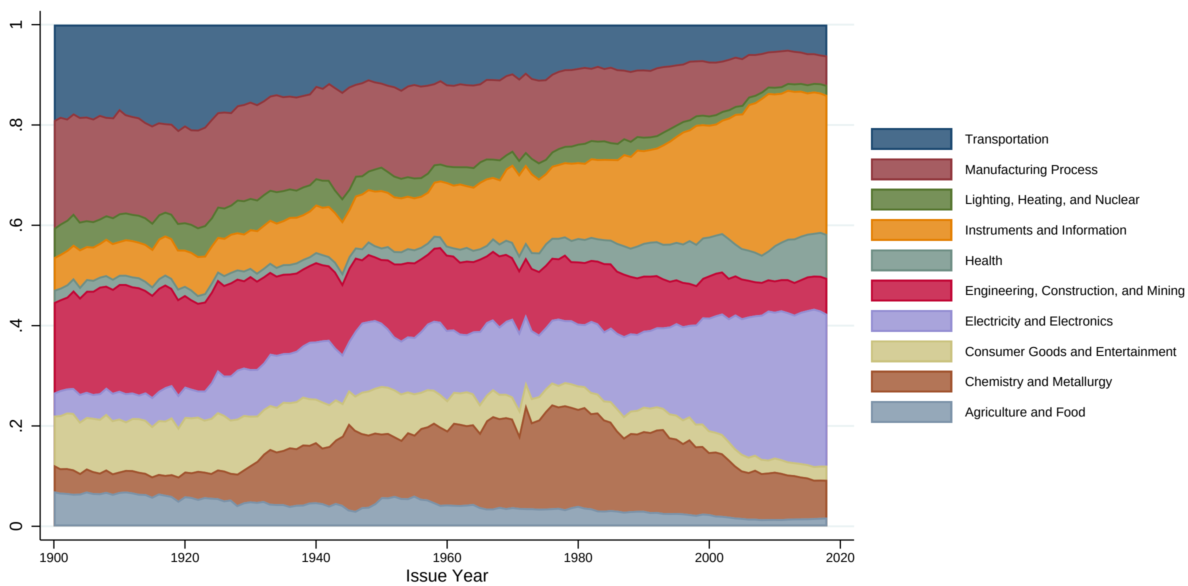
Each panel of the figure is a bin scatter showing the relationship between augmentation exposure and new title emergence within broad occupations. The x-axis plots the logarithm of augmentation patent matches, and the y-axis plots the logarithm of new micro-title counts. Observations with zero new titles or zero augmentation innovations are dropped in this log-log plot. Circles correspond to data for 1940–1980, and triangles correspond to data for 1980–2018.

FIGURE VI
Breakthrough Patent Flows by Class Vs. All Patent Flows by Class

Panel A. The flow of top 10% breakthrough patents, 1900–2002



Panel B. The flow of all patents, 1900-2018

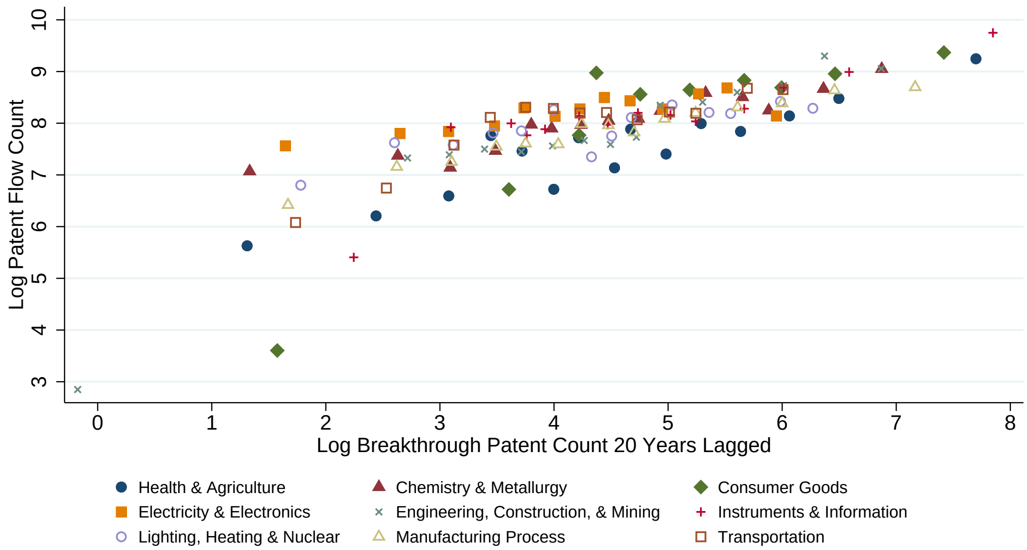


This figure presents the distribution of breakthrough (panel A) and total (panel B) patents by 10 broad technological classes in each year between 1900 and 2000 (breakthroughs) and 1900 and 2018 (all patents). We use the classification of patent 3-digit CPC classes into broad technology groups from Kelly et al. (2021), while excluding patents assigned to the broad classes 'weapons' and 'other', which together make up less than 1 percent of all patents.

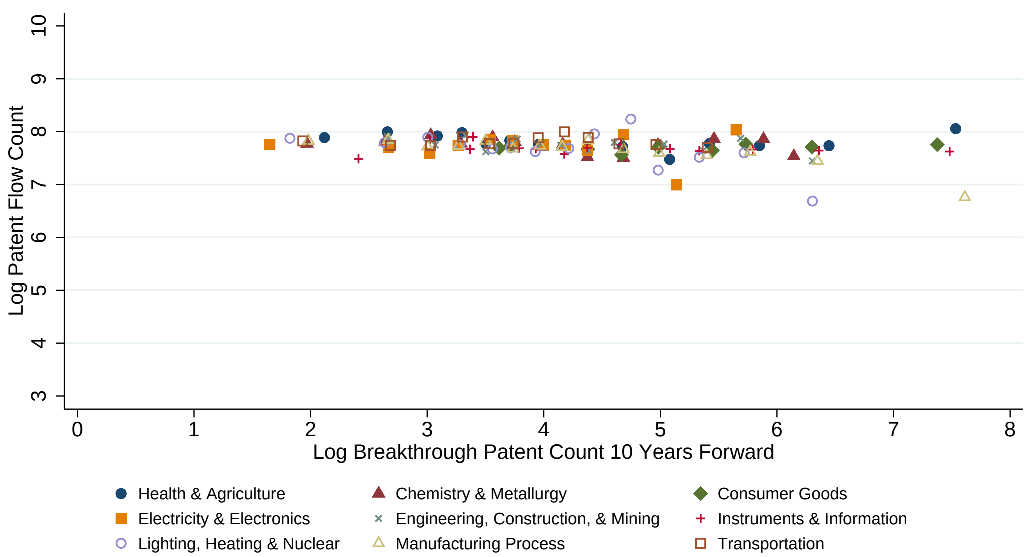
FIGURE VII

Predictive Relationship between Breakthrough Patents and 20-Year Forward Patent Flows and 10-Year Backward Patent Flows by Technology Class

Panel A. Correlation between patent flows and lagged breakthroughs by technology class



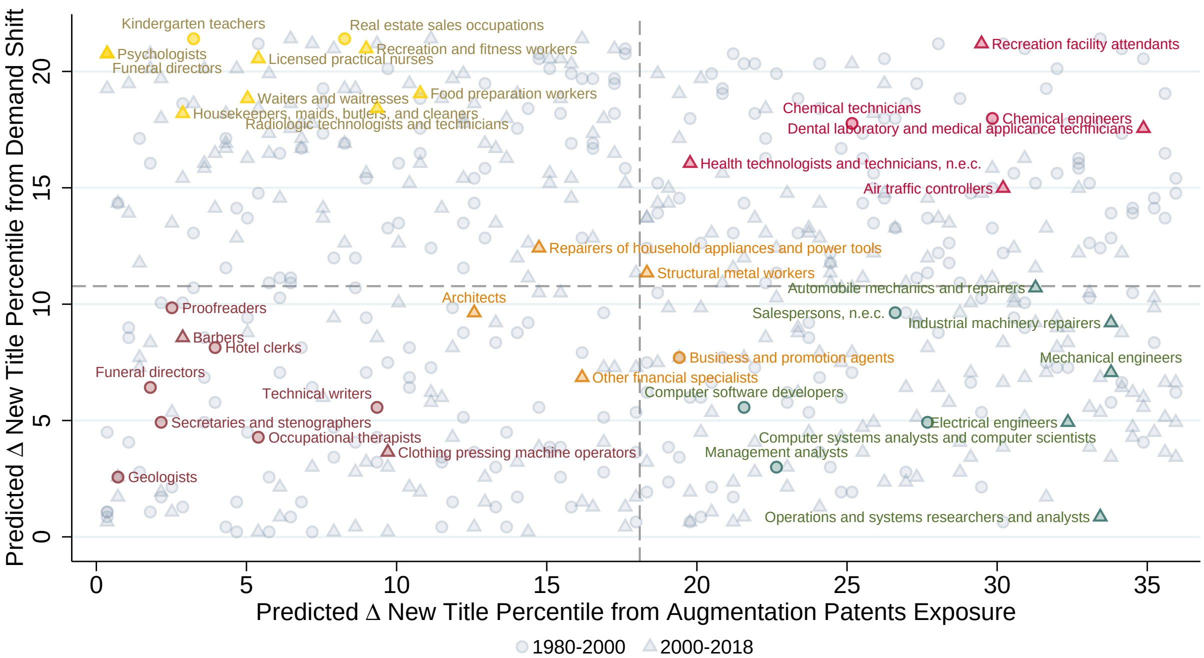
Panel B. Correlation between patent flows and future breakthroughs by technology class



This figure presents conditional correlations between breakthrough technology flows and subsequent patent flows in panel A, and prior patent flows in panel B. Panel A regresses log patent flows by technology class and decade on log breakthrough patent flows lagged by two decades, while controlling for log non-breakthrough patent flow counts lagged by two decades and decade fixed effects: $\hat{\beta} = 0.42, \mathrm{SE} = 0.03, N = 1,051$. Panel B regresses log patent flows by technology class and decade on log breakthrough patent flows one decade forward, while controlling for log non-breakthrough patent flow counts one decade forward and decade fixed effects: $\beta = -0.02, \mathrm{SE} = 0.02$, N=1,044. Standard errors clustered by 3-digit technology class.

FIGURE VIII

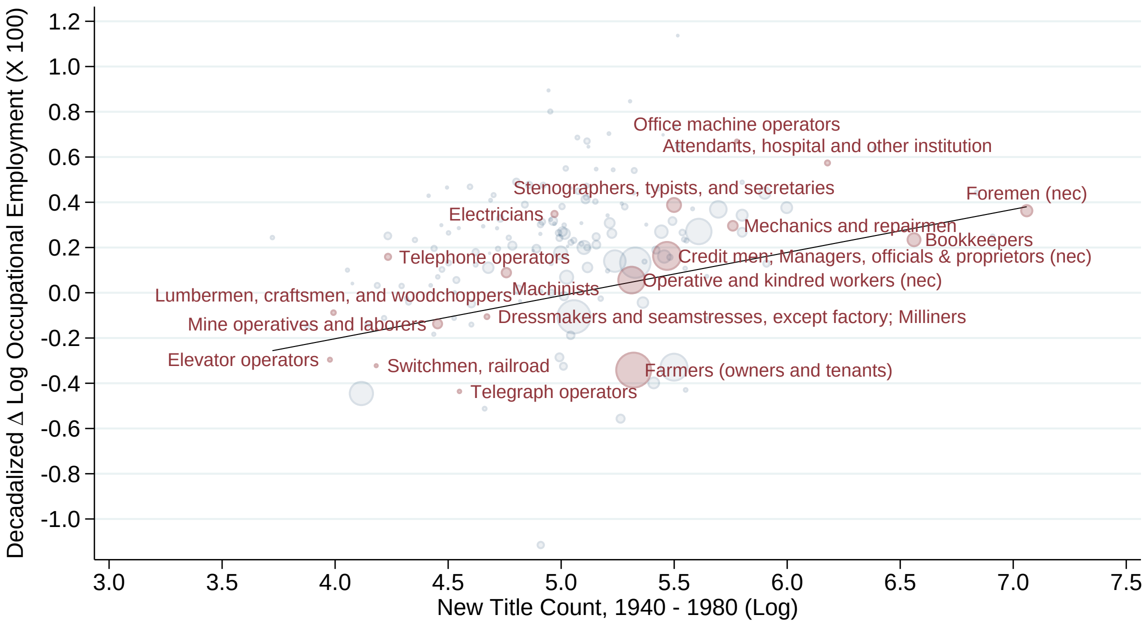
Predicted Δ Occupational New Title Share Percentile From Exposure to Augmentation vs. Exposure to Demand Shifts, 1980–2000 and 2000–2018



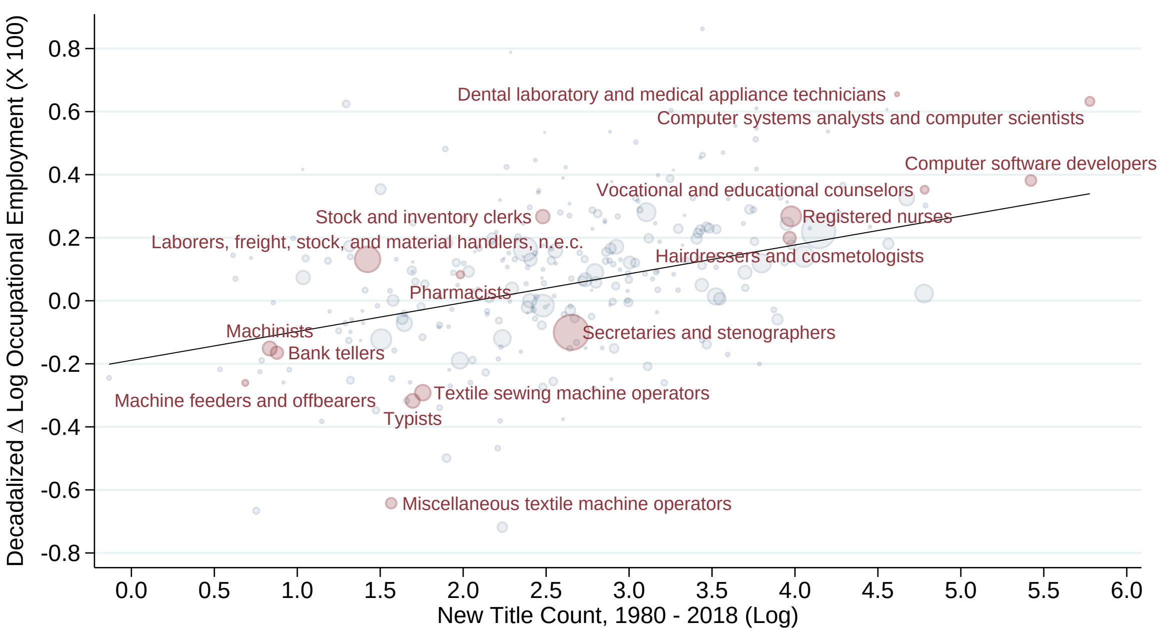
This figure plots partial predicted new title flows from exposure to demand shifts against partial predicted new title flows due to augmentation exposure (both in percentile terms) for consistent Census occupations. Predictions are based on column 4 of Table A.X. Dashed lines indicate median predicted changes.

FIGURE IX
Correlations between Employment Growth and New Title Emergence Rates, 1940–1980 and 1980–2018

Panel A. 1940–1980



Panel B. 1980–2018

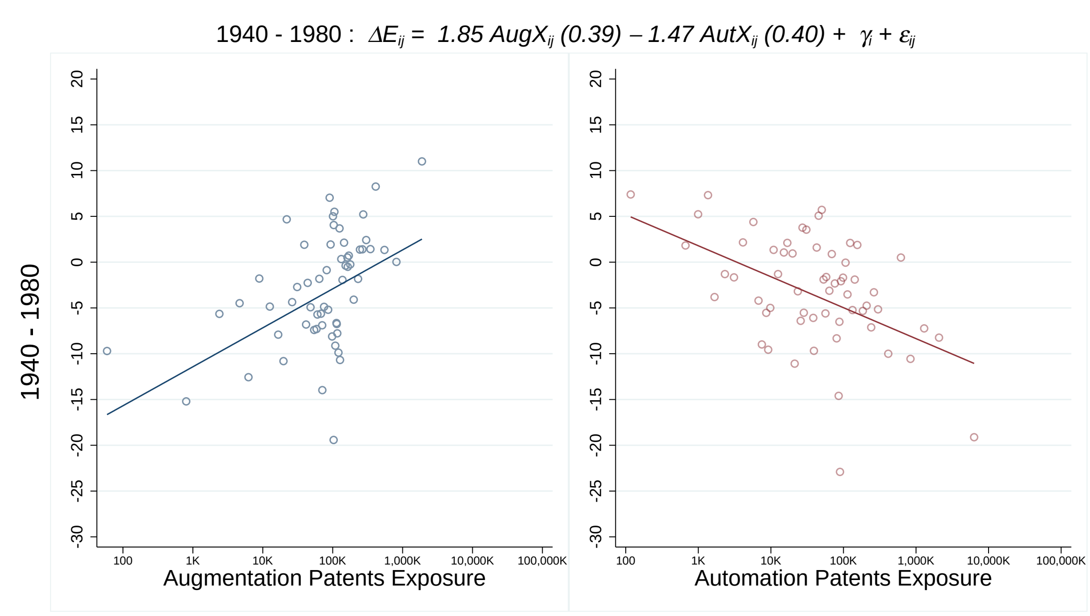


This figure illustrates the relationship between log new title counts and decadalized occupational employment growth, controlling for initial-year log title counts, and using initial-year employment shares as weights. Plotted lines correspond to weighted partial fitted values. The weighted partial correlation between log employment growth and the log of new title flows is 0.333 for 1940–1980 and 0.482 for 1980–2018.

FIGURE X

Conditional Correlations between Automation, Augmentation and Employment Growth (based on OLS regressions)

Panel A. 1940–1980



Panel B. 1980–2018

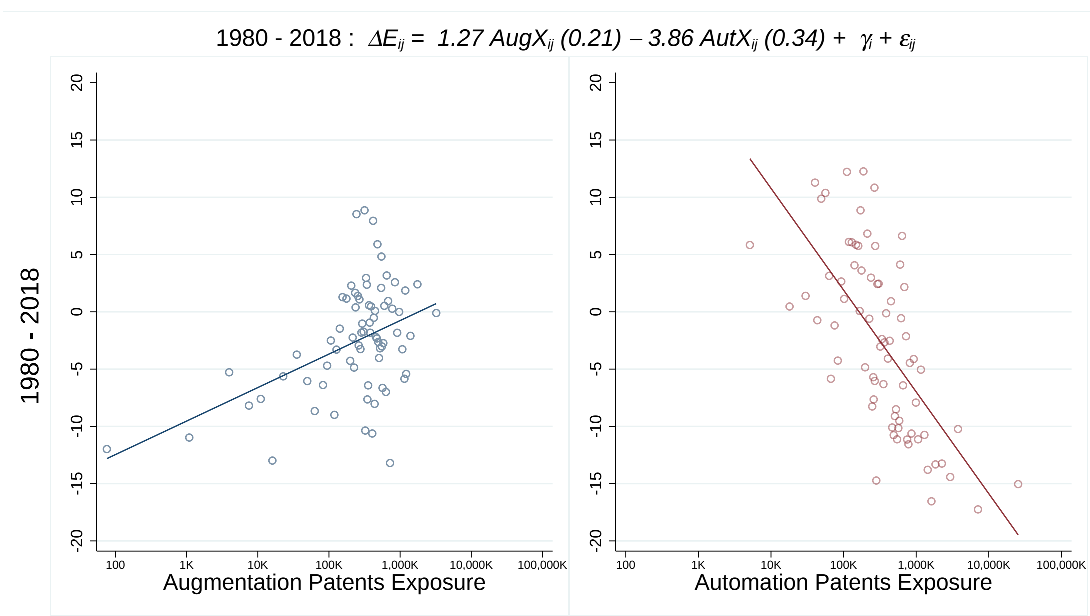


Figure reports bin scatters of the employment-weighted conditional correlation between decadalized percent employment growth and exposure to augmentation innovations (left-hand side) and automation innovations (right-hand side). Panel A corresponds to the regression specification in column (1) panel A of Table VII using consistently defined Census industry×occupation cells over 1940–1980 (N=6,520). Panel B corresponds to the regression specification in column (5) panel A of Table VII using consistently defined Census industry×occupation cells over 1980–2018 (N= 27,380).

TABLE I

EXAMPLES OF NEW TITLES ADDED TO THE *CENSUS ALPHABETICAL INDEX OF OCCUPATIONS* BY VOLUME, 1940–2018

Volume Year	Example Titles Added	
1940	Automatic welding machine operator	Acrobatic dancer
1950	Airplane designer	Tattooer
1960	Textile chemist	Pageants director
1970	Engineer computer application	Mental-health counselor
1980	Controller, remotely-piloted vehicle	Hypnotherapist
1990	Circuit layout designer	Conference planner
2000	Artificial intelligence specialist	Amusement park worker
2010	Technician, wind turbine	Sommelier
2018	Cybersecurity analyst	Drama therapist

Table reports examples of new titles added to the Census Alphabetical Index of Occupations volume of year Y correspond to titles recognized by Census coders between the start of the prior decade and the year preceding the volume's release, e.g., the 1950 CAI volume includes new titles incorporated between 1940 and 1949.

TABLE II
Occupational New Title Emergence and Augmentation versus Automation Exposure, 1940–2018

Dependent Variable: Occupational New Title Count

	(1)	(2)	(3)	(4)	(5)
			A. 1940–2018		
Augmentation Exposure	17.81***	21.46***		16.85***	21.02***
	(3.52)	(3.74)		(3.96)	(3.54)
Automation Exposure			12.75**	1.89	2.35
			(3.93)	(4.52)	(4.07)
			B. 1940–1980		
Augmentation Exposure	23.46***	27.23***		19.48***	26.11***
	(5.09)	(4.80)		(5.79)	(4.45)
Automation Exposure			19.56***	8.15+	9.07+
			(4.21)	(4.85)	(4.87)
			C. 1980–2018		
Augmentation Exposure	8.14*	12.35***		14.87***	13.86***
	(4.08)	(2.31)		(3.72)	(2.08)
Automation Exposure			−1.21	−12.99*	−5.43
			(5.07)	(5.37)	(6.19)
Occ Emp Shares	X	X	X	X	X
Time FE	X		X	X	
Broad Occ × Time FE		X			X

Negative binomial models, coefficients multiplied by 100. N = 1,535 in Panel A; N = 628 in Panel B; N = 907 in Panel C. Twelve broad occupations are defined consistently across all decades. Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE III

First Stage Estimates for New Titles Regressions, 1940–2018

Dependent Variable: Log Patent Count

	(1)	(2)	(3)	
	Aug	Aut	Aug	Aut
		A. 1940–2018		
Augmentation IV	2.26***		2.20***	0.23
	(0.41)		(0.43)	(0.32)
Automation IV		2.40***	0.67	2.37***
		(0.32)	(0.65)	(0.36)
F-stat	30.88	56.59	27.50	28.96
Sanderson-Windmeijer F-stat	30.88	56.59	49.47	35.39
		B. 1940–1980		
Augmentation IV	1.17+		1.70**	−0.01
	(0.62)		(0.54)	(0.33)
Automation IV		3.45***	2.19+	3.58***
		(0.66)	(1.16)	(0.75)
F-stat	3.63	27.21	8.20	11.46
Sanderson-Windmeijer F-stat	3.63	27.21	8.94	9.29
		C. 1980–2018		
Augmentation IV	2.82***		2.94***	0.78**
	(0.35)		(0.48)	(0.30)
Automation IV		1.66***	−0.82	1.35***
		(0.20)	(0.73)	(0.23)
F-stat	65.73	68.15	40.91	29.68
Sanderson-Windmeijer F-stat	65.73	68.15	62.03	30.37
Occ Emp Shares	X	X	X	X
Time FE	X	X	X	X

N = 1,276 in Panel A; N = 589 in Panel B; N = 687 in Panel C. First stage estimates for columns 1, 3, and 4 in Table IV. Specifications additionally control for indicators for being above the cross-sectional median in total augmenting or automating patent matches two decades prior, and whether the occupation has greater than 50% employment in manufacturing industries. Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE IV

2SLS ESTIMATES FOR NEW TITLES REGRESSIONS, 1940–2018

Dependent Variable: Log Occupational New Title Count

	(1)	(2)	(3)	(4)	(5)
			A. 1940–2018		
Augmentation Exposure	24.42**	21.30***		29.54***	19.52**
	(7.46)	(5.81)		(8.58)	(6.66)
Automation Exposure			−3.01	−14.56	4.21
			(11.82)	(11.29)	(9.71)
F-stat (Aug)	30.88	60.16		27.50	37.16
F-stat (Aut)			56.59	28.96	52.05
			B. 1940–1980		
Augmentation Exposure	64.64*	30.17***		56.79**	24.82*
	(29.15)	(8.79)		(19.88)	(11.78)
Automation Exposure			9.80	−17.29	15.75
			(15.16)	(17.23)	(17.74)
F-stat (Aug)	3.63	11.04		8.20	8.72
F-stat (Aut)			27.21	11.46	13.79
			C. 1980–2018		
Augmentation Exposure	7.80+	10.50+		21.08**	15.36*
	(4.67)	(6.31)		(7.11)	(5.93)
Automation Exposure			−22.70	−26.73+	−13.43
			(14.99)	(13.60)	(9.80)
F-stat (Aug)	65.73	52.32		40.91	37.44
F-stat (Aut)			68.15	29.68	35.90
Occ Emp Shares	X	X	X	X	X
Time FE	X		X	X	
Broad Occ × Time FE		X			X

N = 1,276 in Panel A; N = 589 in Panel B; N = 687 in Panel C. Coefficients multiplied by 100. Specifications additionally control for indicators for being above the cross-sectional median in total augmenting or automating patent matches two decades prior, and whether the occupation has greater than 50% employment in manufacturing industries. Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE V

Occupational New Title Emergence and Demand Contractions and Expansions

Dependent Variable: Occupational New Title Count

	(1)	(2)	(3)	(4)	(5)
			A. Negative Trade Shock *Years 1990-2000 & 2000-2018*		
Import Exposure	−9.83	−15.44**	−12.13*	−17.49***	−17.73***
	(6.09)	(5.23)	(5.53)	(5.13)	(5.17)
Augmentation Exposure			7.94+	9.38**	8.32**
			(4.60)	(3.00)	(2.91)
			B. Negative Trade Shock, Placebo Test *Years 1970-1980 & 1980-1990*		
Import Exposure	14.50	3.95	11.77	−2.99	−1.76
	(26.06)	(20.40)	(20.47)	(13.24)	(12.53)
Augmentation Exposure			19.57***	20.00***	20.60***
			(3.15)	(1.77)	(1.92)
			C. Positive Demographic Shock *Years 1980-2000 & 2000-2018*		
Demand Shift Exposure	18.39***	12.83+	18.53***	14.86*	14.75*
	(5.49)	(6.66)	(5.35)	(6.71)	(6.60)
Augmentation Exposure			5.36	11.17***	10.56***
			(4.97)	(2.43)	(2.37)
Time FE	X	X	X	X	X
Occ Emp Shares	X	X	X	X	X
Ind Exposure Control	X	X	X	X	X
Broad Occ FE		X		X	X
Δ Occ Emp Shares					X

Negative binomial models, coefficients multiplied by 100. N = 610 in Panels A and C; N = 588 in Panel B. Panel A estimates the relationship between new titles emerging 1990–2018 and occupational exposure to import competition in the same period. Panel B (a placebo test) estimates the relationship between new titles emerging 1970–1990 and occupational exposure to import competition over the 1990-2018 period. Panel C estimates the relationship between new titles emerging 1980–2000 and 2000—2018 and occupational exposure to demographically-driven demand shocks. Standard errors clustered by occupation in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE VI
The Relationship between Changes in Employment and Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, OLS and 2SLS Stacked Long-Difference Regressions, 1940–2018

	OLS		2SLS	
	(1)	(2)	(3)	(4)
	A. Decadalized ΔLn(Employment)			
Augmentation Exposure	1.51***	1.36***	4.37***	3.63***
	(0.21)	(0.22)	(0.93)	(0.96)
Automation Exposure	−2.27***	−1.00*	−4.02***	−4.21***
	(0.27)	(0.40)	(0.62)	(0.93)
R^2	0.53	0.57		
	B. Decadalized ΔLn(Wagebill)			
Augmentation Exposure	1.50***	1.41***	4.95***	4.04***
	(0.23)	(0.25)	(1.02)	(0.98)
Automation Exposure	−2.21***	−0.97*	−3.17***	−3.43***
	(0.28)	(0.42)	(0.67)	(0.97)
R^2	0.53	0.57		
	C. Decadalized ΔE[Ln(Wagebill)]			
Augmentation Exposure	1.39***	1.30***	4.15***	3.31***
	(0.20)	(0.21)	(0.91)	(0.93)
Automation Exposure	−2.13***	−1.01**	−3.40***	−3.74***
	(0.27)	(0.38)	(0.63)	(0.91)
R^2	0.55	0.58		
	D. Decadalized ΔLn(Adjusted Wagebill)			
Augmentation Exposure	1.62***	1.47***	5.18***	4.36***
	(0.24)	(0.25)	(1.03)	(1.00)
Automation Exposure	−2.34***	−0.96*	−3.79***	−3.90***
	(0.28)	(0.44)	(0.66)	(0.99)
R^2	0.52	0.56		
F-stat (Aug)			127.76	150.56
F-stat (Aut)			202.85	145.10
Ind × Time FE	X	X	X	X
Broad Occ × Time FE		X		X

N = 33,900 changes in employment and wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill changes are winsorized at the 99th percentile. Wagebill variables are all hourly employment and all hourly wages. Hourly wages are imputed to cells where employment but not wages are observed. Specifications additionally control for indicators for being above the cross-sectional median in total augmenting or automating patent matches two decades prior. Standard errors clustered by industry-occupation cell (using Stata command `reghdfe` and `ivreghdfe`) in parentheses. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^+p < 0.10$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE VII

The Relationship between Changes in Employment and Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, OLS and 2SLS Stacked Long-Difference Regressions, 1940–1980 and 1980–2018

	1940–1980				1980–2018			
	OLS		2SLS		OLS		2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Decadalized ΔLn(Employment)							
Augmentation Exposure	1.85***	2.12***	5.78**	5.24*	1.27***	0.78***	3.35***	2.79**
	(0.39)	(0.38)	(2.21)	(2.42)	(0.21)	(0.23)	(0.91)	(0.93)
Automation Exposure	−1.47***	−0.27	−3.97*	−2.79	−3.86***	−1.98***	−5.05***	−5.07***
	(0.40)	(0.64)	(1.77)	(2.95)	(0.34)	(0.46)	(0.66)	(0.92)
R^2	0.63	0.66			0.44	0.48		
	B. Decadalized ΔLn(Adjusted Wagebill)							
Augmentation Exposure	2.04***	2.36***	6.93**	6.40*	1.30***	0.78***	3.86***	3.24***
	(0.45)	(0.45)	(2.33)	(2.54)	(0.23)	(0.23)	(0.96)	(0.96)
Automation Exposure	−1.63***	−0.39	−4.04*	−2.64	−3.81***	−1.74***	−4.80***	−4.86***
	(0.41)	(0.69)	(1.76)	(3.05)	(0.39)	(0.50)	(0.67)	(0.95)
R^2	0.61	0.64			0.44	0.49		
F-stat (Aug)			34.80	46.34			160.12	150.02
F-stat (Aut)			38.68	15.16			169.77	301.79
Ind × Time FE	X	X	X	X	X	X	X	X
Broad Occ × Time FE		X		X		X		X

N=6,520 changes in employment and wagebill in consistently defined Census occupations over 1940–1980, and N=27,380 changes over 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill changes are winsorized at the 99th percentile. Wagebill variables are all hourly employment and all hourly wages. Hourly wages are imputed to cells where employment but not wages are observed. 2SLS specifications additionally control for indicators for being above the cross-sectional median in total augmenting or automating patent matches two decades prior. Standard errors clustered by industry-occupation cell (using Stata command `reghdfe` and `ivreghdfe`) in parentheses. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

New Frontiers: The Origins and Content of New Work, 1940–2018

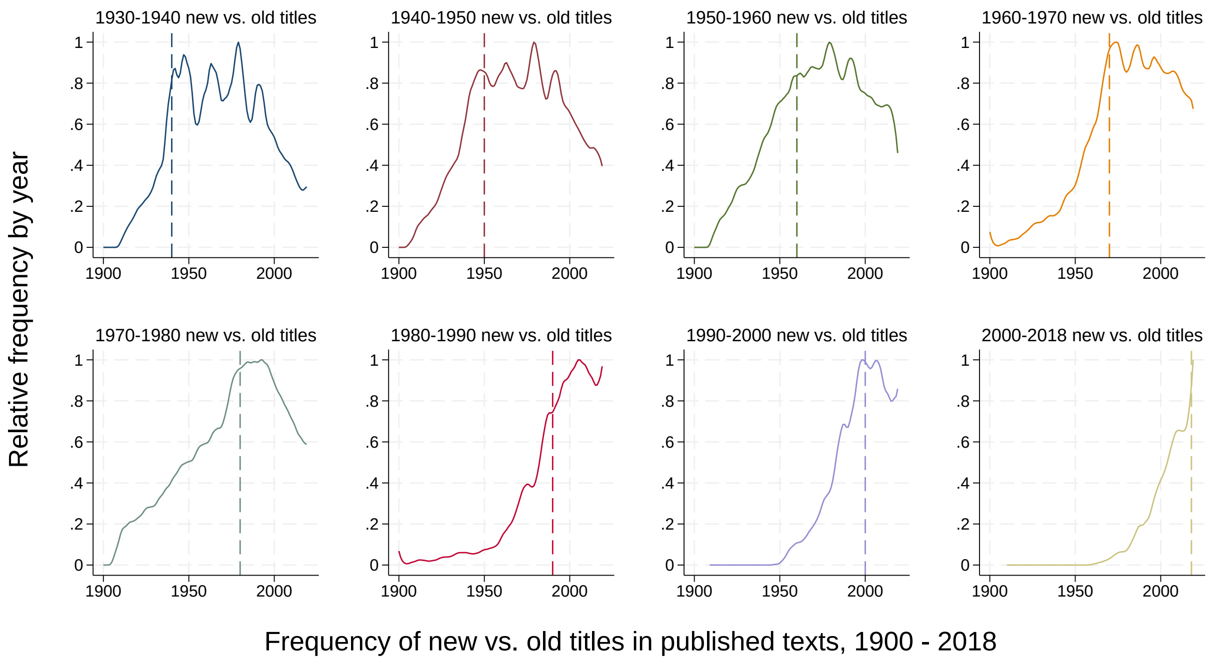
Online Appendix

David Autor
Caroline Chin
Anna Salomons
Bryan Seegmiller

Table of Contents

A Primary appendix figures and tables

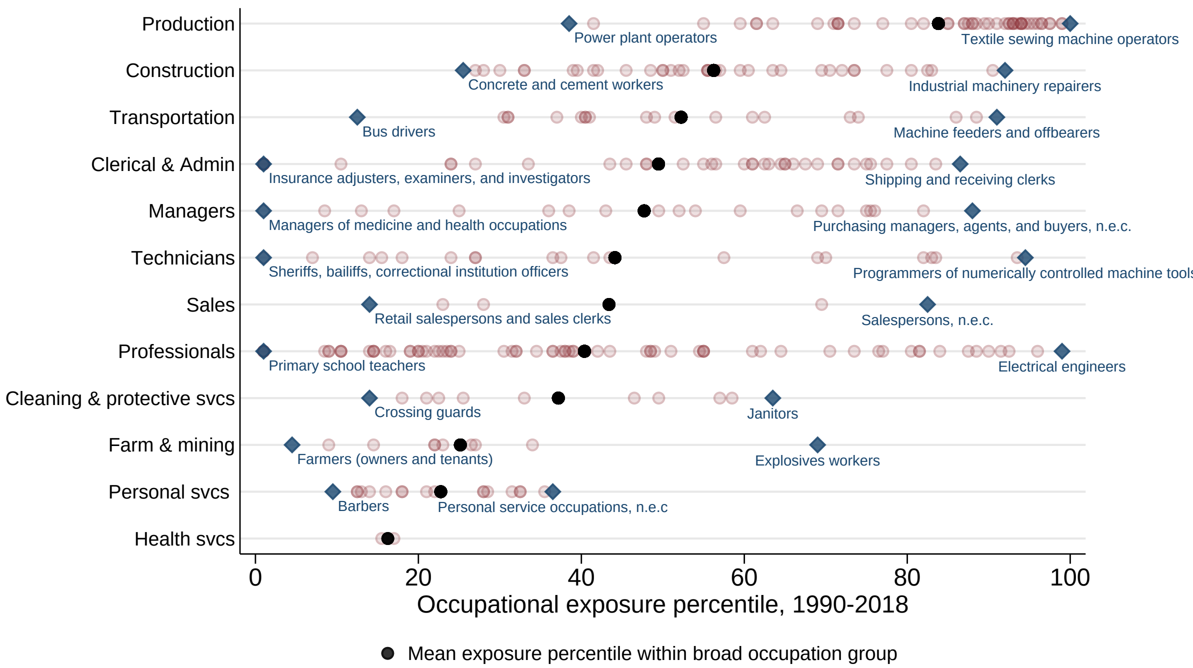
FIGURE A.I
Median Relative Usage Frequency in Published English Language Books of New Occupational Titles added to the *Census Alphabetical Index of Occupations* by Decade, 1940 – 2018



This figure is calculated using Google Ngram Viewer (Michel et al., 2011). It reports the median frequency of each decade's cohort of new titles relative to the median frequency of the decade's cohort of existing titles in digitized English language books published in the United States in every year between 1900 and 2018. Unmatched titles are dropped from the n-gram analysis. The frequency of a title of length n is defined as the number of occurrences of the title in a year as a fraction of the total number of n-grams in that year. The frequencies are normalized by cohort such that the maximum frequency across all years for each cohort is one. We calculate the ratio of normalized frequencies for new and existing titles for each year, then smooth the series of ratios for each cohort using locally weighted regression and, finally, renormalize so that the maximum relative frequency across all years for each cohort is one.

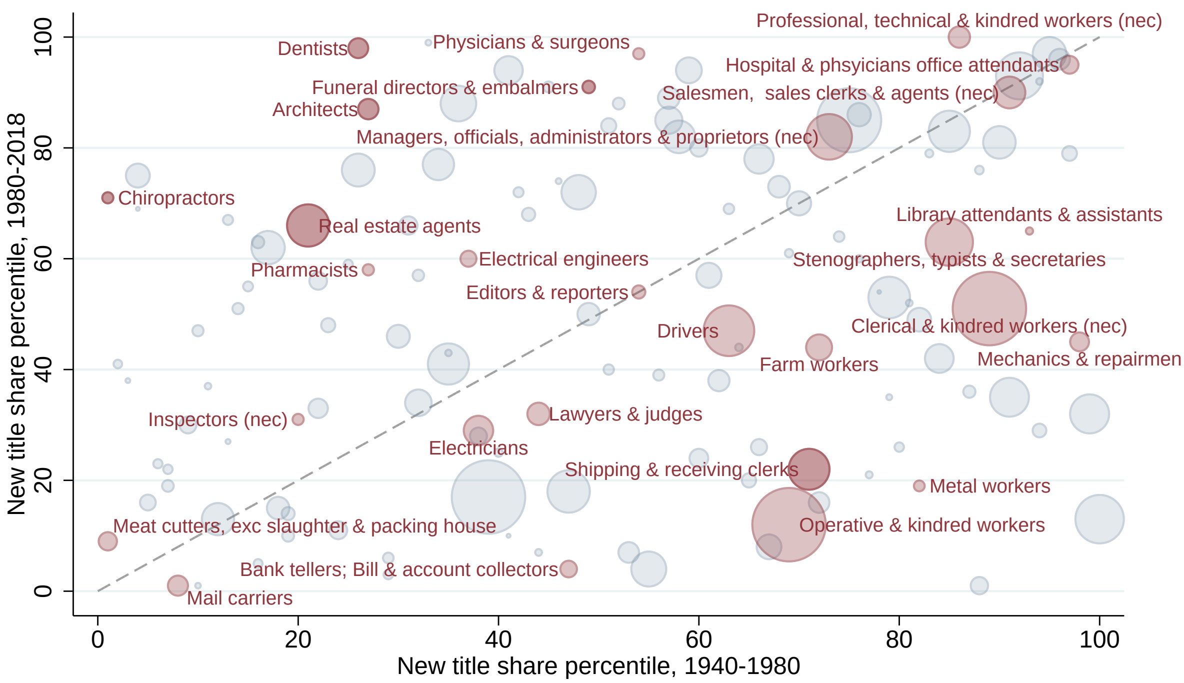
FIGURE A.II

Percentiles of Occupational Exposure to Import Competition from China, by Broad Occupation



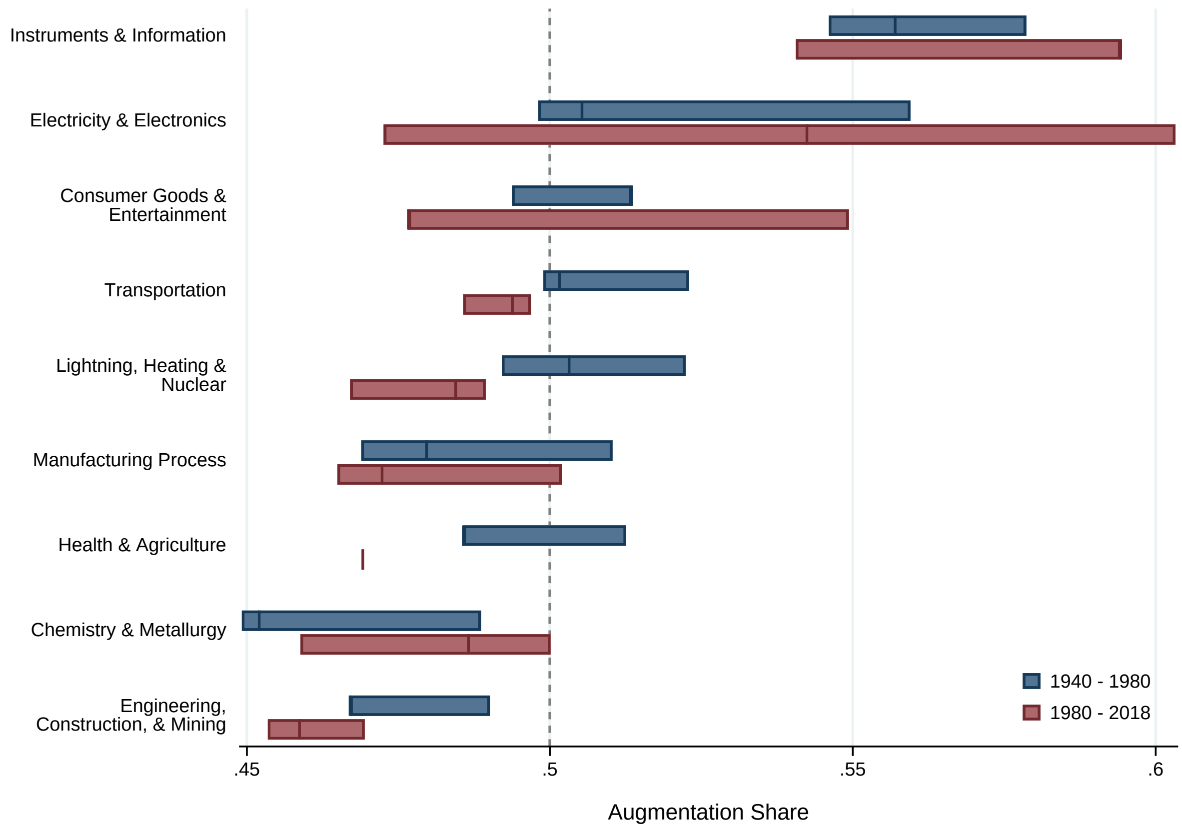
This figure shows percentiles of exposure to import competition averaged over sub-periods 1990–2000 and 2000–2018 for consistent Census occupations classified into twelve broad occupation groups. These groups are ranked vertically from lowest to highest by their mean exposure percentile.

FIGURE A.III
New Title Emergence by Occupation, 1940–1980 vs. 1980–2018



This figure shows percentiles of new title shares (defined as the ratio of the new title count over total title count in a decade), averaged over 1940–1980 (x-axis) and 1980–2018 (y-axis). Each point is a consistently defined Census occupation ($N = 132$), where the size of the circle represents mean occupational employment size over 1940-2018. The employment weighted correlation between 1940-1980 and 1980-2018 new title shares is 0.38. The 45 degree line is plotted with dashes.

FIGURE A.IV
Augmentation Shares by Patent Class, 1940–1980 and 1980–2018



This figure shows a box plot of the share of augmentation patent-occupation matches in each 3-digit CPC technology class over 1940–1980 (blue bars) and 1980–2018 (red bars) relative to its number of automation and augmentation matches. Technology classes are weighted by the citation-scaled number of issued patents in each period. The boxes indicate the interquartile range (25th to 75th percentile), and the vertical bar within each of these boxes indicates the median.

TABLE A.I
Examples of Patents that are Highly Augmenting for Computer Scientists and Highly Automating for Other Occupations

A. Billing Clerks and Related Financial Records Processing
Authentication device
Systems and methods for automated exchange of electronic mail encryption certificates
Authentication system and method thereof for dial-up networking connection via terminal
Enforcing file authorization access
Verification of user communication addresses
Method and apparatus for storing confidential information
Secure end-to-end transport through intermediary nodes
Methods and apparatus for increased security in issuing tokens
Secure end-to-end transport through intermediary nodes
B. Health Record Technologists and Technicians
Method for synchronizing documents for disconnected operation
Document access auditing
Direct connectivity system for healthcare administrative transactions
Method and device for transcoding during an encryption-based access check on a database
Data classification and privacy repository
Monitoring and auditing system
Method and system for validating timestamps'
System and method for user authentication in in-home display
System, method, and article of manufacture for maintaining and accessing a whois database
Methods and systems for consolidating medical information
Synchronization of application documentation across database instances
Method and system for processing a database query by a proxy server
C. Office Machine Operators, Computer and Peripheral Equipment Operators and Other Telecom Operators
Peripheral device for programmable logic controller
Method of authentication user using server and image forming apparatus using the method
System and method for securing data for redirecting and transporting over a wireless network
System and method for securing data
Secret authentication system
Web-enabled mainframe
System and method for authenticating streamed data

Table reports patents issued between 2000 and 2018 that are simultaneously augmentation-linked to *Computer Systems Analysts & Computer Scientists* and automation-linked to *Billing Clerks and Related Financial Records Processing* in panel A, *Health Record Technologists and Technicians* in panel B, and *Office Machine Operators, Computer and Peripheral Equipment Operators* in panel C.

TABLE A.II

New Title Emergence Robustness to Alternative Specifications, 1940–2018

	(1)	(2)	(3)	(4)	(5)
			A. Dep Var: New Title Count (Poisson)		
Augmentation Exposure	59.89***	31.26***		37.44***	32.00***
	(9.91)	(5.16)		(9.20)	(5.71)
Automation Exposure			39.38***	18.85***	−1.90
			(9.81)	(4.48)	(5.52)
			B. Dep Var: Log New Title Count (OLS)		
Augmentation Exposure	20.49***	19.84***		19.96***	18.88***
	(4.91)	(2.98)		(5.47)	(2.91)
Automation Exposure			14.38**	1.22	4.20
			(5.20)	(5.14)	(3.24)
			C. Dep Var: Dummy for New Titles (LPM)		
Augmentation Exposure	0.71*	1.20***		1.12**	1.33***
	(0.36)	(0.30)		(0.41)	(0.33)
Automation Exposure			−0.26	−1.01**	−0.63
			(0.26)	(0.35)	(0.40)
			D. Dep Var: New Title Count (Negative Binomial)		
Augmentation Exposure	12.55**	16.84***		9.46*	16.02***
Start-of-Period CAI	(3.86)	(4.03)		(4.16)	(3.95)
Automation Exposure			12.63**	8.23*	4.67
			(3.98)	(4.16)	(3.84)
			E. Dep Var: New Title Count (Negative Binomial)		
Augmentation Exposure	16.57***	20.03***		15.07***	19.47***
CAI excl. new titles	(3.32)	(3.35)		(3.78)	(3.13)
Automation Exposure			12.79**	3.13	3.23
			(3.93)	(4.46)	(4.01)
Occ Emp Shares	X	X	X	X	X
Time FE	X		X	X	
Broad Occ × Time FE		X			X

N = 1,535 in Panels A and C; N = 1,276 in Panel B; N = 1,496 in Panel D; N = 1,527 in Panel E. Coefficients multiplied by 100. Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE A.III
The Relationship between Changes in Employment and Exposure to Augmentation and Automation within Industry-Occupation Cells,
First Stage Estimates

Dependent Variable: Log Patent Count

	1940–2018		1940–1980		1980–2018	
	Aug	Aut	Aug	Aut	Aug	Aut
Augmentation IV	1.80***	0.27+	1.43***	−0.03	2.03***	0.41***
	(0.11)	(0.15)	(0.17)	(0.34)	(0.11)	(0.08)
Automation IV	0.19	2.03***	0.37	1.92***	0.11	1.88***
	(0.14)	(0.11)	(0.30)	(0.24)	(0.16)	(0.10)
N	33,900	33,900	6,520	6,520	27,380	27,380
F-stat	127.76	202.85	34.80	38.68	160.12	169.77
Sanderson-Windmeijer F-stat	316.38	374.63	96.22	55.43	309.97	305.00
Ind × Time FE	X	X	X	X	X	X

First stage estimates for column 3 in Table VI and for columns 3 and 7 in Table VII. All employment changes are winsorized at the 99th percentile. Standard errors clustered by industry-occupation cell (using Stata command `ivreghdfe`) in parentheses. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE A.IV

The Relationship between Changes in Employment and Adjusted Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, Stacked Long-Difference Regressions, Manufacturing and Non-Manufacturing Sector, First Stage Estimates, 1940–2018

Dependent Variable: Log Patent Count

	Non-Manufacturing		Manufacturing	
	Aug	Aut	Aug	Aut
Augmentation IV	2.04***	0.51*	1.12***	−0.29*
	(0.15)	(0.21)	(0.10)	(0.15)
Automation IV	0.17	2.24***	0.17*	1.42***
	(0.19)	(0.14)	(0.08)	(0.13)
N	21,795	21,795	12,105	12,105
F-stat	90.27	155.38	78.99	58.25
Sanderson-Windmeijer F-stat	226.77	271.48	156.84	110.01
Ind × Time FE	X	X	X	X

First stage estimates for Table A.V. Manufacturing sectors 1980–2018 are classified according to the 1990 Census industrial classification scheme. Manufacturing sectors 1940–1980 are classified according to the 1950 Census industrial classification scheme. All employment and wagebill changes are winsorized at the 99th percentile. Wagebill variables are all hourly employment and all hourly wages. Hourly wages are imputed to cells where employment but not wages are observed. Standard errors clustered by industry-occupation cell (using Stata command `ivreghdfe`) in parentheses. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE A.V

The Relationship between Changes in Employment and Adjusted Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, Stacked Long-Difference Regressions, Manufacturing and Non-Manufacturing Sector, 1940–2018

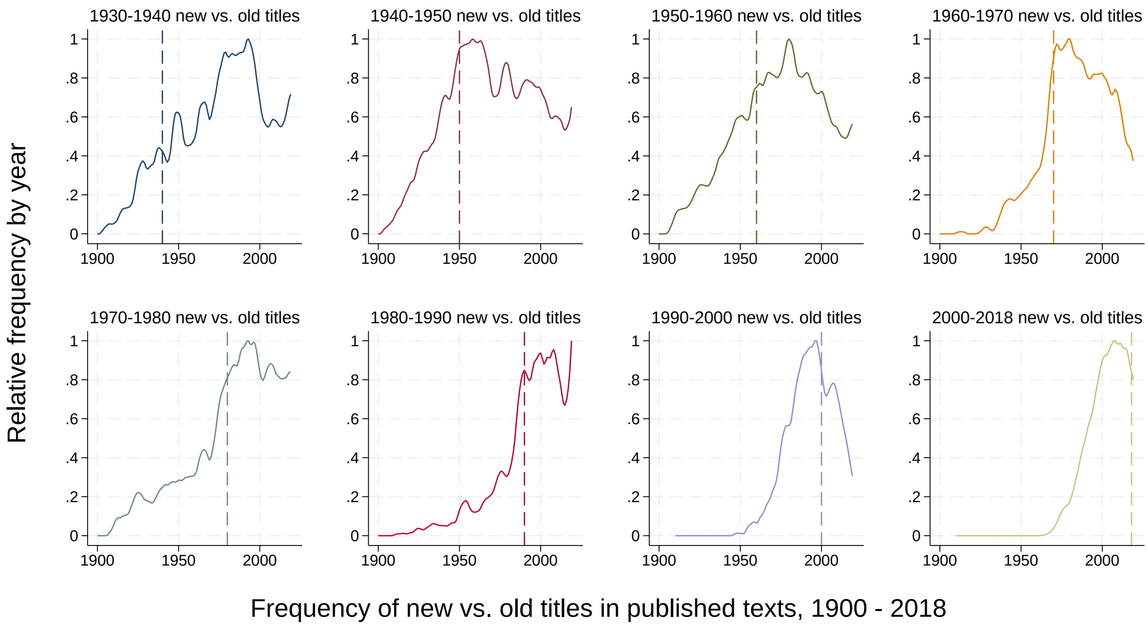
	Decadalized ΔLn(Employment)		Decadalized ΔLn(Adjusted Wagebill)	
	Non-Manuf	Manuf	Non-Manuf	Manuf
	(1)	(2)	(3)	(4)
		A. OLS		
Augmentation Exposure	1.58***	1.16***	1.74***	1.11***
	(0.25)	(0.32)	(0.29)	(0.32)
Automation Exposure	−2.65***	−1.01**	−2.64***	−1.32***
	(0.33)	(0.37)	(0.35)	(0.37)
R^2	0.52	0.55	0.51	0.52
		B. 2SLS		
Augmentation Exposure	4.07***	4.62**	4.93***	4.81**
	(1.11)	(1.77)	(1.21)	(1.76)
Automation Exposure	−3.68***	−6.11***	−3.30***	−6.47***
	(0.70)	(1.10)	(0.74)	(1.11)
F-stat (Aug)	90.27	78.99	90.27	78.99
F-stat (Aut)	155.38	58.25	155.38	58.25
Ind × Time FE	X	X	X	X
N	21,795	12,105	21,795	12,105

Changes in employment and wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Manufacturing sectors 1980–2018 are classified according to the 1990 Census industrial classification scheme. Manufacturing sectors 1940–1980 are classified according to the 1950 Census industrial classification scheme. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill changes are winsorized at the 99th percentile. Wagebill variables are all hourly employment and all hourly wages. Hourly wages are imputed to cells where employment but not wages are observed. 2SLS specifications additionally control for indicators for being above the cross-sectional median in total augmenting or automating patent matches two decades prior. Standard errors clustered by industry-occupation cell (using Stata command `reghdfe` and `ivreghdfe`) in parentheses. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

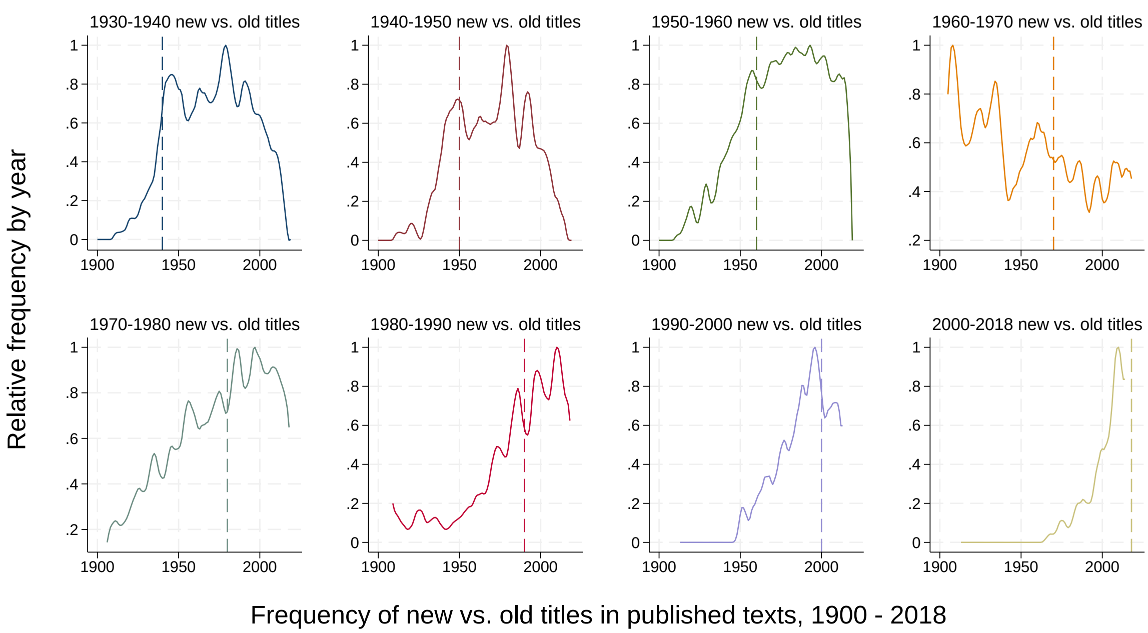
B Supplemental appendix figures and tables

FIGURE A.V
Google Ngram Viewer Occurrence, by Four Major Occupation Groups

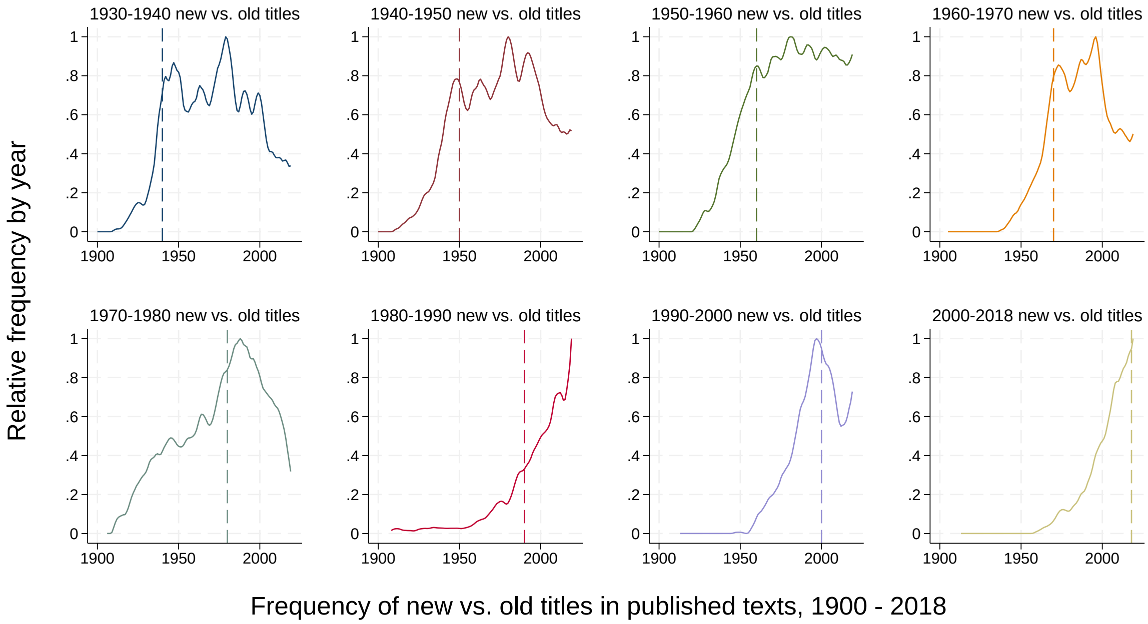
Panel A. Personal service occupations: Health services, Cleaning and protective services, Personal services



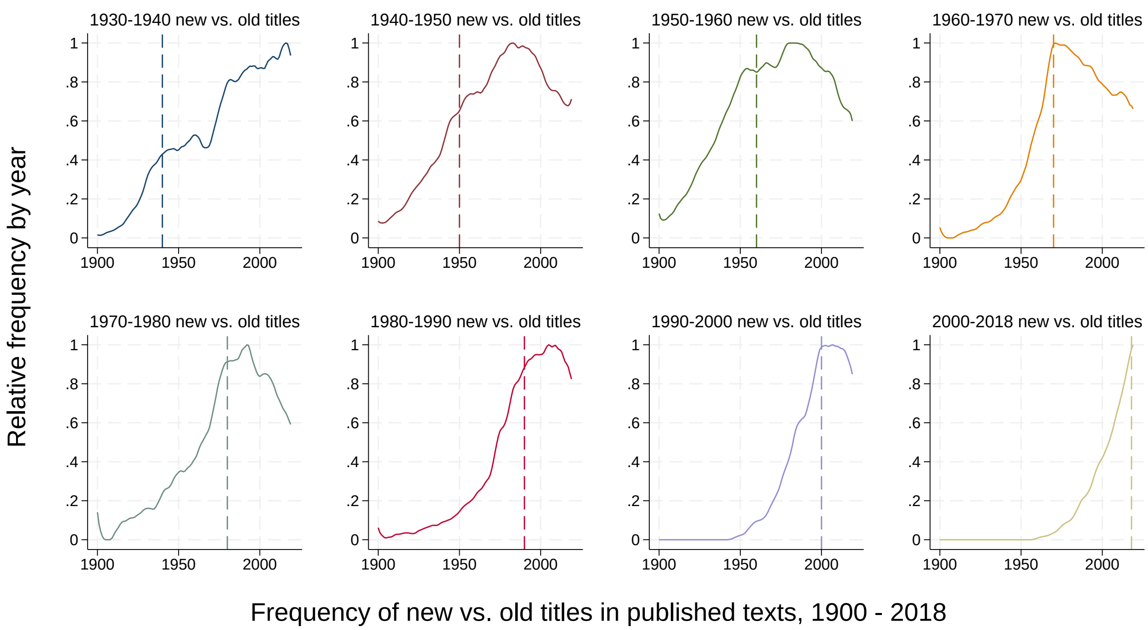
Panel B. Blue collar occupations: Agriculture and mining, Construction and mechanics, Production and operatives



Panel C. Commercial service occupations: Retail sales, Technicians, fire, and police, Transportation



Panel D. Professional and information occupations: Managers and executives, Professionals, Advertising and financial sales, Clerical and administrative support



This figure reports the relative frequency of each decade's cohort of new titles relative to the cohort of existing titles across four broad occupational groups. This figure is constructed using the procedure used for Figure A.I.

FIGURE A.VI
The Occupational Distribution of the Flow of New Work by Education Group, 1940–1980 and 1980–2018

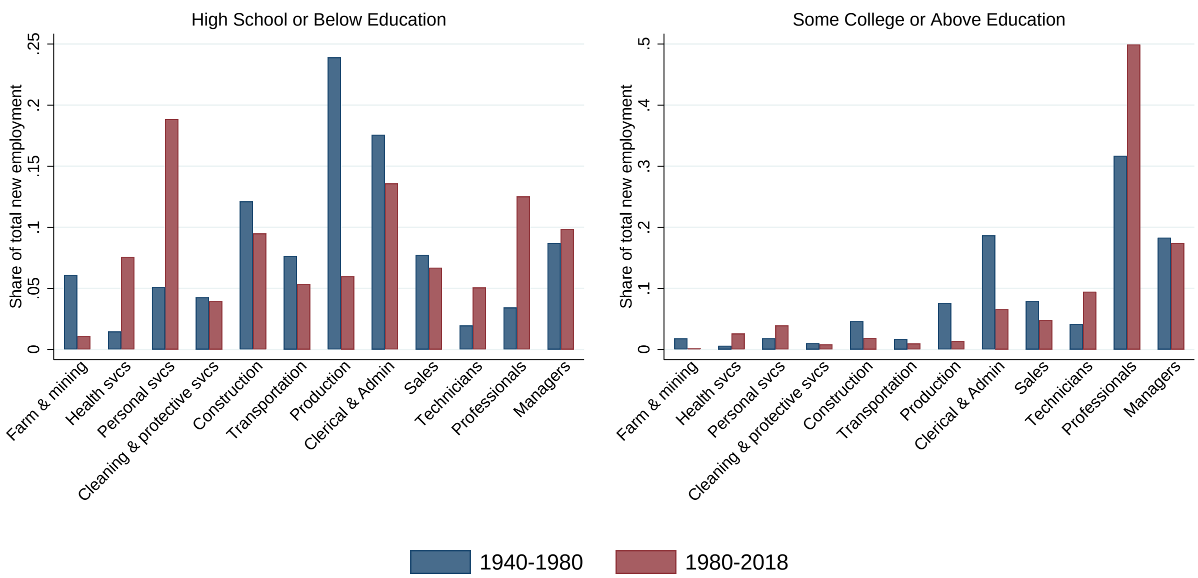
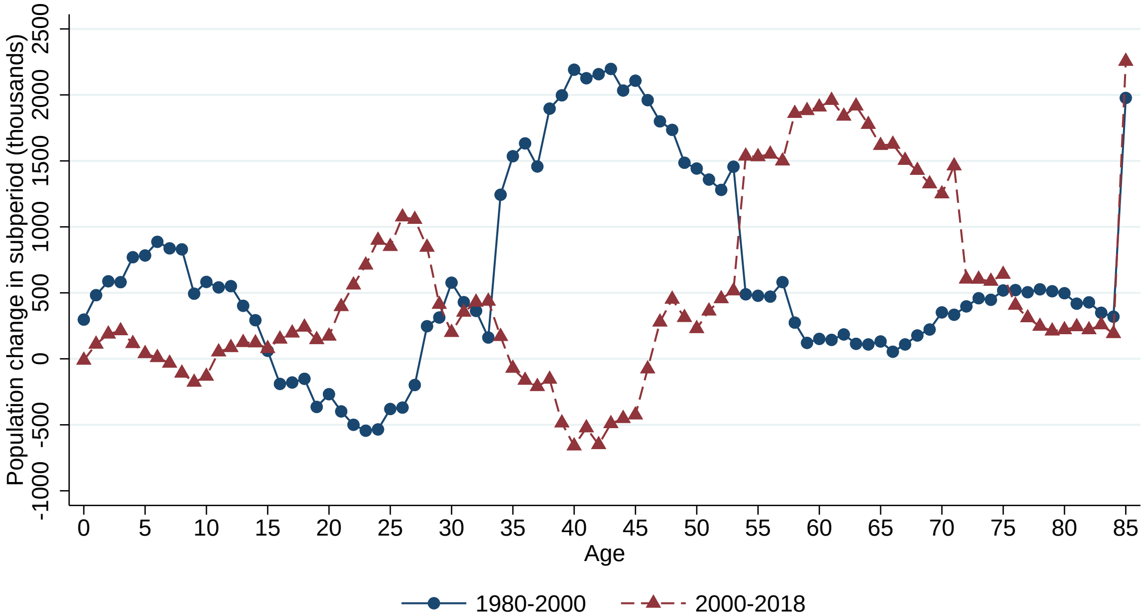


Figure reports the share of employment in new work across broad occupational categories separately for education groups and time periods in twelve exhaustive and mutually exclusive occupational categories ordered from lowest to highest-average earnings. Blue bars represent new work employment shares over 1940–1980, and red bars represent shares over 1980–2018.

FIGURE A.VII
U.S. Population Changes (1,000s) by Single Year of Age, 1980–2000 and 2000–2018



This figure depicts changes in U.S. population counts in 1,000s by single years of age over 1980–2000 (blue line) and 2000–2018 (red line).

TABLE A.VI
Descriptive Statistics for Augmentation and Automation Exposure, 1940–2018, 1940–1980, and 1980–2018

	1940–2018		1940–1980		1980–2018	
	Mean	SD	Mean	SD	Mean	SD
			A. Occupation × Decade			
Augmentation Exposure	9.85	3.54	9.49	3.56	10.32	3.46
Automation Exposure	9.84	2.80	9.18	2.88	10.71	2.45
N	1,535		628		907	
			B. Occupation × Industry × Four-Decade Period			
Augmentation Exposure	11.77	2.84	11.15	2.82	12.38	2.72
Automation Exposure	11.63	2.56	10.59	2.65	12.67	1.98
N	33,900		6,520		27,380	

Augmentation and automation exposure measures correspond to the log of the weighted count of matched patents. All statistics are weighted by start-of-period employment shares.

TABLE A.VII
Descriptive Statistics for Augmentation and Automation Exposure by Sector, 1940–2018

	Manufacturing		Non-Manufacturing	
	Mean	SD	Mean	SD
Augmentation Exposure	13.34	2.07	11.31	2.86
Automation Exposure	13.31	1.99	11.14	2.49
N	12,105		21,795	

Augmentation and automation exposure measures correspond to the log of the weighted count of matched patents at occupation × industry level by four-decade period. All descriptive statistics are weighted by start-of-period employment shares.

TABLE A.VIII

First Stage Estimates for New Titles Regressions, 1940–2018

Dependent Variable: Log Patent Count

	(1)	(2)	(3)	
	Aug	Aut	Aug	Aut
Augmentation IV	2.29***		2.16***	0.24
	(0.33)		(0.39)	(0.25)
Automation IV		2.20***	0.59	2.16***
		(0.23)	(0.47)	(0.25)
F-stat	47.77	91.04	32.91	42.44
Sanderson-Windmeijer F-stat	47.77	91.04	47.11	35.91
Occ Emp Shares	X	X	X	X
Time FE	X	X	X	X

N = 1,535. First stage estimates for columns 1, 3, and 4 in Table A.IX. Specifications additionally control for indicators for being above the cross-sectional median in total augmenting or automating patent matches two decades prior, and whether the occupation has greater than 50% employment in manufacturing industries. Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE A.IX

2SLS Estimates for New Titles Regressions, 1940–2018

Dependent Variable: Dummy for New Titles (LPM)

	(1)	(2)	(3)	(4)	(5)
Augmentation Exposure	0.92	1.99*		1.49	1.91
	(1.22)	(0.97)		(1.44)	(1.44)
Automation Exposure			−1.07	−1.73	−0.16
			(1.04)	(1.41)	(2.67)
F-stat (Aug)	47.77	58.22		32.91	37.34
F-stat (Aut)			91.04	42.44	62.76
Occ Emp Shares	X	X	X	X	X
Time FE	X		X	X	
Broad Occ × Time FE		X			X

N = 1,535. Coefficients multiplied by 100. Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE A.X
Occupational New Title Emergence and Demand Expansions from Demographic Change

Dependent Variable: New Title Percentile

	(1)	(2)	(3)	(4)	(5)
Demand Shift Exposure Percentile	0.252***	0.150*	0.288***	0.214***	0.203***
	(0.074)	(0.068)	(0.074)	(0.060)	(0.060)
Augmentation Exposure Percentile			0.177+	0.360***	0.346***
			(0.092)	(0.059)	(0.061)
R^2	0.19	0.31	0.21	0.38	0.38
Time FE	X	X	X	X	X
Occ Emp Shares	X	X	X	X	X
Ind Exposure Control	X	X	X	X	X
Broad Occ FE		X		X	X
Δ Occ Emp Shares					X

N = 612. Standard errors clustered by occupation × 40-year period in parentheses. Observations are weighted by start-of-period occupational employment shares. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE A.XI

The Relationship between Changes in Employment and Wagebill and Exposure to Augmentation (Using Start-of-Period CAI) and Automation within Industry-Occupation Cells, OLS Stacked Long-Difference Regressions, 1940–2018

	(1)	(2)	(3)	(4)	(5)	(6)
			A. Decadalized ΔLn(Employment)			
Augmentation Exposure	0.41*	0.74***			0.87***	0.87***
	(0.18)	(0.20)			(0.18)	(0.19)
Automation Exposure			−1.76***	−0.49	−2.02***	−0.80*
			(0.27)	(0.39)	(0.26)	(0.38)
R^2	0.51	0.56	0.52	0.56	0.53	0.56
			B. Decadalized ΔLn(Wagebill)			
Augmentation Exposure	0.24	0.63**			0.68***	0.75***
	(0.20)	(0.22)			(0.19)	(0.20)
Automation Exposure			−1.73***	−0.47	−1.93***	−0.74+
			(0.28)	(0.44)	(0.28)	(0.43)
R^2	0.52	0.57	0.52	0.57	0.53	0.57
			C. Decadalized ΔE[Ln(Wagebill)]			
Augmentation Exposure	0.24	0.58**			0.68***	0.70***
	(0.18)	(0.19)			(0.18)	(0.18)
Automation Exposure			−1.68***	−0.54	−1.88***	−0.80*
			(0.27)	(0.38)	(0.27)	(0.38)
R^2	0.53	0.58	0.54	0.58	0.54	0.58
			D. Decadalized ΔLn(Adjusted Wagebill)			
Augmentation Exposure	0.40*	0.79***			0.88***	0.91***
	(0.20)	(0.22)			(0.19)	(0.20)
Automation Exposure			−1.81***	−0.41	−2.08***	−0.74+
			(0.29)	(0.44)	(0.28)	(0.43)
R^2	0.50	0.56	0.51	0.55	0.51	0.56
Ind × Time FE	X	X	X	X	X	X
Broad Occ × Time FE		X		X		X

N = 33,799 changes in employment and wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill changes are winsorized at the 99th percentile. Wagebill variables are all hourly employment and all hourly wages. Hourly wages are imputed to cells where employment but not wages are observed. Standard errors clustered by industry-occupation cell (using Stata command `reghdfe`) in parentheses. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE A.XII
The Relationship between Changes in Employment and Wagebill and Exposure to Augmentation (Using CAI excl. new titles) and Automation within Industry-Occupation Cells, OLS Stacked Long-Difference Regressions, 1940–2018

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Decadalized ΔLn(Employment)					
Augmentation Exposure	0.70***	1.14***			1.33***	1.28***
	(0.19)	(0.19)			(0.19)	(0.19)
Automation Exposure			−1.76***	−0.47	−2.17***	−0.85*
			(0.26)	(0.38)	(0.27)	(0.38)
R^2	0.52	0.56	0.52	0.56	0.53	0.56
	B. Decadalized ΔLn(Wagebill)					
Augmentation Exposure	0.72***	1.20***			1.34***	1.34***
	(0.22)	(0.23)			(0.22)	(0.23)
Automation Exposure			−1.73***	−0.45	−2.14***	−0.85*
			(0.28)	(0.43)	(0.28)	(0.42)
R^2	0.52	0.57	0.52	0.57	0.53	0.57
	C. Decadalized ΔE[Ln(Wagebill)]					
Augmentation Exposure	0.64***	1.10***			1.24***	1.25***
	(0.19)	(0.19)			(0.19)	(0.20)
Automation Exposure			−1.68***	−0.53	−2.06***	−0.90*
			(0.27)	(0.38)	(0.27)	(0.38)
R^2	0.53	0.58	0.54	0.58	0.54	0.58
	D. Decadalized ΔLn(Adjusted Wagebill)					
Augmentation Exposure	0.77***	1.24***			1.43***	1.37***
	(0.22)	(0.23)			(0.22)	(0.22)
Automation Exposure			−1.81***	−0.39	−2.25***	−0.80+
			(0.28)	(0.44)	(0.28)	(0.42)
R^2	0.50	0.56	0.51	0.55	0.52	0.56
Ind × Time FE	X	X	X	X	X	X
Broad Occ × Time FE		X		X		X

N = 33,894 changes in employment and wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill changes are winsorized at the 99th percentile. Wagebill variables are all hourly employment and all hourly wages. Hourly wages are imputed to cells where employment but not wages are observed. Standard errors clustered by industry-occupation cell (using Stata command `reghdfe`) in parentheses. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

C Constructing consistent occupations and industries

Our analyses require occupation-by-industry panels. The specificity of Census 3-digit occupation and industry categories generally rises from decade to decade, with approximately 250 3-digit occupations in 1940 and roughly 500 in 2018. Simultaneously, occupation and industry categories merge, split, and recombine. This makes it infeasible to construct a fully balanced panel of detailed occupations and industries over the eight decades of 1940–2018 without sacrificing substantial resolution. Instead, we construct two separate panels to cover the 1940–2018 period: one set of consistent occupations and industries over 1940–1980, and another over 1980–2018.

The balanced panel for 1940–1980 is based on the IPUMS's harmonization of the Census Bureau's occupation and industry coding scheme for 1950 (Ruggles, Flood, et al., 2022), which we further refine to yield a set of 166 consistent, 3-digit occupations and 116 consistent, 3-digit industries (`occ4080rj` and `ind4080rj`). The balanced panel for 1980–2018 is based on the consistent occupation and industry coding scheme (`occ1990dd` and `ind1990dd`) developed by Dorn (2009), which refines the IPUMS consistent industry/occupation 1990 scheme. We update the Dorn (2009) series for consistency through 2018, yielding a panel of 306 consistent, 3-digit occupations and 206 consistent, 3-digit industries (`occ1990dd_18` and `ind1990dd_18`).

We additionally construct thirteen broad industries and twelve broad occupations which can be consistently defined over the entire 1940–2018 period. The broad industry groups are 1. Manufacturing; 2. Agriculture; 3. Mining; 4. Construction; 5. Transportation; 6. Wholesale; 7. Retail; 8. Finance, Insurance, and Real Estate; 9. Business and Repair Services; 10. Personal Services; 11. Entertainment and Recreation Services; 12. Professional and Related Services; and 13. Public Sector. The broad occupation groups are: 1. Managers and Executives; 2. Professionals (including Financial Advertising and Sales); 3. Technicians, Fire, and Police; 4. Sales (excluding Financial Advertising and Sales); 5. Clerical and Administrative Support; 6. Production and Operatives; 7. Transportation; 8. Construction and Mechanics; 9. Cleaning and protective services; 10. Personal services; 11. Health services; and 12. Agriculture and mining occupations.

In Figure A.V we further aggregate the twelve broad occupations into four groups. Blue collar occupations contain Agriculture and Mining, Construction and Mechanics, and Production and Operatives. Professional and information occupations contain Managers and Executives, Professionals (including Financial Advertising and Sales), and Clerical and Administrative Support. Personal service occupations contain Health Services, Cleaning and Protective Services, and Other Personal Services. Finally, Commercial service occupations contain Retail Sales (minus Financial and Advertising Sales), Technicians, Fire, and Police, and Transportation.

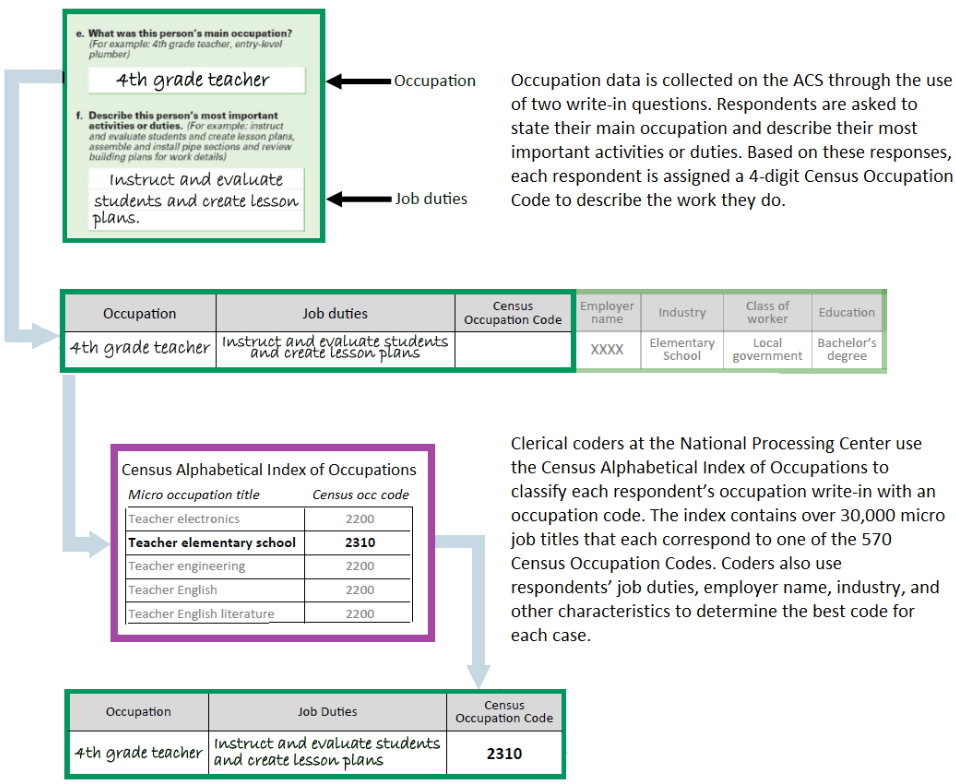
Lastly, for Figure A.III we employ a set of 132 consistent occupations over the 1940–2018 period (`occ4018bw`), obtained by further aggregating all overlapping categories among our consistent occupations for the 1940–1980 and 1980–2018 subperiods and then unwinding some of the overly aggregated categories, though this comes at the cost of some measurement error. We do not use this for our baseline results because it substantially reduces occupational variation—central to all our analyses—and because the 1950 and 1990 IPUMS occupation schemes are profoundly different, leading to substantial information loss when crosswalking.

D Measuring new work

We detail here how we identify new occupation titles and how total employment in new work is constructed.

Figure A.VIII summarizes the Census Bureau's procedure for classifying American Community Survey respondents' free-text occupational write-ins to Census occupation codes. Census clerical coders use the CAI to classify respondent write-ins, while simultaneously flagging new and emerging write-in titles. Census managers review these candidate titles for potential inclusion in subsequent CAI editions.

FIGURE A.VIII
Coding Process of Occupation Write-ins in the ACS



1 Procedure for identifying new occupation titles

To extract new work added to the Census Alphabetical Index of Occupations (CAIO) between Census or ACS years $t-1$ and t we use the following steps:

1. Clean titles in both $t-1$ and t by removing capitalization, punctuation, as well as certain common synonyms and decade-specific format changes we identify from inspection of CAIO volumes. This avoids unnecessarily flagging titles as potentially new ("candidate-new") if they are old titles that have been reformatted or reworded in minor or predictable ways.

   (a) Examples of format and wording changes that we discard are titles like "Accounting Work, Accountant" and "Ad Writer" being added in t when "Accountant" and "Advertising Writer" already exist in $t-1$.

(b) We also unify variations of titles which contain the same terms either in full or abbreviated form, such as "db" for database, "pt" for physical therapy, "pv" for photovoltaics, and "qc" for quality control.

(c) Prior to matching, we reduce -man, -person, -work, -er, -or, -ing, -ist etc. titles to the same word base, e.g. "Salesperson", "Salesman" and "Sales work" are changed to "sales"; "Adviser", "Advisor" and "Advising" are degenerated to "advis", and 'Motorist" is degenerated to "motor".

(d) We clean plural forms, including those ending in "-s" or "-es", and other specific plural forms such as '-ies" when it is a plural of "-y".

(e) We also discard new gender-specific or gender-neutral versions of existing titles, e.g. we treat the titles "Actor" and "Actress" as one and the same; as we do "Waiter", "Waitress", and "Waitstaff"; and we discard "Chipper Operator" as new because it replaced "Chipperman",

(f) We discard word order duplicates that are classified to the same Census occupation (e.g. out of "Television Station Manager", and "Manager, Television Station", we retain only one): these occur because at the time of its conception the alphabetical index was used in printed form– multiple word orders were included to save coders time in looking up entries. We retain any title duplicates classified to different industries or occupations, as this may reflect (increasing) emergence of a type of job (an example is the prevalence of IT-related titles across many industries).

(g) Examples of words we automatically denote as synonyms are "auto" and "automobile"; "equipment operator" and "operator"; "sales", "selling", and "sales representative"; "garbage" and "rubbish"; "aide" and "assistant"; "gage" and "gauge".

2. Exact-match and fuzzy-match of cleaned occupation titles between CAIO_t to CAIO_{t-1}. We drop all exact title duplicates between $t-1$ and t, disregarding any spacing differences in titles. For the remainder, we retain the three most similar $t-1$ title matches for each t title. Specifically:

   (a) For the exact match, we simply match the cleaned titles in t to $t-1$, discard exact matches, and retain the set of unmatched CAIO_t titles as "candidate-new" titles.

   (b) Next, we fuzzy match the CAIO_t candidate-new titles to all CAIO_{t-1} cleaned titles. We use a fuzzy-matching Jaro-Winkler algorithm which matches based on letter-swaps, implemented in R as the package *stringdist* (Loo, 2014). This assigns high similarity (i.e. low distance) scores to titles where a low number of single-character transpositions are required to change one word into the other.[29] It also gives higher similarity to strings matching from the beginning up to some specified length: we set the constant scaling factor determining this at the standard value of 0.1. For example, titles which are identical except for a hand-keying error ("Mechanotheraplst" and "Mechanotherapist") receive a high similarity score.

3. Adjudicate remaining unmatched $t+1$ titles ("candidate-new" titles) by classifying them as new or not new, using a combination of automated assignment and careful manual revision. The large majority of candidate-new titles are manually revised, with only around 1,034

[29]Note that we have already discarded word order duplicate titles prior to implementing this algorithm.

automatically assigned. We observe 273,960 total titles over 1940–2018, of which we identify 28,315 as new over the whole period.

In adjudicating candidate-new titles, our overarching goal is to identify titles that either capture a type of job that was previously nonexistent or reflect further differentiation or specialization of existing work. These latter cases are much more common than are entirely new categories, and can arise from new or specialized work domains, specialization in educational or professional requirements, or the use of specialized work methods (e.g. by hand or using a machine). On the other hand, candidate-new titles are discarded (i.e. marked as not new) when they reflect a renaming, reformatting, or generalization of previously existing work. This (time-consuming) manual revision requires looking beyond fuzzy-match results to search the entire $t-1$ index for comparable work.

We implement these principles with the following specific rules for classifying a candidate-new title as new or not new. While not exhaustive, these rules capture commonly occurring cases. A t candidate-new title is:

1. *New* when it is a differentiation of a t title, e.g. "Clinical Psychologist" is new in 1950 as a differentiation of "Psychologist", and "Assembler, Electrical Controls" is new in 1990 as a differentiation of "Assembler, n.s.". This is by far the most commonly occurring type of new title.

2. *New* when it adds specialized work tools to a $t-1$ title, most commonly 'hand' or 'machine'; or specializes operators and set-up operators. E.g. "Bookkeeping Clerk, Machine" is new in 1970 because before only "Bookkeeping Clerk" was listed; and "Drill-Press Set-Up Operator" is new when it is added to "Drill-Press Operator".

3. *New* when it adds some additional educational or professional differentiation to a $t-1$ title. E.g. "Licensed Addiction Counselor" is new in 2018 as an addition to "Addiction Counselor"; and "Health Therapist, Less Than Associate Degree" is new in 1990 as an addition to "Health Therapist". This is a type of new title that occurs relatively infrequently.

4. *New* when it adds "not specified" or "not elsewhere classified" to a $t-1$ title. This reflects more types of this title are emerging which (for the time being) are listed as n.s. / n.e.c. For example, "Mechanic, Instrument, n. s." is added in 1980.

5. *New* when it bifurcates a $t-1$ title into two separate types, usually marked with "incl", "exc", or "any other". E.g. in 1980 the title "Sitter, exc. Child Care" was new since before only "Sitter" had existed.

6. *Not new* when it simply reorganizes information across various columns of the index for the same title. E.g. "Apprentice Dentist" was discarded as new in 1940 because it already existed in 1930. This is a common reason for discarding candidate-new titles.

7. *Not new* when it is generalization from previously-specified title, e.g. "Ad Taker" is not new in 1980 because it simply subsumes the 1970 titles "Classified-Ad Taker" and "Telephone-Ad Taker"; and "Inspector Agricultural commodities" is not new in 1980 because it subsumes "Inspector Fruit", "Inspector Food", and "Inspector Livestock".

8. *Not new* when it is the same as a $t-1$ title except for filler words; or an (un)abbreviated version. E.g. "Software Applications Developer" is not new in 2018 because the title "Software Developer" already existed before; and "RRT" is not new in 2018 because "Registered respiratory therapist" already existed before.

9. *Not new* when a title is a combination of two existing titles. E.g. "Inker and opaquer" is not new in 1980 because both "Inker" and "Opaquer" already existed in 1970.

Overall, we discard 15.56% of all candidate-new titles we manually revise ($N = 30,332$). This manual revision is necessary since many newly added titles reflect synonyms or rewordings of existing titles, which would not reliably be captured by an algorithm. Further examples of candidate-new titles we discard are "Nanny" (as the synonyms "Governess", and "Baby sitter" already existed); "Server, food" (as the synonym "Waiter" already existed), "Glass worker" (as various more specific glass workers already existed, including "Glass bender", "Glass blower", "Glass cleaner", "Glass cutter"); "Flight instructor" (as the synonym "Teacher, flying" already existed); "Ad setter" (as the synonym "Advertising-layout man" already existed); "Pilot, aircraft" (as the synonym "Aviator" already existed); "Supervisor of police" (as the synonym "Police superintendent" already existed); "Key attendant" (as "Key boy" and "Key girl" already existed and the Census only degendered the title); "Mental health aide" (as "Psychiatric aide" already existed); "Dog groomer" (as the more specific "Dog barber" and "Dog beautician" already existed); "Private eye" (as the synonym "Private detective" already existed); "General-office worker" (as the synonym "Office Employee" already existed); "Bingo usher" (as the synonym "Bingo attendant" already existed); "PT" (as the unabbreviated "Physical therapist" already existed); "O.T." (as the unabbreviated "Occupational therapist" already existed); and "NP" (as the unabbreviated "Nurse practitioner" already existed).

2 Examples of new work

Here, we provide several illustrative examples of the emergence of new work at the occupation level. Measuring the emergence of new work requires a consistent set of occupations over the full 1940–2018 interval, which we have constructed at the cost of some information loss ($N = 132$ occupations), as described in Appendix C. (Note that our main analyses do not use this 80-year consistent series because of the loss of occupational granularity.)

As a first example, consider the occupation *Office machine operators*, where 178 new titles have been added over 1940–2018. In the first decades of our sample, examples of such titles are Recordak operator, Letter opener operator, Check writing machine operator (all added in 1940); Telautograph operator, Microfilm cameraman, Multilith operator, I.B.M. operator (all added in 1950); and Xerox-machine operator, Computer operator, Univac operator, Check-sorter operator (all added in 1960). This new work reflects various innovations in office machinery, including salient examples related to the Recordak system developed by the Eastman Kodak Company in 1928, which photographed documents on 16- or 35-millimeter film[30]; the development of I.B.M. and Xerox office machines; and Universal Automatic Computer (Univac) machines used by, among others, the U.S. Census itself.[31] In later decades, new titles in this occupation become increasingly related to digital technology, including Digital computer operator, Machine bookkeeper (both added in 1970), Cryptographic machine operator, Photocopying machine operator, Wires transfer clerk, (all added in 1980), Computer operator, terminal; Camera operator, microfilm or microfiche; Computer typesetter; Facsimile machine operator (all added in 1990). In the same time period there are also new titles related to data entry, such as Data coder operator, Data processing, data entry or keying, and Data processing, any other specified or not specified (all added in 1980). New work emergence has dwindled in the most recent decades, with only a small number of examples such as

[30] https://www.britannica.com/technology/microform#ref237311

[31] https://en.wikipedia.org/wiki/UNIVAC

Photocomposing keyboard operator and Peripheral electronic data processing equipment operator (both added in 2000); and Flat sorting machine clerk and Mail forwarding system markup clerk (both added in 2018). Specific breakthrough innovations that our method links to this occupation over our period of observation are patents for 'Photocomposing machine', 'Automatic computer', 'Data processing system', 'Electronic check writer', ' Computer terminal', ' Digital typesetter', and ' Electronic mail forwarding system and method'.

A related occupation is *Stenographers, typists and secretaries*, where we identify 85 new titles. As with with Office machine operators, in the first decades many new titles in this occupation reflect the diffusion of various types of new office technology, e.g. Operator dictaphone or dictagraph; Telegraphic typewriter operator, Automatic telegraph operator (all added in 1940), Varitype operator (added in 1950), Robotype operator, Planograph operator (both added in 1960), and Magnetic tape selectric typewriter operator (added in 1970) as well as associated changes in workplace organization, e.g. Typing-pool supervisor (added in 1950). This was accompanied by new titles in specialized typing and transcribing activities, including Court transcriber, Statistical typist, Hearing stenographer, Escrow secretary, Psychiatric secretary, and Medical typist (all added in 1960) and Policy writer, typing (added in 1970). In 1980, the new title of Word processor is added, reflecting the introduction of office computing. In subsequent decades, fewer new titles reflect operating office machines, but instead indicate a role switch from typing to other administrative activities, as evidenced by the additions of Executive secretary, exc. managerial or official duties (added in 1980), Private secretary (added in 2000), Medical scheduler, and Legal administrative assistant (both added in 2018). Breakthrough innovations linked to this occupation include patents for 'Telegraph system', 'Dictation system', 'Typesetting system and apparatus', 'Transcriber system for the automatic generation and editing of text from shorthand machine outlines', 'Computer-assisted documentation system for enhancing or replacing the process of dictating and transcribing', 'Remote medical system', and 'Network based legal services system'.

The professional occupation of *Dentists* has added 25 new titles since 1940. These reflect new practice areas, clientele, educational requirements, or entirely new types of dental services. For example, new titles include Prosthodontist (1940), Endodontist (1970), Maxillofacial surgeon (2000), Dental surgeon bachelors degree or higher (2018), Pediatric orthodontist (2018) and Dentofacial orthopedics dentist (2018), which are all types of dentists distinguished by their specific field of expertise or their credentials. These specialized dentistry titles reflect a deepening of expertise, possibly partially in response to expanding technological knowledge and associated diagnostic and treatment modalities. This is also evident from new titles arguably reflecting novel dentistry services include Radiodontist (1950) and Invisible braces orthodontist (2018). Breakthrough patents linked to this occupation include 'X-ray-contrast medium with a density of 0.8-1.1', 'Orthodontic supporting structure', 'Ring-type implant for artificial teeth', 'Method of treating periodontal disease', 'Surgical implant for oral and maxillofacial implantology', and 'Artificial articular eminence for the mandibular joint'.

Since 1940, the large blue-collar occupation of *Automobile mechanics* has gained 258 new titles. Many added titles capture jobs servicing new automotive technologies: Carburetor man and Speedometer repairman (1940), Dynamometer tester and Wheel alignment mechanic (1950), Hydramatic (an early automatic transmission[32]) specialist (1960); Power brake rebuilder and Auto air conditioning mechanic in 1970; Fuel injection servicer in 1980; Antenna specialist in 2000; and Hybrid car mechanic and GPS car navigation installer in 2018. As these titles underscore, the

[32] https://en.wikipedia.org/wiki/Hydramatic

demand for new expertise is often instantiated by the emerging requirements for mastery of new technologies, tools, and features. These align with several breakthrough technologies linked to this occupation, including a patent for 'Carburetor', various patents for air conditioning systems, a patent for 'GPS car navigation system', and a patent for 'Hybrid energy storage device'.

3 Using new title shares as a measure of employment in new work

To construct total employment in new work by broad occupation in 2018 shown in Figure I, we sum the number of new titles added over 1940–2018, nr_{new}, and divide this by the total number of titles in the 2018 index adjusted for titles that were removed $\tilde{nr}_{2018}$, separately by broad occupation J. The adjustment in the total title count consists of adding in the implied total number of removed titles nr_{dead}, if this number is positive. That is, the cumulative new title share over 1940–2018 is $\frac{nr_{new}}{\tilde{nr}_{2018}}$ where $\tilde{nr}_{2018} \equiv nr_{2018} + nr_{dead} \equiv nr_{2018} + nr_{new} - (nr_{2018} - nr_{1940})$.

For Figure II (and Figure A.VI) we calculate the occupational employment of each education group across all Census macro-occupations (approximately 300) in each decade, and allocate employment within each macro-occupation into new and preexisting work in proportion to the share of titles in that occupation that are newly emergent in that decade, then convert these employment counts into employment shares across 12 broad, consistently defined Census occupations. Differencing these occupational distributions of flows versus stocks by education group in two time periods, 1940–1980 and 1980–2018, gives rise to Figure II. We focus on the *distribution* of new work added by decade rather than the absolute numbers of new titles added, because the latter also depends on available resources at the U.S. Census Bureau for revising the index. By focusing on the occupational distribution of new titles added—representing the flow of new titles between decade t and t_{-1}—we require only that efforts to keep the index representative within a decade are not biased towards any particular set of occupations.

E Identifying augmentation and automation patents

1 Linking patents to occupations and industries

Augmentation innovations

To link patents to the CAI text corpus to capture candidate augmentation innovations, we create a numerical representation of the textual content of each patent and the set of CAI titles falling under a Census occupation and then use these representations to measure textual similarity. We follow Kogan et al. (2021) in representing documents as weighted averages of word embeddings.[33] Word embeddings (Mikolov et al., 2013) are vector representations of individual words, where highly related words have high cosine similarities between their word embeddings. To turn each word into its vector representation we use the pre-estimated set of word embeddings from Pennington, Socher, and Manning (2014).

We first clean and transform each document to be consistent across datasets. We remove common "stop words" (e.g. "is", "the", "above", etc.) with little informative content, retain all nouns and verbs, and lemmatize each word by converting verbs to their present tense and nouns to their singular form. We then extract the word embeddings for each term in the cleaned document and average across them, leaving us with a vector representation of the document's meaning. We use TF-IDF scores to weight the averages.[34] We call the resulting TF-IDF weighted average of word embeddings a "document vector", which we calculate for all CAI occupation descriptions. (Our results are robust to excluding any new titles from the CAI documents prior to patent matching.)

Using these document vectors, we compute the matrix of cosine similarity scores for all patent-occupation pairs within a decadal cohort, where a cohort is a given Census year, the set of titles in the corresponding CAI volume from that year, and the set of patents issued in the preceding decade. To account for the fact that some types of patents have naturally low similarity scores (e.g. those using highly technical terminology such as chemical patents), we normalize these scores by subtracting the median score across occupations for a given patent. We then retain the top 15 percent highest adjusted textual similarity scores across patent p × occupation j pairs according to:

$$I_{p,j} = 1 \text{ if } X_{p,j} \geq \sigma_t, \text{ and zero otherwise,}$$

where $X_{p,j}$ is cosine similarity between patent p and occupation j, and σ_t is the 85th percentile of the similarity score distribution for period t. A period t corresponds to a Census year and also the set of patent issue years we consider for that Census year. Typically this will be the previous 10

[33] A "document" is either the full text of a particular patent or set of micro-titles falling under a Census occupation or industry for a given Census year. A common approach for comparing textual similarity is to represent documents as vectors that count the number of times a given word shows up in the document; textual similarities are then computed by taking the cosine similarity of these vector representations, relying on exact overlap in terms (what is known as the 'bag of words' approach). As discussed in Kogan et al. (2021), the bag of words method for determining document similarity neglects synonyms and is likely to perform poorly in comparing sets of documents that have disparate vocabularies, as is the case when comparing patent texts with lists of CAI titles or DOT task descriptions. Our word embeddings approach overcomes the synonym-blindness problem.

[34] TF-IDF weighting is often used in text analysis to down-weight terms that occur frequently across documents and up-weight terms that occur frequently within a document. The TF-IDF weight of term q in document k is given by $w_{q,k} \equiv TF_{q,k} \times IDF_q$, where $TF_{q,k}$ is the number of times term q occurs in document k divided by the total number of terms in document k, and $IDF_q = \log\left(\frac{\text{N documents in sample}}{\text{N documents that include term } q}\right)$. We compute TF-IDF weights separately for patent documents, CAI titles, and DOT task descriptions.

issue years (so for the 1940 Census t will consist of patents issued over 1930-1939).[35] We find that this method does well in generating substantively appropriate matches between occupations and patents; Table A.XIII provides examples of patents linked to Census occupations.

To aggregate individual patent-occupation matched pairs to an occupation-level (or occupation-industry level, where appropriate) measure of technological exposure, we take the citation-weighted sum over patents issued in period t to obtain patent counts by occupation over time:

$$\text{Npatents}_{jt} = \sum_{p \in \Gamma(t)} \omega_p \times I_{p,j} \quad \text{with} \quad \omega_p \equiv \frac{\text{Ncites}_p}{\text{AvgNcites}_{y(p)}},$$

where $\Gamma(t)$ denotes the set of all patents issued in period t and $y(p)$ denotes the issue year cohort of patent p. Thus the sum of citation weights ω_p is normalized to one in each issue year cohort. Our results are qualitatively identical when simply taking the sum of $I_{p,j}$ without weighting by citations. We have also experimented with using different thresholds for σ_t, including top 20%, top 10%, top 5%, and top 1%. Our results are similar in all cases, though signal strength weakens somewhat when applying thresholds smaller than 5%.

When studying occupation-by-industry cell-level outcomes such as employment and wagebill, we use patents linked to both occupation and industry cells: we refer to these linkages as 'industry-occupation-linked' patents. An occupation-by-industry cell, (j,i), is linked to a patent p if the average of the adjusted occupation-patent similarity score $(X_{p,j})$ and industry-patent similarity score $(X_{p,i})$ is among the top 15 percent highest adjusted textual similarity scores across all patent $\times$ occupation $\times$ industry cells within a decadal cohort. The occupation-level document is constructed using lists of occupation titles from the CAI, as described above. The industry-level document is constructed similarly using lists of industry titles from the CAI.[36]

Automation innovations

The *automation* exposure measure identifies technologies that may automate existing labor-using job tasks. Construction of this measure is identical to above but we replace CAI micro-titles with occupational task descriptions from the DOT, using the 1939 DOT volume for 1940–1980 patent-occupation matches and the 1977 DOT volume for 1980–2018 patent-occupation matches.[37] Our procedure closely follows Kogan et al. (2019) and Webb (2020), who use the textual similarity between occupational task content and patent texts to measure the ability of new technologies to perform occupational tasks. Kogan et al. (2019) study the universe of such patents over a far longer (150 year) time period, linking them to occupational task descriptions in the 1991 DOT volume. We stress that the procedures used here for measuring augmentation and automation innovations and their links to occupations are constructed using fully parallel procedures. As shown in sections IV and V, they have substantively distinct predictive content for new work emergence and occupational demand shifts.

[35] When analyzing time-consistent occupation definitions in the post-1980 period, we skip Census year 2010 (focusing on the long change between 2000 and 2018); therefore in this case t corresponds to patent issue years 2000-2017 and the 2018 Census/ACS year.

[36] Due to the large number of industry $\times$ occupation $\times$ patent cells (the 2000-2018 period has over 100 billion cells), we use a 5% sample of patents to approximate the 85th percentile threshold of the distribution of patent *times* occupation-industry similarity scores. The 85th percentile threshold is calculated using only industry-occupation pairs with non-zero employment counts.

[37] Unlike the CAI, the DOT only has occupation-level textual information. Consequently, the automation exposure measures is always defined at the occupation level only.

2 Examples of the textual content of patent-occupation linkages

Patents are classified as augmentation or automation innovations based on their match scores with textual descriptions of occupational outputs (augmentation) and inputs (automation). This section illustrates the textual content that gives rise to these classifications. As a concrete example, Figure A.X summarizes the textual information contained in the top-100 (i.e., highest match score) augmentation patents that are not automation matched and top-100 automation patents that are not augmentation matched to the Computer Systems Analysts & Computer Scientists occupation over the years 1980–1989. The word clouds in the figure are arranged in four quadrants, with columns containing terms drawn from occupational descriptions (left) and from matched patents (right); and with rows containing terms associated with augmentation (top) and with automation (bottom). To illuminate the occupational data that are used for these textual matches, Figure A.XI reports the word corpora employed for forming patent linkages for the Computer Systems Analysts & Computer Scientists occupation, with the left-hand panel containing text describing their occupational outputs (from the CAI) and the right-hand panel containing text describing their task inputs (from the DOT).

As shown in the top left quadrant of Figure A.X, the occupational output terms for the Computer Systems Analysts & Computer Scientists occupation with the highest TF-IDF weighted cosine similarity to matched augmentation patents include *"system"*, *"computer"*, *"data"*, and *"analyst"*, *"software"*, *"process"*, and *"management"*. The word cloud in the bottom left quadrant reports occupational input terms for this occupation that have the highest TF-IDF weighted cosine similarity to matched automation patents. These include *"input"*, *"program"*, *"computer"*, *"data"*, and *"system"*. Unsurprisingly, there is overlap in these terms: *"computer"* and *"system"* receive high similarity scores in both augmentation and automation matches. However, while the augmentation-associated and automation-associated terms have considerable semantic overlap, the augmentation terms tend to reflect complex roles and services (e.g., management, coordinator, or specialist) while the automation terms tend to encompass concrete inputs (e.g., program, computer).

The two right-hand word clouds of Figure A.X summarize the patent (rather than occupation) terms that most closely link to the Computer Systems Analysts & Computer Scientists occupation, with top terms from augmentation patents on the top right and top terms from automation patents on the bottom right. The patent terms from matched augmentation patents with the highest cosine similarity to the occupation corpus include *"customer"*, *"telephone"*, *"data"*, and *"credit"*. The corresponding top linked terms from automation patents include *"input"*, *"output"*, *"circuit"*, *"voltage"*, and *"amplifier"*. As was the case with the top linked occupation terms, the top linked augmentation and automation terms from patents have substantial qualitative differences, with augmentation terms capturing complex services and automation terms capturing concrete tasks. In comparing the occupation-linked terms in the left-hand column with the patent-linked terms in the right-hand column, it bears note that one should *not* expect the top terms in occupational descriptions to precisely match the top terms in patent descriptions. Rather, the linkages among these terms derive from their similarity in word embedding space.

Figure A.XII illustrates the context in which these patent terms are used by reporting sentences from two top-100 patents linked to Computer Systems Analysts & Computer Scientists. The augmentation-linked patent is titled *US4310727A Method of processing special service telephone calls*, and includes terms such as *"services"*, *"telephone"*, *"customer"* and *"services"*. The automation-linked patent is *US4847683A Diagonal correction in composite video decoder*, and includes terms such as *"input"*, *"output"*, *"signal"*, and *"circuit"*. This automation patent supplies a computer-based apparatus for detecting and correcting for color errors in video signals.

TABLE A.XIII

Examples of Individual Patents Linked to Census Occupations

Patent name	Linked Occupation
A. Examples of Occupation Linkages for 1940	
Thermal insulating material	Asbestos and insulation workers
Pie making process	Bakers
Towing dolly	Chauffeurs and drivers, bus, taxi, truck, and tractor
Process for the production of antiseptic agents	Chemical engineers
Corn popper	Cooks–except private family
Lever locking device	Cranemen, hoistmen, and construction machinery operators
Toecap for toe dancing shoes	Dancers, dancing teachers, and chorus girls
Spring roller venetian blind	Decorators and window dressers
Multiple elevator system	Elevator operators
Artificial fish bait	Fishermen and oystermen
Fruit squeezer	Fruit and vegetable graders and packers–except in cannery
Combination hand weeder and cultivator	Gardeners–except farm and groundskeepers
Mail covering	Mail carriers
Variable speed power transmission mechanism	Mechanics and repairmen–railroad and car shop
Chord finder for tenor banjos	Musicians and music teachers
Roof sump or floor drain	Plumbers and gas and steam fitters
Guided transmission of ultra high frequency waves	Radio and wireless operators
Telephone and telegraph signaling system	Telegraph operators
B. Examples of Occupation Linkages for 2018	
Systems and methods for unmanned aerial vehicle navigation	Aircraft pilots and flight engineers
Stabilised supersaturated solids of lipophilic drugs	Chemists and materials scientists
Telepresence robot with a camera boom	Communications equipment operators, all other
Systems and methods for detecting malware on mobile platforms	Computer programmers
Method of treating Attention Deficit Hyper-Activity Disorder	Counselors, all other
Document revisions in a collaborative computing environment	Editors
Mobile personal fitness training	Exercise trainers and group fitness instructors
Broccoli based nutritional supplements	Food cooking machine operators and tenders
Insulation with mixture of fiberglass and cellulose	Insulation workers
Determining text to speech pronunciation based on an utterance from a user	Interpreters and translators
Rotary drill bit including polycrystalline diamond cutting elements	Jewelers and precious stone and metal workers
Method and system for navigating a robotic garden tool	Landscaping and groundskeeping workers
Cuticle oil dispensing pen with ceramic stone	Manicurists and pedicurists
Adaptive audio conferencing based on participant location	Meeting, convention, and event planners
Identification and ranking of news stories of interest	News analysts, reporters, and journalists
Fumigation apparatus	Pest control workers
Low profile prosthetic foot	Podiatrists
Invertible trimmer line spool for a vegetation trimmer apparatus	Tree trimmers and pruners

F Constructing demand shifts using Chinese import competition and population aging

Here, we detail the construction of occupation-level demand shifts using (a) exogenous changes in Chinese import competition; and (b) exogenous shifts in consumption stemming from changing population demographics.

1 Chinese import competition

We obtain import data for manufacturing industries classified by consistent SIC 87 codes (`SIC87_dd`) over 1991–2014 from Autor, Dorn, and Hanson (2013). (Because the China trade shock had run its course by 2014 (Autor, Dorn, and Hanson, 2021), we use trade exposure data for 1991 to 2014.) We crosswalk these SIC 87-coded data to consistent Census industries (`ind1990ddx`) using 1991 value-added weights obtained from the NBER-CES Manufacturing Industry Database. To correspond with our Census data, which are coded in `ind1990dd_18` format, we create a classification (`ind1990ddx_18`) that aggregates manufacturing categories as needed to yield a balanced panel of 69 manufacturing industries.

We retain years 1991, 2000, and 2014, and construct long differences over 1991–2000 and 2000–2014, scaling these to match the time periods of our primary data (1990–2000 and 2000–2018). For each industry, this gives us two changes in import competition, defined as changes in Chinese imports for other developed countries ($\Delta M_{i,t}^{OC}$) divided by the industry's U.S. market size in 1988 (U.S. industry output plus imports minus exports, $Y_{i,1988} + M_{i,1988} - E_{i,1988}$). We use these industry-level changes in import exposure to construct occupational exposure to changes in import competition, as seen in equation (A1) below. Note that for non-manufacturing industries, the China exposure values are zero by definition. Hence, an occupation's exposure to the China trade shock depends on (1) the share of its employment that is found in manufacturing; and (2) the occupation's employment distribution across manufacturing industries that differ in their China trade exposure.[38]

$$\text{DemandX}_{j,t}^{C} = 100 \times \sum_{i} \frac{E_{ij,t-10}}{E_{j,t-10}} \times \frac{\Delta M_{i,t}^{OC}}{Y_{i,88} + M_{i,88} - X_{i,88}}. \tag{A1}$$

In this expression, the first term to the right of the equal sign is the share of employment of occupation j across industries i in the decade prior to the shock. The numerator of the second term, $\Delta M_{i,t}^{OC}$, is the change in each industry i's imports from China among a set of developed countries other than the United States over the periods 1991–2000 and 2000–2014, following Autor, Dorn, Hanson, and Song (2014) and subsequent papers. The denominator, $Y_{i,88} + M_{i,88} - X_{i,88}$, is initial domestic absorption of industry i's output, equal to the sum of the real value of industry shipments and industry imports minus industry exports, each measured in the initial, pre-shock year of 1988. Summing the product of the first and second terms across industries and multiplying by 100 yields an estimate of each occupation j's total exposure to the trade shock, expressed as a percentage point change relative to baseline.

[38] All regression models control for occupational employment shares in manufacturing so that identification is not driven by the simple manufacturing/nonmanufacturing contrast.

2 Population aging

We use Bureau of Labor Statistics' Consumption Expenditure (CE) data over 2002–2018 combined with Census population data over 1980–2018 to predict annual demand for each Uniform Commercial Code (UCC) product category over 1970–2018 (largely following DellaVigna and Pollet 2007, who use population aging to predict long-run stock market price changes among firms impacted by demographic change), and then crosswalk these predictions to consistent industries (`ind1990dd_18`) to obtain predicted consumption by industry.

Figure A.VII shows changes in the population by age over 1980–2000 and 2000–2018, highlighting the importance of the aging Baby Boom generation. In the first period, this cohort was prime-aged and having children, also leading to an increase for ages 0 to 10. Over the subsequent two decades, the entry of this cohort into middle- and late-adulthood created a large spike at ages 55 and above, as well as a smaller increase (echo) in the number of young adults.

For each UCC category (k), we take the following steps.

1. **Annualizing consumption.** The CE rotational design provides an unbalanced panel in which each consumption unit (CU)—effectively a household—appears in a subset of months. We use all twelve monthly CE surveys within each calendar year and scale up the recorded consumption of each CU by [12 ÷ (number of months the CU appears in the survey)].

2. **Pooling data.** We pool the annualized CE data across 2002–2018. For UCC categories that are not present in all years, we scale consumption by the number of years the UCC is observed. This yields c_{ik}, the average annual consumption for consumption unit i and UCC product category k.

3. **Estimate age-consumption profiles.** We estimate the age-consumption profile relating consumption by consumption unit i and product category k to the household structure observed in the CE data as:

$$c_{ik} = \sum_j \beta_{jk} H_{ij} + \sum_j \gamma_{jk} S_{ij} + \sum_j \delta_{jk} O_{ij} + \varepsilon_{ik},$$

   where H_{ij} is the dummy indicating whether household i has a head in age bin j, S_{ij} is a dummy indicating whether household i has a spouse in age bin j, and O_{ij} is the number of other people (i.e. other than head or spouse) of household i in age bin j, and ε_{ik} is the error term. Note that this regression has no intercept, such that the coefficients can be interpreted as consumption per household member. We estimate this model separately for each UCC product category and weight models by population weights. Note that pooling data across years assumes consumption profiles by age are stable over time: this is supported by DellaVigna and Pollet 2007's analysis.

4. **Calculating household age shares.** We estimate year-averaged shares of head, spouse,

and other household members using population weights available in CE data:

$$h_j = \frac{\sum_i \text{Nr of heads in CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}{\sum_i \text{Nr of total members of CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}$$
$$s_j = \frac{\sum_i \text{Nr of spouses in CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}{\sum_i \text{Nr of total members of CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}$$
$$o_j = \frac{\sum_i \text{Nr of other people in CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}{\sum_i \text{Nr of total members of CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}$$

5. **Predicting consumption.** We combine the estimated age-consumption coefficients and household share data (constructed above in steps 3 and 4, respectively) with Census population data over 1980–2018 to obtain aggregate predictions of consumption:

$$\widehat{c}_{kt} = \sum_j N_{j,t} \times (\widehat{\beta}_{j,k} h_j + \widehat{\gamma}_{j,k} s_j + \widehat{\delta}_{jk} o_j),$$

where $N_{j,t}$ is the total U.S. population within the age bin j in year t. As such, $\widehat{c}_{kt}$ is predicted consumption for product category k in year t, based on the changing age distribution of the population over 1980–2018.

6. **Crosswalking predictions to consistent Census industries.** We crosswalk these predictions to consistent Census industries using the following crosswalk path: CE → PCE 2017 → BEA commodity 2012 → BEA industry 2012 → NAICS 2012 → NAICS 2007 → CIC 2010 → CIC 1990 → ind1990dd → ind1990dd_18, where

   - CE is the Consumer Expenditure Survey;
   - PCE 2017 are 2017 Personal Consumption Expenditures;
   - BEA commodity 2012 are 2012 Bureau of Economic Analysis commodity codes;
   - BEA industry 2012 are 2012 Bureau of Economic Analysis industry codes;
   - NAICS 2012 are 2012 North American Industry Classification System codes;
   - NAICS 2007 are 2007 North American Industry Classification System;
   - CIC 2010 are 2010 Census industry codes;
   - CIC 1990 are 1990 Census industry codes;
   - ind1990dd are consistent industry codes constructed by David Dorn; and
   - ind1990dd_18 are our modified version of these codes to allow extension of the panel to 2018.

   To link CE to PCE we use the weights indicated by the BLS. For PCE categories that match to multiple BEA commodities, we allocate across categories using producer value as weights. This allows us to manually include the trade and transportation margins from the BEA use table when crosswalking PCE to BEA commodity codes without dropping retail and wholesale commodities. In all other crosswalks, expenditures are split evenly when one category matches to multiple categories. In our baseline demand shift results shown in Table V, we used full input-output adjustments (since industry demands intrinsically have an input-output component). Our results are robust to using demand shifts without input-output linkages, i.e. equating BEA commodity and industry codes.

We finally apply these predicted consumption levels by industry to construct occupational exposure to demographically-induced demand shifts for 1980–2000 and 2000–2018 as follows:

$$\text{DemandX}_{j,t}^{D} = 100 \times \sum_{i} \frac{E_{ij,t-1}}{E_{j,t-1}} \times \tilde{\Delta} \ln \text{demand}_{i,t}. \tag{A2}$$

Here $\tilde{\Delta} \ln \text{demand}_{i,t}$ is 100 times the predicted log change in demand for output of industry i in time interval t, $E_{ij,t-1}$ is the start-of-period employment of occupation j in industry i, and $E_{j,t-1}$ is the start-of-period employment of occupation j across all industries.

G Constructing composition-adjusted wages

Böhm, Gaudecker, and Schran (2022) and Autor and Dorn (2009) show that contracting occupations tend to retain more experienced workers—and workers with relatively high earnings given experience—while the opposite occurs in expanding occupations. This induces a negative correlation between occupational employment changes and wage changes. These compositional shifts—akin to quantity rather than price changes in an earnings equation—cloud inference on the earnings of workers of given skill levels. We address this issue by estimating the effect of augmentation and automation innovations on composition-constant wages in three steps.

In step one, we estimate cross-sectional log hourly wage regressions in each census year to obtain predicted wages in the primary Census and ACS samples:

$$w_{nt} = \alpha_{nt} + \mathbf{S_n}'\boldsymbol{\beta_{1t}} + (\mathbf{S_n} \times A_n)'\,\boldsymbol{\beta_{2t}} + \left(\mathbf{S_n} \times A_n^2\right)'\boldsymbol{\beta_{3t}} + e_{nt}. \tag{A3}$$

Here, w_{nt} is the log hourly earnings of worker n, $\mathbf{S_n}$ is a vector of dummies for completed schooling categories, and A_n is years of age. To account flexibly for education-experience profiles, equation (A3) includes a quadratic in age fully interacted with the vector of schooling levels. This model is fit separately for each of eight demographic groups (male/female $\times$ white/Black/Hispanic/other) in each time period to form a predicted wage for each worker, $\tilde{w}_{nt}$.

In step two, we collapse predicted and observed log wage levels into means within consistent industry-by-occupation cells. Combining these estimates with cell-level employment allows us to calculate observed wagebills ($W_{ij,t}$), predicted wagebills ($\widehat{W}_{ij,t}$) (means of fitted values of equation A3), and composition-adjusted wagebills ($\widetilde{W}_{ij,t}$), where the composition-adjusted wagebill is equal to the observed log change in employment plus the log difference between the observed and predicted wage in an industry-occupation cell.[39]

The final step estimates the relationship between augmentation exposure, automation exposure, and occupational wagebill changes using equation (3) above, where the dependent variable is the log change in an occupation-industry's wagebill ($\Delta W_{ij,t}$), expected wagebill ($\Delta\widehat{W}_{ij,t}$), or composition-adjusted wagebill ($\Delta\widetilde{W}_{ij,t}$).

[39] An industry-occupation cell with no employment has an undefined wagebill. A small subset of industry-occupation cells have positive employment in the IPUMS samples but no valid wage data because all workers in the cell are either self-employed or have invalid wage reports. We impute the predicted wage for these cells using fitted values from equation (A3).

H Individual-level employment in new titles using 1940 Census Complete Count data

Because we do not observe individual workers employed in new and preexisting micro occupation titles, we use occupational new title shares to approximate employment in new work in Section II.B. In this appendix, we apply 1940 Census Complete Count data to document that occupational new title shares are informative about the occupational distribution of individual-level employment in new work; and to study the characteristics of workers employed in new vs. preexisting job titles. We stress that our *primary* analyses predicting both the emergence of new titles and innovation-induced shifts in occupational labor demand do not make any assumptions about the number of workers employed in new versus preexisting titles. The exercise here is primarily relevant for calibrating the relationship between emergence of new titles and the count of workers employed in these titles.

1 Constructing individual-level employment in new work

We use individual-level data from the 1940 Census Complete Count (CCC) file, where workers' self-reported job titles ('write-ins') are unmasked and keyed. Self-reported titles are frequently vague and replete with misspellings. We employ a natural language inference model to link 99.8% of our worker sample in 1940.[40]

We begin by linking write-ins that match exactly with a CAI microtitle within the same macro-occupation—this assigns a CAI occupation title to 65% of the worker sample. For titles without an exact match, we use Facebook's *bart-large-mnli* model (Lewis et al., 2019), which estimates the strength of relationships between two sentences.

In particular, for each Census write-in, we obtain a list of CAI micro-titles within the same macro-occupation. We then construct two statements:

- Statement 1: This person's job is (Census write-in)
- Statement 2: This person's job is (CAI micro-title)

The Census write-in is then classified to the CAI title with the highest relationship score as determined by the model. Overall, 92% of micro-titles in the Census Alphabetical Index are linked to at least one CCC worker.[41]

To obtain the share of workers employed in new titles in each occupation, we aggregate individual employment counts in matched occupation titles to the 'macro' (3-digit) occupation level. Some CAI micro-titles contain industry restrictions. A small share of these (approximately 2.4%) are marked as "new" for some industry restrictions but "preexisting" for others. Since we do not use industry data to link CCC respondents, we assign the value of new employment for these respondents to be the share of micro-occupational titles marked as new across all industry restrictions within a macro-occupation.

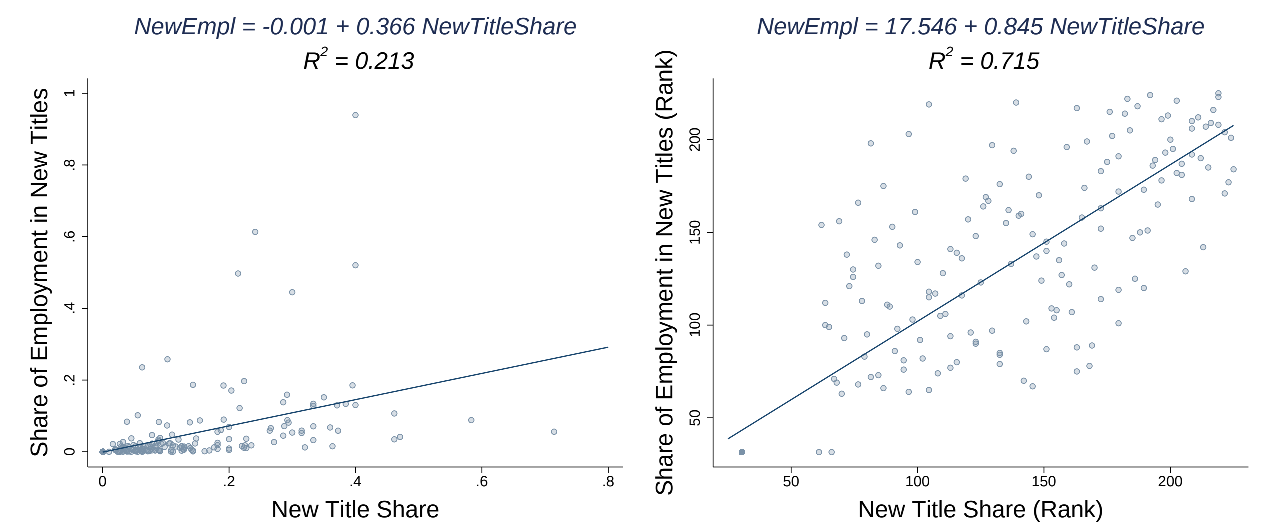
[40] The 0.2% unlinked observations either have missing write-ins, or contain write-ins with only one character or exclusively numeric characters.

[41] Since obsolete titles are retained in the CAI, we would not expect 100% of extant 1940 titles to be populated.

2 Comparison of occupational new title shares with individual-level employment in new work

Figure A.IX reports a pair of scatter plots depicting the relationships between occupations' new title shares (x-axis) and observed employment in new titles in those occupations. In the left-hand panel, the new work employment measure is the count of workers in new titles divided by total occupational employment. The relationship between new title shares and new work employment shares is positive and highly significant with a slope of 0.366 and a p-value below 0.01. The right-hand panel replaces the new work share measure with the *rank* of that measure. The coefficient on the new title share measure rises from 0.366 to 0.845 in this specification, and the fit of the regression is much tighter (the R^2 value is 0.213 in the first panel and 0.715 in the second).[42] These plots demonstrate that the new title share within macro-occupations is a strong positive correlate of the share of employment in new work within these occupations. Our regression exercises in the body of the paper relating new title emergence to both augmentation and automation innovations and to demand shifts (Section IV) do not impose any relationship between new title flows and employment in new work. These analyses rather establish that new title flows are elastic to these forces. Section V then demonstrates that forces that spur new title flows—augmentation innovations specifically—also raise occupational employment and wagebills.

FIGURE A.IX
Comparison of New Title Shares and Employment in New Work, 1940



$N = 225$ 1940 3-digit Census occupations. The left panel shows the relationship between the share of new titles in a 3-digit ('macro') occupation and the share of employment in new titles in that occupation in 1940, both measured on the unit interval. The right panel replaces the new work share measure with the *rank* of that measure (ranging from 1 to 225), where the lowest rank represents the occupation with the smallest share.

[42]These relationships, particularly in the rank-based specification, are partially driven by occupations with no new micro-occupation titles, which always have zero employment in new titles by definition. Limiting the sample to occupations with positive new title shares decreases the coefficient to 0.632 and R^2 to 0.390 in the rank specification.

While the slope in the left panel of Figure A.IX suggests that employment in new titles is substantially lower than employment in existing titles, there are at least two sources of bias that may lead to an underestimate of the employment count in new titles. First, new titles often represent the specialization of an existing title. In these cases, Census respondents who are employed in specialized fields may choose to report the more general version of their occupation, while workers with broad occupational responsibilities are unlikely to report specialized occupations. Second, new titles appear to be more difficult to link to Census write-ins than existing titles, likely as a result of new titles being more specialized and thus more specific. Concretely, we find that respondents with write-in titles that match exactly to titles in the CAI have lower rates of employment in new titles than workers matched to the CAI using more flexible matching procedures. This leads us to suspect that the true rate of employment in new titles exceeds the employment shares depicted in the left-hand panel of Figure A.IX.

3 Characteristics of workers employed in new and preexisting work

We next use the occupational write-in data in the Census Complete Count file to consider whether workers employed in new work in 1940 are more educated than and earn a wage premium relative to workers employed in preexisting titles, as we would expect if new work demands scarce expertise. Among workers who are matched to an occupation title from the 1940 Census Alphabetical Index, 2.60% are employed in titles that newly emerged between 1930 and 1940.

Table A.XIV reports the most common new CAI titles that are distinctive of new work emerging for three broad education groups: workers with less than a 9th grade education, workers with a high school degree, and workers with some college or higher education. To find these titles, we first identify the top 5% most commonly matched CAI titles in each education category s. We then calculate the share of all workers linked to these titles who have education level s relative to the share of workers with education s in our sample population. Titles with the highest relative share are listed in Table A.XIV. New work distinctive for college-educated workers reflects various specialized professional jobs, including “druggist (pharmacist)”, “patent attorney”, “geophysicist”, and “four-h agent”.[43] New clerical specialties such as “billing machine operator”, “bookkeeping machine operator”, “teletype operator”, and “social secretary” are dominant among high school graduates, while new titles among workers without a high school degree tend to be concentrated in production, such as “electrical layout man”, “caterpillar operator”, and “chemical operator”, as well as in (personal) services, e.g. “sandwich maker”, “delivery helper”, and “car hop”.

Panel A of Table A.XV regresses an indicator variable for employment in a new title (multiplied by 100 such that coefficients reflect percentage points) on a set of education dummies to explore whether better-educated workers are more likely to be employed in new titles, potentially reflecting greater demand for expertise. The data support this conjecture. Relative to workers with less than a 9th grade education, better-educated workers are substantially more likely to be found in new work. This partly reflects the fact that new work is more prevalent in higher-educated occupations. Column (2) accounts for this fact by adding a full set of 3-digit occupation fixed effects. Within 3-digit occupations, workers with college and some college are at least 0.6% more likely to be employed in new work than workers with only elementary educations. (Recall that the base rate of employment in new work is 2.60%, so these are large effects.) In column 3, we further control for

[43] 4-H (Head, Heart, Hands, and Health) is a network of youth organizations established around 1914, which in the US is administered by the National Institute of Food and Agriculture of the United States Department of Agriculture. https://en.wikipedia.org/wiki/4-H

worker demographics: age, age-squared, sex, race, geography, and 3-digit occupation. Accounting for these covariates has surprisingly little impact on the coefficients of interest: workers with at least some college education are more likely to be employed in new work relative to their high school graduate counterparts, who are in turn more likely to be in new work compared to those who did not obtain a high school degree. This education-new-work gradient for otherwise similar workers employed within the same 3-digit occupation is consistent with the hypothesis that new work requiring greater skill or expertise. (All estimates in Table A.XV are highly significant due to the vastness of the 1940 Census Complete Count file.)

We explore the wage returns associated with new work in Panel B of Table A.XV. Here, we regress workers' log annual earnings on an indicator for being employed in new work. Column 1 shows that workers employed in new titles earn 33% more than their counterparts in preexisting jobs. Even with a complete set of 3-digit occupation fixed effects included in column 2, this difference remains substantial at 10%. Column (3) further accounts for a full set of demographics as in panel A (age and age squared, sex, race, education, geography, and 3-digit occupation). The wage premium remains sizable at 6%.

In summary, workers employed in new work are more educated and higher-paid—even conditional on education and other worker observables—than workers employed in preexisting titles in the same detailed occupational categories. This pattern suggests—though does not prove—that new work may be more skilled, specialized, and better-remunerated than preexisting work.

TABLE A.XIV
Most Common New CAI Titles Distinctive for Each Education Group

Rank	Less Than 9th Grade	High School	At Least Some College
1	car hop	apprentice pressman	admission worker
2	chemical operator	attendant elevator	community chest official
3	delivery helper	billing machine operators	druggist (pharmacist)
4	electrical layout man	bookkeeping machine operator	engineer draftsman
5	route supervisor	caf manager	engineer sales or salesman
6	form man or setter	operative (machine operator)	four-h agent
7	layer-out	verifying machine operator	geophysicist
8	apprentice mechanic	social secretary	maintenance engineer (technical)
9	sandwich maker	teletype operator	naval official
10	caterpiller operator	apprentice mechanic	patent attorney

This table shows the most common new titles for workers of each education group, compared to how common they are among workers in the other education groups.

TABLE A.XV

The Relationship between Earnings Levels, Education, and the Probability of Employment in a New Occupational Title in 1940

	(1)	(2)	(3)
		A. Dep Var: 100 × Employment in New Work	
Education level (Reference category: Less than 9th grade education)			
Some high school	0.400***	0.231***	0.239***
	(0.006)	(0.006)	(0.006)
High school	0.376***	0.294***	0.337***
	(0.006)	(0.006)	(0.006)
Some college	3.747***	0.657***	0.651***
	(0.012)	(0.010)	(0.010)
College	8.249***	0.609***	0.576***
	(0.015)	(0.013)	(0.014)
R^2	0.022	0.284	0.284
		B. Dep Var: Log Annual Earnings	
Employed in New Work	0.323***	0.103***	0.060***
	(0.001)	(0.001)	(0.001)
R^2	0.002	0.274	0.399
Occupation FE		X	X
Full Controls			X

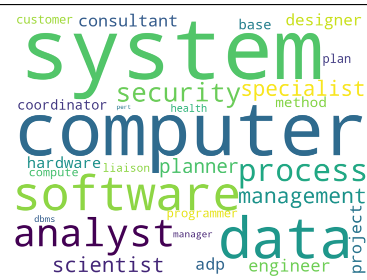
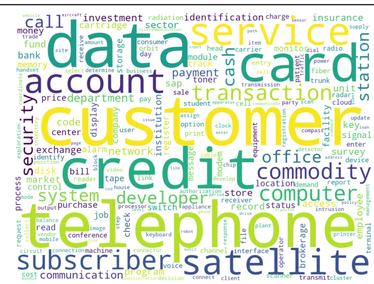
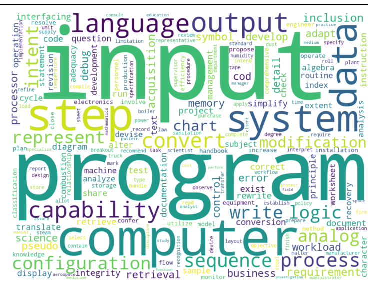
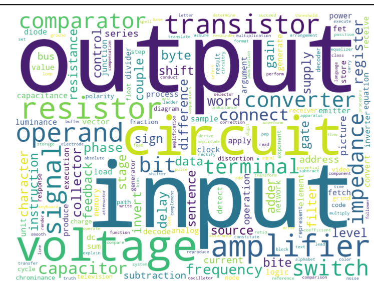
N = 32,631,552. Panel A estimates a linear probability model that compares the probability of employment in new work by educational attainment relative to workers who have less than a 9th grade education level. Mean probability of employment in new work = 0.026. Panel B shows the log differences in annual earnings among workers employed in new vs. preexisting work. Mean annual earnings = $1,276 in 1940 dollars. Column 3 includes controls for occupation, age, age-squared, sex, race, state, and urban/rural status. Column 3 in Panel B includes additional controls for education level. Sample includes employed workers $\geq$ 25 years old with reported education, non-zero reported income, and worked at least one week in the previous year. Robust standard errors reported in parentheses.

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

FIGURE A.X

Word Clouds for Textual Linkages Between Occupations and Patents

Computer Systems Analysts & Computer Scientists

<table>
<tr><th></th><th>Occupation terms</th><th>Patent terms</th></tr>
<tr><td>Augmentation Linked</td><td>
</td><td>
</td></tr>
<tr><td>Automation Linked</td><td>
</td><td>
</td></tr>
</table>

This figure depicts word clouds of terms that drive occupation-patent matches computer systems analysts & computer scientists. The left column contains terms from occupation documents with high similarity to patent documents. The right column contains terms from matched patents with high similarity scores to occupation documents. The size of included terms are proportional to their TF-IDF weighted cosine similarity to terms in either matched patents (left column) or matched occupations (right column).

FIGURE A.XI

Example Occupation Text for Computer Systems Analysts & Computer Scientists

1990 CAIO Micro-Titles	1977 DOT Task Description
64 Computer Systems Analysts & Computer Scientists • ADP customer liaison • ADP planner • ADP system coordinator • ADP systems security • Automatic data processing planner • Automatic data processing system coordinator • Computer analyst • Computer consultant • Computer programmer analyst • Computer scientist • Computer security information specialist • Computer security specialist • Computer specialist, software • Computer systems analyst • Computer systems designer • Computer-systems planning • Computing-systems analyst • DBMS specialist • Data base management system specialist • Data processing consultant • Data-processing-systems analyst • Data-processing-systems-project-planner • Digital-computer systems analyst • Engineer, Computer hardware • Engineer, Computer software • Engineer, Computer systems • Engineer, Software • Health-systems analyst, computer • Information scientist • Information systems consultant • Information systems specialist • Management scientist • Manager, Data systems • Methods analyst, computer • Pert analyst • Software designer • Software specialist • Systems analyst, computer systems • Systems analyst, data processing	**020.187-010 PROGRAMER, INFORMATION SYSTEM (profess. & kin.)** **Develops and writes natural and artificial language computer programs to store, locate, and retrieve specific documents, data, and information: Develops computer programs for input and retrieval of physical science, engineering or medical information, text analysis, and language, law, military, or library science data. Writes programs for classification, indexing, input, storage, and retrieval of data and facts, display devices, and interfacing with other systems equipment. Devises sample input data to test adequacy of program and observes or runs test of program, using sample or actual data. Corrects program errors by altering program steps and sequence. Confers with INFORMATION SCIENTIST (profess. & kin.) to resolve questions of program intent, input data acquisition, time sharing, output requirements, coding use and modification, and inclusion of internal checks and controls for system integrity.**

This figure reports the text corpora used to create occupation-level documents for Computer Systems Analysts & Computer Scientists. The left panel lists micro occupational titles from the 1990 CAI. The right panel shows the 1977 task description for "Programmer, Information System", an occupation within the Computer Systems Analysts & Computer Scientists macro-occupation.

FIGURE A.XII

Excerpts from Patents Linked to Computer Systems Analysts & Computer Scientists

Augmentation Linked	*US4310727A Method of processing special service telephone calls*
	• Direct signaling will allow the introduction of a host of new calling **services** to **telephone customers**.
	• A technical advance is achieved by a method of processing customized **telephone** calls in a system comprising a plurality of **telephone** offices, voice facilities, and a **data** communications network interconnecting prescribed ones of the offices.
	• This invention relates generally to communication call processing and to a method for providing customized call **services** to subscribers by means of the CCIS (Common Channel Interoffice Signaling) system.
Automation Linked	*US4847683A Diagonal correction in composite video decoder*
	• An apparatus for reducing cross color errors in luminance and chrominance **signals** recovered from an encoded video **signal** comprising: means for detecting the presence of high frequency diagonal lines in the recovered luminance **signal**.
	• The **output** of the luminance delay line 20 is **input** to another luminance delay line 24, a multiplier 26 and a shift register 28.
	• The removed luminance and chrominance signals, Yd (LUMA) and Cd (CHROMA), corrected for cross luminance, are **input** to a diagonal correction **circuit** 12 to produce the corrected chrominance and luminance signals CHROMA' and LUMA'.
	• The **output** of the demodulator is the corrected and demodulated chrominance **signal** CHROMA' in time with the corrected luminance **signal** LUMA' from the summer 50.

This figure reports example sentences from two selected patents that are augmentation-linked to Computer Systems Analysts & Computer Scientists in the top panel, and automation-linked to Computer Scientists Analysts & Computer Scientists in the bottom panel. Patent terms with high cosine similarity to words in the occupation documents are bolded.

I Comparison of new title models using Jeffrey Lin's (2011) new title measures

To probe the robustness of our new title results (Table II), we replace our measure of new work with the one constructed by Lin (2011) over 1980–2000. Lin (2011) constructs a new work inventory based on comparisons of Dictionary of Occupational Titles (DOT) records from 1965, 1977, and 1991; and Census Alphabetical Index (CAI) data from 1990 and 2000. He directly obtains new occupation titles for (approximately) 1980 and 1990 from the Conversion Table of Code and Title Changes (U.S. Department of Labor, 1979, 1991) which accompanies the DOT. For new titles in 2000, he uses a method similar to ours—a string match algorithm combined with manual review—to identify new titles. Our new title count shares (and new title counts) are positively correlated with Lin's across Census occupations, with Spearman rank correlations of 0.21 (0.53) in 1980, 0.19 (0.29) in 1990, and 0.50 (0.54) in 2000. The larger discordance for the first two decades likely arises from the differences in data sources and methods. The DOT contains many fewer new titles than the CAI (around 13,000 versus around 30,000). Furthermore, the data intervals covered are not fully aligned: ours CAI data always compares decadal intervals, whereas Lin compares 1977 and 1965 title lists for the 1970–1980 interval, and compares he 1991 and 1977 title lists for the 1980–1990 interval.

In Table A.XVI, we re-estimate the relationship between new title emergence and technology exposure (equation 1) for Census occupations by decade over 1980–2000 using as the dependent variable either our new title measure (columns 1–3) or Lin's (columns 4–6). Our augmentation and automation estimates are this 20-year time interval are highly comparable to our 40-year (1980–2018) estimates in Table II. When using Lin's measure, we find that augmentation exposure robustly predicts new work emergence across each decadal interval for which a comparison is feasible. Although point estimates are somewhat lower when using Lin's data relative to our primary data, coefficients are all positive and significant. Patterns for automation exposure are likewise consistent: while automation has a (weak) positive relationship with new work emergence on its own (columns 2 and 4), this disappears when augmentation is included in the specification (columns 3 and 6). Estimates for 2000 (panel D) are the most directly comparable since they are based on the same data source, method, and time interval; however, the comparisons for 1980 and 1990 highlight that our results hold also when using new titles identified from the DOT, a data source not used for linking augmentation patents. These findings bolster confidence in our results.

TABLE A.XVI
Occupational New Title Emergence and Augmentation
versus Automation Exposure, 1980-2000

Dependent Variable: Occupational New Title Count

	Benchmark				Lin	
	(1)	(2)	(3)	(4)	(5)	(6)
			A. 1980–2000			
Augmentation Exposure	21.43***		23.51***	18.41***		18.68***
	(3.22)		(3.71)	(3.23)		(3.61)
Automation Exposure		13.29***	−4.04		13.91***	−0.49
		(3.84)	(4.15)		(4.15)	(4.27)
			B. 1980			
Augmentation Exposure	32.99***		29.78***	28.15***		30.41***
	(4.90)		(5.14)	(6.98)		(7.67)
Automation Exposure		26.23***	5.90		17.40*	−3.40
		(5.81)	(5.03)		(7.53)	(7.41)
			C. 1990			
Augmentation Exposure	13.07*		24.30***	15.51***		17.47***
	(5.47)		(6.42)	(4.23)		(3.82)
Automation Exposure		3.31	−16.85**		12.25*	−3.08
		(5.71)	(6.48)		(5.43)	(5.30)
			D. 2000			
Augmentation Exposure	16.67***		19.91***	13.84**		13.69**
	(3.64)		(4.37)	(4.44)		(4.67)
Automation Exposure		5.35	−9.59		10.02	0.47
		(5.58)	(6.19)		(7.81)	(8.01)
Occ Emp Shares	X	X	X	X	X	X
Time FE	X	X	X	X	X	X

Negative binomial models, coefficients multiplied by 100. N = 759 in Panel A; N = 158 in Panel B; N = 297 in Panel C; N = 304 in Panel D. Twelve broad occupations are defined consistently across all decades. Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the log of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

J Theoretical model

1 Model

This section presents the model, with proofs given in Appendix 2.

Environment

We begin with two sectors, producing skill-intensive and skill-non-intensive goods or services, Y_S and Y_U. The subscripts denote the respective sectors. A representative household consumes goods Y_U and Y_S according to:

$$U(Y_U, Y_S) = Y_U^{\beta} Y_S^{1-\beta}, \tag{A4}$$

where $\beta \in (0,1)$; P_j is the price of good j with $j = U, S$; P the ideal price index; Y is total utility; and $P_U Y_U + P_S Y_S = PY$. Let Y be the numéraire so that $P \equiv 1$.[44] We will later allow for exogenous changes in β, reflecting demographic or taste shocks that shift preferences for consumption between skill-intensive and skill non-intensive services. We simplify the structure of consumption by assuming that there is no leisure and hence labor supply is inelastic.

Each sector produces a unique final output by combining a unit measure of tasks $i \in [N_j - 1, N_j]$:

$$Y_j = \left[\int_{N_j - 1}^{N_j} y_j(i)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}, \tag{A6}$$

where $y_j(i)$ is the output of task i in sector j; σ is the elasticity of substitution between tasks (assumed identical across sectors $j \in \{U, S\}$).

Each task is produced by combining labor composite of high- and low-skill types, $n_j(i)$, or capital, $k_j(i)$ with a task-specific intermediate $q_j(i)$. The production function for task i is given by:

$$y_j(i) = \begin{cases} B_j q_j(i)^{\eta} k_j(i)^{1-\eta} \text{ if } i \in [N_j - 1, I_j] \\ B_j q_j(i)^{\eta} [\gamma_j(i) n_j(i)]^{1-\eta} \text{ if } i \in (I_j, N_j] \end{cases}, \tag{A7}$$

where $B_j \equiv \psi_j^{\eta} [1-\eta]^{\eta-1} \eta^{-\eta}$ for notational convenience; the parameter $\eta \in (0,1)$ is the share of output paid to intermediates; $\gamma_j(i)$ is the productivity of the labor composite $n_j(i)$ (relative to capital); and I_j and N_j are the equilibrium thresholds for automation and new task creation, respectively, meaning tasks from $N_j - 1$ to I_j are produced by machines and those from I_j to N_j are produced by labor. We make the following assumption:

Assumption 1 *$\gamma_j(i)$ is strictly increasing.*

Assumption 1 implies that in each sector, labor has strict comparative advantage in tasks with a

[44] Given that consumption is Cobb-Douglas, P is given by:

$$P(P_U, P_S) = \left[\frac{P_U}{\beta} \right]^{\beta} \left[\frac{P_S}{1-\beta} \right]^{1-\beta} = 1. \tag{A5}$$

higher index. This assumption guarantees that, in equilibrium, tasks with lower indices will be automated in each sector, while those with higher indices will be produced with labor.

Task-specific intermediates $q_j(i)$ embody the technology used for the production of each task i. The automation of an existing task or creation of a new task requires the production of a corresponding new intermediate. We start by assuming that intermediates are supplied competitively and are produced using ψ_j units of the final good (and hence are priced at ψ_j in units of final output). When we model endogenous innovation responses below, profit opportunities in intermediate creation serve to allocate the supply of intermediate-creating entrepreneurs across sectors.

The measures of high-skill and low-skill labor are given by $H > 0$ and $L > 0$, respectively. The labor composite $n_j(i)$ in each sector is a Cobb-Douglas combination of H and L labor:

$$n_j(i) = l_j(i)^{\alpha_j} h_j(i)^{1-\alpha_j}. \tag{A8}$$

Both types of labor are used in each sector, but H labor is used more intensively in S sector, and L labor is used more intensively in the U sector ($0 < \alpha_S < \alpha_U < 1$). Let L_U, L_S, H_U, and H_S be the equilibrium labor allocations to each sector. Then, $L_U + L_S = L$ and $H_U + H_S = H$. We define here a wage index reflecting the price of the sectoral labor composite, $W_j \equiv \alpha_j^{-\alpha_j}(1-\alpha_j)^{\alpha_j - 1} W_L^{\alpha_j} W_H^{1-\alpha_j}$, where W_L and W_H equal the economy-wide wage for L and H labor, respectively. Finally, capital is sector-specific, with sectoral capital stocks K_U and K_S taken as given, and R_j is the capital rental rate for sector-specific capital.

Equilibrium

Before characterizing the equilibrium in our model, we simplify with two assumptions.

Assumption 2 *We have $K_j < \bar{K}_j$, where $\bar{K}_j$ is such that $R_j = \frac{W_j}{\gamma(N_j)}$ for $j \in \{U, S\}$.*

This ensures that the capital rental rate is sufficiently high in each sector that new tasks will be adopted immediately and will increase aggregate output. If Assumption 2 were not satisfied, new tasks would be more expensive to produce than the tasks that they potentially displace, i.e., the lowest index tasks, so that new tasks would either reduce productivity or would simply not be adopted.

The next assumption simplifies the determination of the automation threshold, I_j. Because labor has a strict comparative advantage in tasks with a higher index, i.e. i, there is a unique threshold $\tilde{I}_j$ in each sector such that

$$\frac{W_j}{R_j} = \gamma_j(\tilde{I}_j). \tag{A9}$$

For all tasks $i \leq \tilde{I}_j$, we have that $R_j \leq W_j/\gamma_j(i)$, so these tasks are potentially more cheaply produced with capital. However, if $I_j < \tilde{I}_j$, then the state of automation acts as a constraint on which tasks are accomplished by capital. In particular, the threshold task that will be performed by capital is $I_j^* = \min\{\tilde{I}_j, I_j\}$. We simplify the set of cases considered by invoking the following assumption:

Assumption 3 *We have that $I_j^* = I_j \leq \tilde{I}_j$, so that the threshold task in each sector is constrained by the state of automation.*

Assumption 3 implies that when a new automation technology is introduced, it is always adopted. With this assumption and the fact that tasks are competitively supplied, the price of task i, $p_j(i)$, is given by:

$$p_j(i) = \begin{cases} R_j^{1-\eta} \text{ if } i \in [N_j - 1, I_j] \\ [W_j/\gamma_j(i)]^{1-\eta} \text{ if } i \in (I_j, N_j] \end{cases} \tag{A10}$$

Combining equations (A4) and (A6), the demand for sectoral task output $y_j(i)$ is:

$$y_j(i) = [P_j/p_j(i)]^{\sigma} Y_j = \beta_j Y P_j^{\sigma-1} p_j(i)^{-\sigma}, \tag{A11}$$

where $j \in \{U, S\}, \beta_U = \beta$ and $\beta_S = 1 - \beta$. Together with the fact that the supply of $y_j(i)$ is a Cobb-Douglas aggregate of labor, capital, and intermediates, we can obtain the sectoral demands for capital and labor for each task i, respectively:

$$k_j(i) = \begin{cases} [1-\eta]\beta_j Y P_j^{\sigma-1} R_j^{-\hat{\sigma}} \text{ if } i \in [N_j - 1, I_j] \\ 0 \text{ if } i \in (I_j, N_j] \end{cases} \tag{A12}$$

and

$$l_j(i) = \begin{cases} 0 \text{ if } i \in [N_j - 1, I_j] \\ [1-\eta]\beta_j Y P_j^{\sigma-1} \frac{1}{\gamma_j(i)} \left[\frac{W_j}{\gamma_j(i)}\right]^{-\hat{\sigma}} \text{ if } i \in (I_j, N_j] \end{cases} \tag{A13}$$

where $\hat{\sigma} \equiv 1 - (1-\eta)(1-\sigma)$.

We can define a static equilibrium in a similar way to Acemoglu and Restrepo (2018b): Given a range of tasks $[N_j - 1, N_j]$, automation technology $I_j \in (N_j - 1, N_j]$, and a capital stock K_j for each sector j, a static equilibrium is summarized by a set of factor prices W_L, W_H, and R_j; threshold tasks $\tilde{I}$ and I^*; employment levels, L_j and H_j; and aggregate output, Y_j, for each sector j, such that

- $\tilde{I}_j$ is determined by equation (A9) and $I_j^* = \min\{I_j, \tilde{I}_j\}$, which is equal to I_j by Assumption 3;
- The capital and labor markets clear in each sector, so that

$$\int_{N_j-1}^{I_j} [1-\eta]\beta_j Y P_j^{\sigma-1} R_j^{-\hat{\sigma}} di = K_j \tag{A14}$$

$$\int_{I_j}^{N_j} [1-\eta]\beta_j Y P_j^{\sigma-1} \frac{1}{\gamma_j(i)} \left[\frac{W_j}{\gamma_j(i)}\right]^{-\hat{\sigma}} di = \mathcal{L}_j \tag{A15}$$

  where $\mathcal{L}_U \equiv L_U^{\alpha_U} H_U^{1-\alpha_U}$ and $\mathcal{L}_S \equiv (L - L_U)^{\alpha_S}(H - H_U)^{1-\alpha_S}$;

- Factor prices satisfy the ideal price index condition:

$$P_j^{1-\sigma} = [I_j - N_j + 1] R_j^{1-\hat{\sigma}} + W_j^{1-\hat{\sigma}} \int_{I_j}^{N_j} \gamma_j(i)^{\hat{\sigma}-1} di. \tag{A16}$$

Proposition 1 (Sectoral output) *In the static equilibrium defined above, aggregate output of sector j is given by:*

$$[1-\eta]Y_j = P_j^{\frac{\eta}{1-\eta}} \left[[I_j - N_j + 1]^{\frac{1}{\hat{\sigma}}} K_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}} + \left[\int_{I_j}^{N_j} \gamma_j(i)^{\hat{\sigma}-1} di \right]^{\frac{1}{\hat{\sigma}}} \mathcal{L}_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}} \right]^{\frac{\hat{\sigma}}{\hat{\sigma}-1}} \tag{A17}$$

Proof See Appendix 2.

Innovation and employment

We now consider the consequences of changes in the task structure in each sector, specifically, the effects of task automation and task augmentation, on sectoral employment and wagebills. Automation occurs when previously labor-using tasks are taken over by capital, corresponding to a rise in the sectoral automation threshold, I_j. Augmentation refers to the introduction of new labor-using tasks in a sector, corresponding to a rise in N_j. In a single-sector model, the effect of augmentation and automation on labor demand depend solely on substitution and scale effects in that sector. In our multi-sector setting with labor mobility and heterogeneous skills, augmentation and automation in either sector affects labor demand in both sectors, causing sectoral labor reallocation.

Proposition 2 (Employment effects of automation and augmentation) *Automation in sector U (a rise in I_U) increases the range of sector U tasks produced by capital, which decreases employment of both high-skill and low-skill workers in that sector. These workers move to sector S. Augmentation in sector U (a rise in N_U) has the converse effect: by introducing new labor-using tasks in sector U, it increases employment of both high-skill and low-skill workers in that sector, drawing away these workers from sector S. That is,*

$$\begin{array}{ll} \frac{\partial L_U}{\partial I_U}, \frac{\partial H_U}{\partial I_U} < 0, & \frac{\partial L_S}{\partial I_U}, \frac{\partial H_S}{\partial I_U} > 0 \\ \frac{\partial L_U}{\partial N_U}, \frac{\partial H_U}{\partial N_U} > 0, & \frac{\partial L_S}{\partial N_U}, \frac{\partial H_S}{\partial N_U} < 0. \end{array}$$

These derivatives have the opposite sign when augmentation or automation occurs in sector S.

Proof See Appendix 2.

This proposition, a key testable implication of the conceptual framework, reveals the direction of labor flows in response to automation and augmentation. All else equal, automation in a sector leads to the contraction of that sector by reducing employment of both types of workers, whereas augmentation in a sector attracts workers of both types.

Three mechanisms jointly underlie the co-movement of low- and high-skill workers across sectors in response to automation or augmentation. First, tasks are gross substitutes in each sector ($\sigma > 1$), so automation in a given sector implies a fall in that sector's labor share (and conversely for augmentation). Second, high- and low-skill labor are combined in Cobb-Douglas fashion in each sector, so the wagebill paid to each skill group by a sector is proportional to that sector's labor share. Finally, the share of aggregate expenditure devoted to each sector is fixed by the utility function (equation A4). Hence, automation in a sector spurs a decline in the sector's labor share, yielding an inward shift in both high- and low-skill sectoral labor demand relative to the other sector.

We directly test the implication that sectoral employment rises with sector-specific augmentation and falls with sector-specific automation in Section V, where we equate occupations in the empirical analysis with sectors in the model.

Remark 1 *Automation necessarily reduces employment in the automating sector, despite countervailing scale and substitution effects, due to the assumed Cobb-Douglas structure of consumer preferences (eqn A4). If instead, consumption goods were gross substitutes (i.e., with a substitution elasticity exceeding unity), automation's effect on sectoral employment would be ambiguous. In either case, distinct from automation, augmentation would necessarily increase employment in the directly affected sector due to substitution and scale effects that are both positive.*

Naturally, changes in sectoral labor demands alter the economy-wide skill premium, W_H/W_L, as explained in the next corollary.

Corollary 1 (Sectoral innovations and the aggregate skill premium) *Automation in the U sector raises the skill premium, W_H/W_L, by reducing labor demand in the low-skill intensive sector. Augmentation in the U sector lowers the skill premium by increasing labor demand in the low-skill intensive sector. Conversely, automation in the S sector lowers the skill premium while augmentation in the S sector raises the skill premium. Formally,*

$$\frac{\partial(W_H/W_L)}{\partial N_U}, \frac{\partial(W_H/W_L)}{\partial I_S} < 0, \quad \frac{\partial(W_H/W_L)}{\partial I_U}, \frac{\partial(W_H/W_L)}{\partial N_S} > 0.$$

This corollary spells out general equilibrium implications of innovations that reallocate the distribution of tasks between labor and capital in either sector. Our empirical analysis does not focus on these general equilibrium empirical implications, and the next corollary explains why.

Corollary 2 (Changes in sectoral wagebills by skill group) *Due to the law of one price for skill, the effect of innovation on the log sectoral wagebill of a skill group relative to its wagebill in the non-innovating sector is identical to its effect on the log relative sectoral employment of that skill group. Formally:*

$$\begin{aligned}\frac{\partial \ln(W_L L_U/W_L L_S)}{\partial I_U} &= \frac{\partial \ln(L_U/L_S)}{\partial I_U}, & \frac{\partial \ln(W_L L_U/W_L L_S)}{\partial N_U} &= \frac{\partial \ln(L_U/L_S)}{\partial N_U}\\ \frac{\partial \ln(W_H H_U/W_H H_S)}{\partial I_U} &= \frac{\partial \ln(H_U/H_S)}{\partial I_U}, & \frac{\partial \ln(W_H H_U/W_H H_S)}{\partial N_U} &= \frac{\partial \ln(H_U/H_S)}{\partial N_U}\end{aligned}$$

and similarly for innovation in the S sector.

This corollary, which echoes Proposition 3 in Hsieh et al. (2019) and follows from the mobility of labor across sectors, provides a testable implication that we examine in Section V: the impact of sectoral innovations—which we measure using augmentation and automation patents—on the sectoral wagebill by skill group will mirror those for sectoral employment.

Remark 2 *The prediction that wagebills expand or contract equiproportionately with employment in the affected sector follows from the assumption that high- and low-skill labor are combined Cobb-Douglas (in different proportions) in each sector, while the law of one price for skills prevails across* *sectors. More generally, with sector- or occupation-specific skills, or an elasticity of substitution greater than one across skill groups in each sector, wagebills could rise and fall more than proportionately with employment. In our empirical analysis in Section V, a finding that wagebills rise (fall) by at least as much employment in occupations exposed to augmentation (automation) is sufficient to establish that these effects capture net demand rather than net supply shifts.*

Shifts in consumer demand and innovation

To understand the interaction between shifts in consumer demand and innovation, we work with a simple, one-period framework, which utilizes the general results above but endogenizes the supply of intermediates that embody task-specific technologies. At the start of the period, two state variables determine the equilibrium variables: factor prices and output. A measure of entrepreneurs of exogenous supply E, where E is some large number, choose whether to supply labor to each of four sector-innovation cells: automation in sector U, new task creation in sector U, automation in sector S, and new task creation in sector S. We denote the number of entrepreneurs in each sector-innovation cell E_I^U, E_N^U, E_I^S, and E_N^S, respectively.

In each sector, entrepreneurs generate new intermediates that embody augmentation and automation technologies according to

$$\Delta I^j = E_I^j \tag{A18}$$

$$\Delta N^j = E_N^j \tag{A19}$$

ΔI^j and ΔN^j are realized immediately.

Entrepreneurs have utility given by

$$U_{j,z}^m = \max_{m \in \{I,N\},\ j \in \{U,S\}} \{w_j^m + \nu \epsilon_{j,z}^m\} \tag{A20}$$

where $U_{j,z}^m$ is the (period) utility of entrepreneur z working on innovation $m \in \{I, N\}$ in sector $j \in \{U, S\}$. The idiosyncratic preference terms $\epsilon_{j,z}^m$ are independent Type-I Extreme Value draws with zero mean, and the parameter ν scales the variance of these idiosyncratic terms. Entrepreneurs choose the sector and innovation activity that delivers the highest utility.

Under the distributional assumptions above, the share of entrepreneurial labor supplied to each sector-innovation cell has a closed-form analytical expression. Denote by π_j^m the fraction of entrepreneurs that move to sector j to work on innovation m. Then,

$$\pi_j^m = \frac{\exp{(w_j^m)}^{1/\nu}}{\sum_j \sum_m \exp{(w_j^m)}^{1/\nu}} \tag{A21}$$

Thus, $1/\nu$ can be interpreted as a labor supply elasticity, as in Caliendo, Dvorkin, and Parro (2019). Applying the law of large numbers, the measure of entrepreneurs supplying labor to sector j working on innovation m is

$$E_j^m = \pi_j^m E \tag{A22}$$

Competition among entrepreneurs to become technology monopolists implies wages as follows:

$$w_j^m = V_j^m \tag{A23}$$

where V_j^m is the value of innovation m in sector j.

Demand for intermediate $q_j(i)$ is given by:

$$q_j(i) = \begin{cases} \psi_j^{-1}\eta Y_j P_j^{\sigma} R_j^{(1-\eta)(1-\sigma)} & \text{if } i \in [N_j - 1, I_j] \\ \psi_j^{-1}\eta Y_j P_j^{\sigma}\left(\frac{W_j}{\gamma_j(i)}\right)^{(1-\eta)(1-\sigma)} & \text{if } i \in (I_j, N_j] \end{cases} \tag{A24}$$

Gross profit from automating task I are

$$\pi(I) = \begin{cases} (1-\mu)\psi_j q_j(I) = (1-\mu)\eta Y_j P_j^{\sigma} R_j^{(1-\eta)(1-\sigma)} & \text{if } I \text{ is produced with capital} \\ (1-\mu)\psi_j q_j(I) = (1-\mu)\eta Y_j P_j^{\sigma}\left(\frac{W_j}{\gamma_j(I)}\right)^{(1-\eta)(1-\sigma)} & \text{if } I \text{ is produced with labor} \end{cases} \tag{A25}$$

where $1-\mu$ represents the per-unit profit relative to the marginal cost of the intermediate-good firm.[45] Note that $(1-\eta)(1-\sigma) \equiv 1-\hat{\sigma}$. Hence the sectoral value of automating task I and of creating task N are given by, respectively:

$$V_j^I = (1-\mu)\eta Y_j P_j^{\sigma}\left[R_j^{1-\hat{\sigma}} - \left(\frac{W_j}{\gamma_j(I)}\right)^{1-\hat{\sigma}}\right] \tag{A26}$$

$$V_j^N = (1-\mu)\eta Y_j P_j^{\sigma}\left[\left(\frac{W_j}{\gamma_j(N)}\right)^{1-\hat{\sigma}} - R_j^{1-\hat{\sigma}}\right] \tag{A27}$$

Having determined the sectoral value of automating task I and of creating task N, we study how these incentives for automation and new task creation change in sector j in response to a demand expansion, i.e., an increase in β if $j = U$, or an increase in $(1-\beta)$ if $j = S$.

Lemma 1 *In equilibrium, we have that $V_j^N = V_j^I$.*

Lemma 1 implies that entrepreneurs are initially indifferent between creating new automation or augmentation intermediates; otherwise the initial allocation of entrepreneurs was not in equilibrium. Note that in this equilibrium there are still positive productivity gains from additional task automation and from additional new task creation, by Assumptions 2 and 3.

As these incentives given by V_j^I and V_j^N determine wages, the employment share and changes in I and N for each sector naturally follow by consulting (A21) and the innovation production functions (A18) and (A19).

Proposition 3 *A demand shift towards a given sector unambiguously increases new task creation relative to automation in that sector, while decreasing new task creation relative to automation in*

[45] We follow Acemoglu and Restrepo (2018b) to assume that a firm with entrepreneurs has a proportionally lower marginal cost, $\mu\psi$, compared to ψ, the marginal cost of a firm without entrepreneurs.

the other sector.

$$\frac{\partial \Delta N_U}{\partial \beta} > \frac{\partial \Delta I_U}{\partial \beta}, \quad \frac{\partial \Delta N_U}{\partial (1-\beta)} < \frac{\partial \Delta I_U}{\partial (1-\beta)}$$
$$\frac{\partial \Delta N_S}{\partial \beta} < \frac{\partial \Delta I_S}{\partial \beta}, \quad \frac{\partial \Delta N_S}{\partial (1-\beta)} > \frac{\partial \Delta I_S}{\partial (1-\beta)}$$

Proof. See Appendix 2.

This proposition indicates a positive relationship between demand shifts and new task creation. When there is a positive demand shift in a given sector j, the incentives for new task creation in that sector increase as a result of movement on two margins: on the demand side, both output and price increases, and on the factor side, the price of capital increases more than that of effective labor as capital supply is inelastic, increasing the price differential between the two factors. This increased price differential raises the potential returns to new task creation, which assigns tasks from capital to labor.[46] Section IV.C corroborates an implications of this proposition: outward demand shifts accelerate new task emergence whereas inward demand shifts decelerate it.

2 Proofs of propositions

Proof of proposition 1

The price index P_j is the minimum cost for buying an additional unit of good j in equilibrium. Given that equation (A6) is CES, the corresponding expression for the marginal cost of producing Y_j is given by:

$$P_j = \left[\int_{N_j-1}^{N_j} p_j(i)^{1-\sigma} di\right]^{\frac{1}{1-\sigma}} \tag{A28}$$

Using equation (A10), equation (A28) can be written as:

$$\begin{aligned} P_j^{1-\sigma} &= \int_{N_j-1}^{I_j} R_j^{[1-\eta][1-\sigma]} di + \int_{I_j}^{N_j} \left[\frac{W_j}{\gamma_j(i)}\right]^{[1-\eta][1-\sigma]} di \\ &= [I_j - N_j + 1] R_j^{1-\hat{\sigma}} + W_j^{1-\hat{\sigma}} \int_{I_j}^{N_j} \gamma_j(i)^{\hat{\sigma}-1} di \end{aligned} \tag{A29}$$

with $[1-\eta][1-\sigma] = 1 - \hat{\sigma}$ given that $\hat{\sigma} \equiv [1-\eta]\sigma + \eta$.

We can rewrite the factor-market clearing conditions (equations (A14) and (A15)) to solve for W_j and R_j in equilibrium:

$$W_j = \left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{\mathcal{L}j} \int_{I_j}^{N_j} \gamma_j(i)^{\hat{\sigma}-1} di\right]^{\frac{1}{\hat{\sigma}}} \tag{A30}$$

[46] This result relies on the fact that tasks are gross substitutes in each sector, $\sigma > 1$, so that a change in the sectoral relative price of capital versus labor increases the profitability of innovations that expand usage of the factor whose relative price has fallen.

and

$$R_j = \left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{K_j}[I_j - N_j + 1]\right]^{\frac{1}{\hat{\sigma}}}. \tag{A31}$$

Substituting the expressions for W_j and R_j from equations (A30) and (A31) into equation (A29) gives:

$$\begin{aligned} P_j^{1-\sigma} =& [I_j - N_j + 1]\left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{K_j}[I_j - N_j + 1]\right]^{\frac{1-\hat{\sigma}}{\hat{\sigma}}} \\ &+ \left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{\mathcal{L}_j}\int_{I_j}^{N_j}\gamma_j(i)^{\hat{\sigma}-1}di\right]^{\frac{1-\hat{\sigma}}{\hat{\sigma}}}\int_{I_j}^{N_j}\gamma_j(i)^{\hat{\sigma}-1}di \end{aligned}$$

Bringing the P_j on the left-hand side to the right-hand side and using that $\frac{\sigma-\hat{\sigma}}{\hat{\sigma}-1} = \eta/[1-\eta]$ gives:

$$[1-\eta]Y_j = P_j^{\frac{\eta}{1-\eta}}\left[[I_j - N_j + 1]^{\frac{1}{\hat{\sigma}}}K_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}} + [\int_{I_j}^{N_j}\gamma_j(i)^{\hat{\sigma}-1}di]^{\frac{1}{\hat{\sigma}}}\mathcal{L}_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}}\right]^{\frac{\hat{\sigma}}{\hat{\sigma}-1}}$$

Proof of proposition 2

Wagebills must satisfy:

$$W_L L_U = \alpha^U s_U^L P_U Y_U = \alpha^U s_U^L \beta Y \tag{A32}$$
$$W_L L_S = \alpha^S s_S^L P_S Y_S = \alpha^S s_S^L [1-\beta] Y \tag{A33}$$
$$W_H H_U = (1-\alpha^U) s_U^L P_U Y_U = (1-\alpha^U) s_U^L \beta Y \tag{A34}$$
$$W_H H_S = (1-\alpha^S) s_S^L P_S Y_S = (1-\alpha^S) s_S^L [1-\beta] Y \tag{A35}$$

where W_L and W_H are the respective wages for low-skilled and high-skilled workers. Acemoglu and Restrepo (2019) show that the labor share in sector j is given by:

$$s_j^L = \left[1 + \left[\frac{1-\Gamma_j}{\Gamma_j}\right]^{\frac{1}{\sigma}}\left[\frac{K_j}{\mathcal{L}_j}\right]^{\frac{\sigma-1}{\sigma}}\right]^{-1} \tag{A36}$$

with σ the elasticity between tasks in goods production and

$$\Gamma_j \equiv \frac{\int_{I_j}^{N_j}\gamma^L(i)^{\sigma-1}di}{[I_j - N_j + 1]^{\sigma-1} + \int_{I_j}^{N_j}\gamma_j(i)^{\sigma-1}di}. \tag{A37}$$

Step 1. Wage determination. From assuming perfect competition in the low-skilled and the high-skilled labor market, the marginal product of low-skilled and high-skilled labor to produce one unit of efficient labor (labor composite) in both the skill-non-intensive and the skill-intensive sector must be equal to their corresponding marginal costs, which are the low-skilled wage and the high-skilled wage, respectively. Formally,

$$W_L = \alpha_U L_U^{\alpha_U - 1} H_U^{1-\alpha_U} W_U = \alpha_S (L - L_U)^{\alpha_S - 1}(H - H_U)^{1-\alpha_S} W_S \tag{A38}$$

$$W_H = (1-\alpha_U) L_U^{\alpha_U} H_U^{-\alpha_U} W_U = (1-\alpha_S)(L - L_U)^{\alpha_S}(H - H_U)^{-\alpha_S} W_S \tag{A39}$$

Rewriting using the expressions for $\mathcal{L}_U$ and $\mathcal{L}_S$ above, we have

$$W_L = \alpha_U \frac{\mathcal{L}_U}{L_U} W_U = \alpha_S \frac{\mathcal{L}_S}{L - L_U} W_S \tag{A40}$$

$$W_H = (1-\alpha_U) \frac{\mathcal{L}_U}{H_U} W_U = (1-\alpha_S) \frac{\mathcal{L}_S}{H - H_U} W_S \tag{A41}$$

Thus,

$$\frac{\alpha_U}{1-\alpha_U} \frac{L - L_U}{L_U} = \frac{\alpha_S}{1-\alpha_S} \frac{H - H_U}{H_U} \tag{A42}$$

Therefore, we know that L_U, H_U, and thus their product $\mathcal{L}_U$ always move in the same direction, and this is opposite in sign to the movements of $L - L_U$, $H - H_U$, and their product $\mathcal{L}_S$.

Given the co-movements of labor allocations, in the following steps, it is sufficient to analyze the relationship between high-skilled labor supply in sector S and the relative wage ratio to pin down the entire labor allocation. The responses to innovations follow from there.

Step 2: Sectoral Labor Supply Condition. From (A42), we can solve for L_U in terms of H_U:

$$L_U = \frac{L}{\frac{1-\alpha_U}{\alpha_U} \frac{\alpha_S}{1-\alpha_S} \frac{H - H_U}{H_U} + 1} \tag{A43}$$

This gives us an expression for the relative wage ratio of high- and low-skilled labor:

$$\frac{W_H}{W_L} = \frac{1-\alpha_U}{\alpha_U} \frac{L_U}{H_U} \tag{A44}$$

Or alternatively,

$$H_S = \left(\frac{L}{\frac{W_H}{W_L}} - \frac{\alpha_U}{1-\alpha_U} H \right) \frac{(1-\alpha_S)(1-\alpha_U)}{\alpha_S - \alpha_U} \tag{A45}$$

Therefore, since $\alpha_U > \alpha_S$, the supply of high-skilled labor in sector S, $H_S \equiv H - H_U$, is increasing in the relative wage ratio, $\frac{W_H}{W_L}$. We call this upward-sloping relationship the Sectoral Labor Supply Condition, indicating that when $\frac{W_H}{W_L}$ increases, both types of labor flow into the skill-intensive sector, S.

Step 3: Sectoral Labor Demand Condition. Combining equations (A32), (A33), (A34), and (A35) we have

$$\frac{\mathcal{L}_U}{\mathcal{L}_S} = C\frac{s_U^L}{s_S^L} \tag{A46}$$

$$= C\frac{1+\left[\frac{1-\Gamma_S}{\Gamma_S}\right]^{\frac{1}{\sigma}}\left[\frac{K_S}{\mathcal{L}_S}\right]^{\frac{\sigma-1}{\sigma}}}{1+\left[\frac{1-\Gamma_U}{\Gamma_U}\right]^{\frac{1}{\sigma}}\left[\frac{K_U}{\mathcal{L}_U}\right]^{\frac{\sigma-1}{\sigma}}} \tag{A47}$$

where $C \equiv \frac{W_S}{W_U}\frac{\beta}{1-\beta} = (\frac{W_H}{W_L})^{\alpha_U-\alpha_S}\frac{\alpha_U^{\alpha_U}(1-\alpha_U)^{1-\alpha_U}\beta}{\alpha_S^{\alpha_S}(1-\alpha_S)^{1-\alpha_S}(1-\beta)}$.

Therefore, combining the implication from (A42), the fact that $\sigma > 1$, and the fact that $\alpha_U > \alpha_S$, we know $\mathcal{L}_U$ must be increasing with $\frac{W_H}{W_L}$ and it is the reverse for $\mathcal{L}_S$. Therefore, H_S is also decreasing in $\frac{W_H}{W_L}$. We call this downward-sloping relationship the Sectoral Labor Demand Condition. This downward-sloping relationship captures the fact that rise in $\frac{W_H}{W_L}$ increases the output price for the skill-intensive sector S, which causes the representative consumer to substitute away from S and towards U. As a result, labor demand for both H and L fall in S and rise in U.

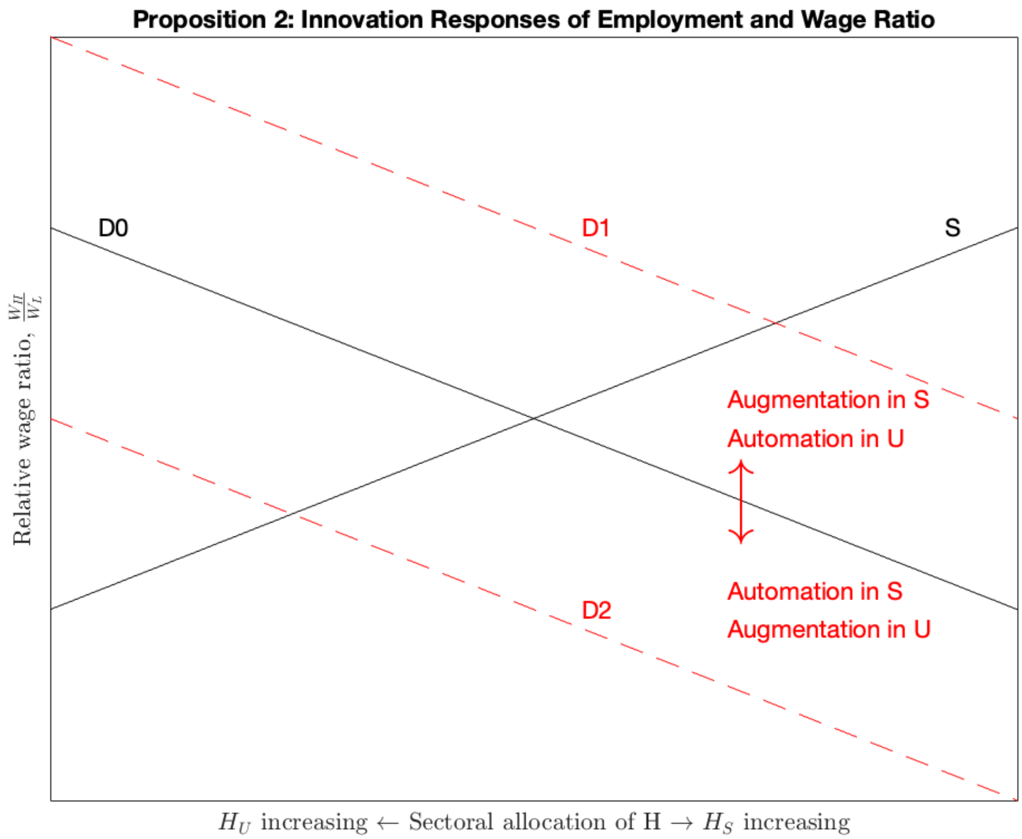
We use a graph with H_S as the x-axis and $\frac{W_H}{W_L}$ as the y-axis to illustrate that:

- Automation in sector U decreases Γ_U, so that relative demand for labor in sector S shifts outward, and in equilibrium, $\mathcal{L}_U$, L_U, and H_U decrease, $\mathcal{L}_S$, L_S, and H_S increase, and $\frac{W_H}{W_L}$ increases;

- Augmentation in sector U increases Γ_U, so that the relative demand for labor in sector S shifts inward, $\mathcal{L}_U$, L_U, and H_U increase, $\mathcal{L}_S$, L_S, and H_S decrease, and $\frac{W_H}{W_L}$ decreases;

- Automation in sector S decreases Γ_S, so that the relative demand for labor in sector S shifts inward, $\mathcal{L}_U$, L_U, and H_U increase, $\mathcal{L}_S$, L_S, and H_S decrease, and $\frac{W_H}{W_L}$ decreases;

- Augmentation in sector S increases Γ_S, so that the relative demand for labor in sector S shifts outward, $\mathcal{L}_U$, L_U, and H_U decrease, $\mathcal{L}_S$, L_S, and H_S increase, and $\frac{W_H}{W_L}$ increases



Proof of corollary 1

Included in the proof for Proposition 2.

Proof of corollary 2

Follows directly from taking ratios among (A32), (A33), (A34), and (A35).

Proof of proposition 3

Assumption 3 implies

$$\frac{W_j}{\gamma(I^*)} > R_j > \frac{W_j}{\gamma(I(N))}$$

so that it is always profitable to automate and create new tasks. In the initial equilibrium, we will have that $V_j^N = V_j^I$ (see Lemma 1). Differentiating the value functions of sector j with respect to β, we obtain the effect of a positive demand shift on incentives for automation and new task creation in sector j:

$$\frac{\partial V_j^I}{\partial \beta} = \underbrace{\frac{\partial Y_j P_j^\sigma}{\partial \beta}}_{A} \times \underbrace{(1-\mu)\eta \left[R_j^{1-\hat{\sigma}} - \left(\frac{W_j}{\gamma_j(I)}\right)^{1-\hat{\sigma}}\right]}_{B} + \underbrace{\frac{\partial \left[R_j^{1-\hat{\sigma}} - \left(\frac{W_j}{\gamma_j(I)}\right)^{1-\hat{\sigma}}\right]}{\partial \beta}}_{C} \times \underbrace{(1-\mu)\eta Y_j P_j^\sigma}_{D} \tag{A48}$$

$$\frac{\partial V_j^N}{\partial \beta} = \underbrace{\frac{\partial Y_j P_j^\sigma}{\partial \beta}}_{A} \times \underbrace{(1-\mu)\eta \left[\left(\frac{W_j}{\gamma_j(N)}\right)^{1-\hat{\sigma}} - R_j^{1-\hat{\sigma}}\right]}_{E} + \underbrace{\frac{\partial \left[\left(\frac{W_j}{\gamma_j(N)}\right)^{1-\hat{\sigma}} - R_j^{1-\hat{\sigma}}\right]}{\partial \beta}}_{F} \times \underbrace{(1-\mu)\eta Y_j P_j^\sigma}_{D} \tag{A49}$$

Equations (A48) and (A49) cover two cases. First, when $j = U$, term (A) is *positive* as output and prices increase from the outward demand shift. For the value of additional task automation, A multiplies a term (B) which is *positive* by Assumption 3, as an increase in the range of tasks that are automated increases productivity. Similarly, for the value of additional augmentation, A multiplies a term (E) which is *positive* by Assumption 2, as an increase in new tasks also increases productivity. Since rental rates rise more strongly than do wages, term C is *negative*, and term F is *positive*; and both C and F multiply a *positive* term (D). Therefore, the incentive for new task creation in the sector with the demand expansion are unambiguously positive and exceed the incentive for task automation in this sector, $\frac{\partial V_U^N}{\partial \beta} > \frac{\partial V_U^I}{\partial \beta}$, if in the initial equilibrium $V_j^N = V_j^I$ since this implies $B = E$.

Second, when $j = S$, terms B, D, and E remain positive, while the sign of terms A, C, and F reverse. Hence, in response to a demand contraction in sector U, the incentive for new task creation in sector S is reduced: both overall, $\frac{\partial V_S^N}{\partial \beta} < 0$, and relative to automation in the sector $\frac{\partial V_S^N}{\partial \beta} < \frac{\partial V_S^I}{\partial \beta}$.

We can summarize the relative magnitudes of the changes in the value of innovations in response to changes in demand as follows:

$$\frac{\partial V_U^N}{\partial \beta} > \frac{\partial V_U^I}{\partial \beta}, \frac{\partial V_S^N}{\partial \beta} < \frac{\partial V_S^I}{\partial \beta}$$

Since in a two-sector model, a demand expansion in one sector implies a relative demand contraction in the other, the responses of innovation incentives to a positive demand shift in sector S (a fall in β) follows directly from above:

$$\frac{\partial V_U^N}{\partial (1-\beta)} < \frac{\partial V_U^I}{\partial (1-\beta)}, \frac{\partial V_S^N}{\partial (1-\beta)} > \frac{\partial V_S^I}{\partial (1-\beta)}$$

Because entrepreneurs' wages in a given sector-innovation cell are equal to the value of the innovations they create intermediates for ($w_j^m = V_j^m$), and since $\Delta I_j = E_j^I$ and $\Delta N_j = E_j^N$, we obtain that

$$\begin{aligned} \tfrac{\partial \Delta N_U}{\partial \beta} &> \tfrac{\partial \Delta I_U}{\partial \beta}, \tfrac{\partial \Delta N_U}{\partial (1-\beta)} &< \tfrac{\partial \Delta I_U}{\partial (1-\beta)} \\ \tfrac{\partial \Delta N_S}{\partial \beta} &< \tfrac{\partial \Delta I_S}{\partial \beta}, \tfrac{\partial \Delta N_S}{\partial (1-\beta)} &> \tfrac{\partial \Delta I_S}{\partial (1-\beta)}, \end{aligned}$$

which corresponds to our proposition.