The Journal of Economic Perspectives

A journal of the
American Economic Association

Spring 2019

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@jepjournal.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed by LSC Communications, Owensville, Missouri, 65066, USA.
No responsibility for the views expressed by the authors in this journal is assumed by the editors or by the American Economic Association.
THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Spring 2019, Vol. 33, No. 2. The *JEP* is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $22.00, $32.00, or $42.00 depending on income; for an additional $15.00, you can receive this journal in print. The journal is freely available online. For details and further information on the AEA go to https://www.aeaweb.org/.
Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 33 • Number 2 • Spring 2019

Symposia

Automation and Employment

Daron Acemoglu and Pascual Restrepo, "Automation and New Tasks: How Technology Displaces and Reinstates Labor" 3

Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb, "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction" 31

Jeremy Atack, Robert A. Margo, and Paul W. Rhode, "'Automation' of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study" 51

Hong Cheng, Ruixue Jia, Dandan Li, and Hongbin Li, "The Rise of Robots in China". 71

Fiscal Policy

Valerie A. Ramey, "Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" 89

Pierre Yared, "Rising Government Debt: Causes and Solutions for a Decades-Old Trend" 115

Alberto Alesina, Carlo Favero, and Francesco Giavazzi, "Effects of Austerity: Expenditure- and Tax-based Approaches" 141

The Problems of Men

Ariel J. Binder and John Bound, "The Declining Labor Market Prospects of Less-Educated Men" 163

Courtney C. Coile and Mark G. Duggan, "When Labor's Lost: Health, Family Life, Incarceration, and Education in a Time of Declining Economic Opportunity for Low-Skilled Men" 191

Kathryn Edin, Timothy Nelson, Andrew Cherlin, and Robert Francis, "The Tenuous Attachments of Working-Class Men" 211

Features

Samuel Hollander, "Retrospectives: Ricardo on Machinery". 229

Timothy Taylor, "Recommendations for Further Reading" 243

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 3–30

Automation and New Tasks: How Technology Displaces and Reinstates Labor

Daron Acemoglu and Pascual Restrepo

The implications of technological change for employment and wages are a source of controversy. Some see the ongoing process of automation—as exemplified by computer numerical control machinery, industrial robots, and artificial intelligence—as the harbinger of widespread joblessness. Others reason that current automation, like previous waves of technologies, will ultimately increase labor demand, and thus employment and wages.

This paper presents a task-based framework, building on Acemoglu and Restrepo (2018a, 2018b) as well as Acemoglu and Autor (2011), Autor, Levy, and Murnane (2003), and Zeira (1998), for thinking through the implications of technology for labor demand and productivity. Production requires tasks, which are allocated to capital or labor. New technologies not only increase the productivity of capital and labor at tasks they currently perform, but also impact the allocation of tasks to these factors of production—what we call the *task content of production*. Shifts in the task content of production can have major effects for how labor demand changes as well as for productivity.

Automation corresponds to the development and adoption of new technologies that enable capital to be substituted for labor in a range of tasks. Automation changes the task content of production adversely for labor because of a *displacement effect*—as capital takes over tasks previously performed by labor. The displacement

■ *Daron Acemoglu is Elizabeth and James Killian Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Pascual Restrepo is Assistant Professor of Economics, Boston University, Boston, Massachusetts. Their emails are daron@mit.edu and pascual@bu.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.3 doi=10.1257/jep.33.2.3

effect implies that automation reduces the labor share of value added. Historical examples of automation are aplenty. Many early innovations of the Industrial Revolution automated tasks performed by artisans in spinning and weaving (Mantoux 1928), which led to widespread displacement, triggering the Luddite riots (Mokyr 1990). The mechanization of agriculture, which picked up speed with horse-powered reapers, harvesters, and plows in the second half of the 19th century and with tractors and combine harvesters in the 20th century, displaced agricultural workers in large numbers (Rasmussen 1982; Olmstead and Rhode 2001). Today too we are witnessing a period of rapid automation. The jobs of production workers are being disrupted with the rise of industrial robots and other automated machinery (Graetz and Michaels 2018; Acemoglu and Restrepo 2018b), while white-collar workers in accounting, sales, logistics, trading, and some managerial occupations are seeing some of the tasks they used to perform being replaced by specialized software and artificial intelligence.

By allowing a more flexible allocation of tasks to factors, automation technology also increases productivity, and via this channel, which we call the *productivity effect*, it contributes to the demand for labor in non-automated tasks. The net impact of automation on labor demand thus depends on how the displacement and productivity effects weigh against each other.

The history of technology is not only about the displacement of human labor by automation technologies. If it were, we would be confined to a shrinking set of old tasks and jobs, with a steadily declining labor share in national income. Instead, the displacement effect of automation has been counterbalanced by technologies that create new tasks in which labor has a comparative advantage. Such new tasks generate not only a positive productivity effect, but also a *reinstatement effect*—they reinstate labor into a broader range of tasks and thus change the task content of production in favor of labor.[1] The reinstatement effect is the polar opposite of the displacement effect and directly increases the labor share as well as labor demand.

History is also replete with examples of the creation of new tasks and the reinstatement effect. In the 19th century, as automation of some tasks was ongoing, other technological developments generated employment opportunities in new occupations. These included jobs for line workers, engineers, machinists, repairmen, conductors, managers, and financiers (Chandler 1977; Mokyr 1990). New occupations and jobs in new industries also played a pivotal role in generating labor demand during the decades of rapid agricultural mechanization in the United States, especially in factories (Rasmussen 1982; Olmsted and Rhode 2001) and in clerical occupations, both in services and manufacturing (Goldin and Katz 2008; Michaels 2007). Although software and computers have replaced labor in some white-collar tasks, they have simultaneously created many new tasks. These include tasks related

[1] There are also new tasks in which capital has a comparative advantage (for example, automated detection). Throughout our focus is on "labor-intensive" new tasks, and for brevity, we will simply refer to these as "new tasks."

to programming, design, and maintenance of high tech equipment, such as software and app development, database design and analysis, and computer-security-related tasks, as well as tasks related to more specialized functions in existing occupations, including administrative assistants, analysts for loan applications, and medical equipment technicians (Lin 2011). In Acemoglu and Restrepo (2018a, using data from Lin 2011), we show that about half of employment growth over 1980–2015 took place in occupations in which job titles or tasks performed by workers changed.

Our conceptual framework offers several lessons. First, the presumption that *all* technologies increase (aggregate) labor demand simply because they raise productivity is wrong. Some automation technologies may in fact reduce labor demand because they bring sizable displacement effects but modest productivity gains (especially when substituted workers were cheap to begin with and the automated technology is only marginally better than them). Second, because of the displacement effect, we should not expect automation to create wage increases commensurate with productivity growth. In fact, as we noted already, automation by itself always reduces the labor share in industry value added and tends to reduce the overall labor share in the economy (meaning that it leads to slower wage growth than productivity growth). The reason why we have had rapid wage growth and stable labor shares in the past is a consequence of other technological changes that generated new tasks for labor and counterbalanced the effects of automation on the task content of production. Some technologies displaced labor from automated tasks while others reinstated labor into new tasks. On net, labor retained a key role in production. By the same token, our framework suggests that the future of work depends on the mixture of new technologies and how these change the task content of production.

In the second part of the paper, we use our framework to study the evolution of labor demand in the United States since World War II and explain how industry data can be used to infer the behavior of the task content of production and the displacement and reinstatement effects. We start by showing that there has been a slowdown in the growth of labor demand over the last three decades and an almost complete stagnation over the last two. We establish this by studying the evolution of the economy-wide wage bill, which combines information on average wages and total employment and is thus informative about changes in overall labor demand. We then use industry data to decompose changes in the economy-wide wage bill into productivity, composition and substitution effects, and changes in the task content of production. All technologies create productivity effects that contribute to labor demand. The composition effect arises from the reallocation of activity across sectors with different labor intensities. The substitution effect captures the substitution between labor- and capital-intensive tasks within an industry in response to a change in task prices (for instance, caused by factor-augmenting technologies making labor or capital more productive at tasks they currently perform). We estimate changes in the task content of production from residual changes in industry-level labor shares (beyond what can be explained by substitution effects). We further decompose changes in the task content of production into displacement effects caused by automation and reinstatement effects driven by new tasks.

We provide external support for this decomposition by relating estimated changes in the task content of production to a battery of measures of automation and introduction of new tasks across sectors.

Our decomposition suggests that the evolution of the US wage bill, especially over the last 20 years, cannot be understood without factoring in changes in the task content of production. In particular, we find that the sharp slowdown of US wage bill growth over the last three decades is a consequence of weaker-than-usual productivity growth and significant shifts in the task content of production against labor. By decomposing the change in the task content of production, we estimate stronger displacement effects and considerably weaker reinstatement effects during the last 30 years than the decades before. These patterns hint at an acceleration of automation and a deceleration in the creation of new tasks. They also raise the question of why productivity growth has been so anemic while automation has accelerated during recent years. We use our framework to shed light on this critical question.

An online Appendix available with this paper at the journal website contains a more detailed exposition of our framework, proofs, additional empirical results, and details on the construction of our data.

Conceptual Framework

Production requires the completion of a range of tasks. The production of a shirt, for example, starts with a design, then requires the completion of a variety of production tasks, such as the extraction of fibers, spinning them to produce yarn, weaving, knitting, dyeing, and processing, as well as additional nonproduction tasks, including accounting, marketing, transportation, and sales. Each one of these tasks can be performed by human labor or by capital (including both machines and software). The allocation of tasks to factors determines the task content of production.

Automation enables some of the tasks previously performed by labor to be produced by capital. As a recent example, advances in robotics technologies since the 1980s have allowed firms to automate a wide range of production tasks in manufacturing, such as machining, welding, painting, and assembling, that were performed manually (Ayres and Miller 1983; Groover, Weiss, Nagel, and Odrey 1986; Acemoglu and Restrepo 2018b). The set of tasks involved in producing a product is not constant over time, and the introduction of new tasks can be a major source of labor demand as well as productivity. In textiles, examples of new labor-intensive tasks include computerized designs, new methods of market research, and various managerial activities for better targeting of demand and cost saving. By changing the allocation of tasks to factors, both automation and the introduction of new tasks affect the task content of production.

Tasks are thus the fundamental unit of production, and the factors of production contribute to output by performing these tasks. In contrast, the canonical approach in economics bypasses tasks and directly posits a production function of the form $Y = F\ (A^K K, A^L L)$, which additionally imposes that all technological change takes a factor-augmenting form. There are three related reasons we prefer our conceptual framework. First, the canonical approach lacks descriptive realism. Advances in robotics, for example, do not make capital or labor more productive, but expand the set of tasks that can be produced by capital. Second, capital-augmenting technological change (an increase in A^K) or labor-augmenting technological change (an increase in A^L) corresponds to the relevant factor becoming *uniformly more productive in all tasks*, which, we will show, ignores potentially important changes in the task content of production. Third, and most importantly, we will also see that the quantitative and qualitative implications of factor-augmenting technological advances are different from those of technologies that change the task content of production. Focusing just on factor-augmenting technologies can force us into misleading conclusions.

Tasks and Production

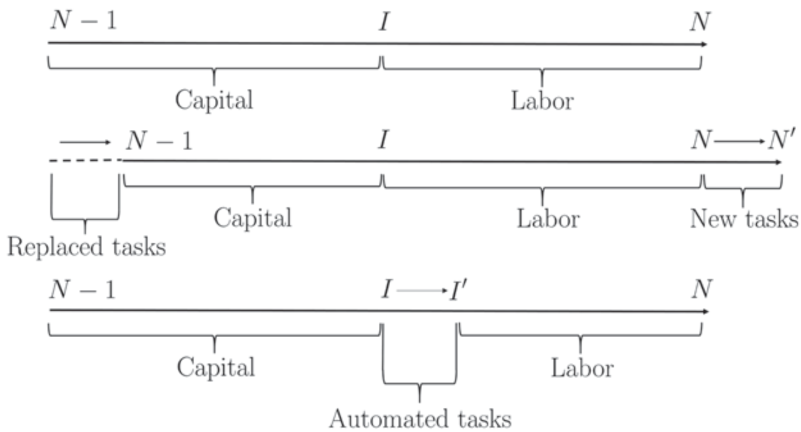
We present our task-based framework by first describing the production process in a single-sector economy.[2] Suppose that production combines the output of a range of tasks, and that the tasks are indexed by z and normalized to lie between $N - 1$ and N, as shown in Figure 1.[3] Tasks can be produced using capital or labor. Tasks with $z > I$ are not automated, and can only be produced with labor, which has a wage rate W. Tasks $z \leq I$ are automated and can be produced with capital, which has a rental rate R, as well as labor. We assume that labor has both a comparative and an absolute advantage in higher indexed tasks. An increase in I therefore represents the introduction of an automation technology, or *automation* for short. An increase in N, on the other hand, corresponds to the introduction of new labor-intensive tasks or *new tasks* for short. In addition to automation (I) and introduction of new tasks (N), the state of technology for this sector depends on A^L (labor-augmenting technology) and A^K (capital-augmenting technology), which increase the productivities of these factors in all tasks.

Let us assume that it is cost-minimizing for firms to use capital in all tasks that are automated (all $z \leq I$) and to adopt all new tasks immediately. This implies an allocation of tasks to factors as summarized in Figure 1, which also shows how automation and new tasks impact this allocation.

[2] This also describes the production process in a sector situated in a multisector economy, with the only difference being that, in that case, changes in technology impact relative prices and induce reallocation of capital and labor across sectors. We discuss these relative price and reallocation effects below.

[3] Namely, the production function takes the form $Y = \left(\int_{N-1}^{N} Y(z)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, where $Y(z)$ is the output of task z. The assumption that tasks lie between $N - 1$ and N is adopted to simplify the exposition. Nothing major changes if we instead allow tasks to lie on the interval between 0 and N. The online Appendix presents more detail on underlying assumptions and on derivations of results that follow throughout the discussion.

Figure 1
The Allocation of Capital and Labor to the Production of Tasks and the Impact of Automation and the Creation of New Tasks



Source: Authors.
Note: The figure summarizes the allocation of tasks to capital and labor. Production requires the completion of a range of tasks, normalized to lie between $N - 1$ and N. Tasks above I are not automated, and can only be produced with labor. Tasks below I are automated and will be produced with capital. An increase in I represents the introduction of automation technology or automation for short. An increase in N corresponds to the introduction of new labor-intensive tasks or new tasks for short.

Following the same steps as in Acemoglu and Restrepo (2018a), output can be represented as a constant elasticity of substitution (CES) function of capital and labor:

$$Y = \Pi(I, N)\left(\Gamma(I, N)^{\frac{1}{\sigma}}(A^L L)^{\frac{\sigma-1}{\sigma}} + (1 - \Gamma(I, N))^{\frac{1}{\sigma}}(A^K K)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}.$$

As in the canonical model, we have production as a function of the quantities of labor and capital, L and K. The labor-augmenting technology term A^L and the capital-augmenting term A^K increase the productivity of labor and capital in all tasks they currently produce. The elasticity of substitution between tasks, σ, determines how easy it is to substitute one task for another, and is also the (derived) elasticity of substitution between capital and labor.

The crucial difference from the canonical model is that the share parameters of this constant-elasticity-of-substitution function depend on automation and new tasks. The share parameter for labor, $\Gamma(I, N)$, is the labor task content of production, which represents the share of tasks performed by labor relative to capital (adjusted for differences in labor and capital productivity across these tasks). Conversely, $1 - \Gamma(I, N)$ is the capital task content of production. Hence, an increase in $\Gamma(I, N)$

shifts the task content of production in favor of labor and against capital. In the special case where $\sigma = 1$, $\Gamma(I, N) = N - I$. More generally, $\Gamma(I, N)$ is always increasing in N and decreasing in I. This, in particular, implies that automation (greater I) shifts the task content of production against labor because it entails capital taking over tasks previously performed by labor. In contrast, new labor-intensive tasks shift the task content of production in favor of labor.[4] Finally, automation and new tasks not only change the task content of production but also generate productivity gains by allowing the allocation of (some) tasks to cheaper factors. The term $\Pi(I, N)$, which shows up as total factor productivity, represents these productivity gains.

The labor share, given by wage bill (WL) divided by value added (Y), can be derived as:

$$s^L = \frac{1}{1 + \frac{1 - \Gamma(I, N)}{\Gamma(I, N)} \left(\frac{R/A^K}{W/A^L} \right)^{1-\sigma}} .$$

This relationship, which will be relied upon extensively in the rest of the paper, clarifies the two distinct forces shaping the labor share (in an industry or the entire economy). As is standard, the labor share depends on the ratio of effective factor prices, W/A^L and R/A^K. Intuitively, as effective wages rise relative to effective rental rates of capital, the price of tasks produced by labor increases relative to the price of tasks produced by capital, and this generates a *substitution effect* across tasks. This is the only force influencing the labor share in the canonical model. Its magnitude and size depend on whether σ is greater than or less than 1. For example, when tasks are complements ($\sigma < 1$), an increase in the effective wage raises the cost share of tasks produced by labor. The opposite happens when $\sigma > 1$. When $\sigma = 1$, we obtain a Cobb–Douglas production function and the substitution effect vanishes because the share of each task in value added is fixed.

More novel are the effects of the task content of production, $\Gamma(I, N)$, on the labor share. Intuitively, as more tasks are allocated to capital instead of labor, the task content shifts against labor and the labor share will decline unambiguously. Our model thus predicts that, independently from the elasticity of substitution σ, automation (which changes the task content of production against labor) will reduce the labor share in the industry, while new tasks (which alter the task content of production in favor of labor) will increase it.

[4] Our exposition assumes that the task content of production does not depend on factor-augmenting technologies or the supply of capital or labor. This will be the case when it is cost-minimizing for firms in this sector to use capital in all tasks that are automated (all $z \leq I$) and use all new tasks immediately. The online Appendix presents the underlying assumptions on technology and factor supplies that ensures this is the case. When this assumption does not hold (for example, because of very large changes in factor-augmenting technologies or factor supplies), the allocation of tasks to factors will change with factor supplies and factor-augmenting technologies. Even in this case, the impact of factor-augmenting technologies on the task content will be small relative to the productivity gains from these technologies.

Technology and Labor Demand

We now investigate how technology changes labor demand. We focus on the behavior of the wage bill, WL, which captures the total amount employers pay for labor. Recall that

$$\text{Wage bill} = \text{Value added} \times \text{Labor share}.$$

Changes in the wage bill will translate into some combination of changes in employment and wages, and the exact division will be affected by the elasticity of labor supply and labor market imperfections, neither of which we model explicitly in this paper (for discussion, see Acemoglu and Restrepo 2018a, 2018b).

We use this relationship to think about how three classes of technologies impact labor demand: automation, new tasks, and factor-augmenting advances. Consider the introduction of new automation technologies (an increase in I in Figure 1). The impact on labor demand can be represented as:

$$\text{Effect of automation on labor demand} = \text{Productivity effect} + \text{Displacement effect}.$$

The *productivity effect* arises from the fact that automation increases value added, and this raises the demand for labor from non-automated tasks. If nothing else happened, labor demand of the industry would increase at the same rate as value added, and the labor share would remain constant. However, automation also generates a *displacement effect*—it displaces labor from the tasks previously allocated to it—which shifts the task content of production against labor and always reduces the labor share. Automation therefore increases the size of the pie, but labor gets a smaller slice. There is no guarantee that the productivity effect is greater than the displacement effect; some automation technologies can reduce labor demand even as they raise productivity.[5]

Hence, contrary to a common presumption in popular debates, it is not the "brilliant" automation technologies that threaten employment and wages, but "so-so technologies" that generate small productivity improvements. This is because the positive productivity effect of so-so technologies is not sufficient to offset the decline in labor demand due to displacement. To understand when this is likely to be the case, let us first consider where the productivity gains from automation are coming from. These are not a consequence of the fact that capital and labor are becoming more productive in the tasks they are performing, but follow from the ability of firms to use cheaper capital in tasks previously performed by labor. The productivity effect of

[5] Indeed, in Acemoglu and Restrepo (2018b), we show that industrial robots, a leading example of automation technology, are associated with lower labor share and labor demand at the industry level and lower labor demand in local labor markets exposed to this technology. This result is consistent with a powerful displacement effect that has dominated the productivity effect from this class of automation technologies.

automation is therefore proportional to cost-savings obtained from such substitution. The greater is the productivity of labor in tasks being automated relative to its wage and the smaller is the productivity of capital in these tasks relative to the rental rate of capital, the more limited the productivity gains from automation will be. Examples of so-so technologies include automated customer service, which has displaced human service representatives but is generally deemed to be low quality and thus unlikely to have generated large productivity gains. They may also include several of the applications of artificial intelligence technology to tasks that are currently challenging for machines.

Different technologies are accompanied by productivity effects of varying magnitudes, and hence, we cannot presume that one set of automation technologies will impact labor demand in the same way as others. Likewise, because the productivity gains of automation depend on the wage, the net impact of automation on labor demand will depend on the broader labor market context. When wages are high and labor is scarce, automation will generate a strong productivity effect and will tend to raise labor demand. When wages are low and labor is abundant, automation will bring modest productivity benefits and could end up reducing labor demand. This observation might explain why automation technologies adopted in response to the scarcity of (middle-aged) production workers in countries where the labor force is aging rapidly, such as Germany, Japan, and South Korea, appear to have more positive effects than in the United States (on cross-country patterns, see Acemoglu and Restrepo 2018e; on the effect of robots in the United States, see Acemoglu and Restrepo 2018b; in Germany, see Dauth, Findeisen, Suedekum, and Woessner 2018). It also suggests a reinterpretation of the famous Habakkuk hypothesis that the faster growth of the 19th-century US economy compared to Britain was due to its relative scarcity of labor (Habakkuk 1962; for a similar argument in the context of the British Industrial Revolution, see also Allen 2009). Labor scarcity encourages automation, and the high wages it causes help explain why this automation process led to rapid productivity and further wage growth.

Consider next the effect of the introduction of new tasks on the wage bill, which is captured by an increase in N in our framework. This expands the set of tasks in which humans have a comparative advantage, and its effect can be summarized as:

$$\text{Effect of new tasks on labor demand} = \text{Productivity effect} \\ + \text{Reinstatement effect.}$$

The *reinstatement effect* captures the change in the task content of production, but now in favor of labor as the increase in N reinstates labor into new tasks. This change in task content always increases the labor share. It also improves productivity as new tasks exploit labor's comparative advantage. The resulting productivity improvement, together with the change in task content, ensures that labor demand always increases following the introduction of new tasks.

Finally, as we claimed previously, the implications of factor-augmenting technologies are very different from those of automation and new tasks, because they do not change the task content of production. In particular,

Effect of factor-augmenting technologies on labor demand = Productivity effect
+ Substitution effect.

With factor-augmenting technological improvements, either labor or capital becomes more productive in all tasks, making the productivity effect proportional to their share in value added.

Factor-augmenting technologies also impact labor demand via the substitution effect introduced above, which changes the labor share but does not alter the task content of production. Available estimates of σ place this parameter to be less than but close to 1, which implies that the substitution effects of factor-augmenting technologies are small relative to their productivity effects.

In summary, in contrast to automation and new tasks that can generate significant displacement and reinstatement effects, factor-augmenting technologies affect labor demand mostly via the productivity effect and have a relatively small impact on the labor share. As a result, they are unlikely to generate a lower labor demand from technological advances: capital-augmenting technologies always increase labor demand, and labor-augmenting technologies do the same for plausible parameter values, in particular, so long as $\sigma > 1 - s^L$ (Acemoglu and Restrepo 2018c).[6]

Tasks, Production, and Aggregate Labor Demand

We now embed the model of tasks and production in an economy with multiple industries and investigate how technology changes aggregate labor demand by characterizing the behavior of the (economy-wide) wage bill. In our multisector economy we have:

$$\text{Wage bill} = \text{GDP} \times \sum_{i \in \mathcal{I}} \text{Labor share sector } i \times \text{Share of value added in sector } i.$$

The multisector perspective offers an additional margin for adjustment in response to automation, which we refer to as the *composition effect*. Following automation in sector i (an increase in I for that sector) we have:

Effect of automation in i on aggregate labor demand = Productivity effect
+ Displacement effect
+ Composition effect.

[6] Many other technologies share the feature that they do not impact the task content of production. For example, improvements in the quality or productivity of equipment in any subset of already-automated tasks in $(N - 1, I)$ (what, in Acemoglu and Restrepo 2018d, we call a "deepening of automation") will have an impact on labor demand identical to capital-augmenting technologies. These technologies do not change the allocation of tasks to factors (as a new piece of equipment is replacing an older one), and so they affect labor demand mostly through the productivity effect.

The first two effects are the same as above—the productivity effect represents the impact of automation in sector i on GDP, while the displacement effect represents the change in the task content of production sector i (which affects the labor share within this sector). These effects are scaled by the size of sector i, since larger sectors will have larger aggregate effects.

The composition effect, which was absent when we were focusing on the effect of automation in a one-sector economy, captures the implications of sectoral reallocations (changes in the share of value-added across sectors). For example, automation in sector i may reallocate economic activity towards sector j (depending on demand elasticities and input-output linkages). This reallocation contributes positively to aggregate labor demand when sector j has higher labor share than the contracting sector i, and negatively when the opposite holds.

A similar decomposition applies to new tasks. Following the introduction of new tasks in sector i (an increase in N for that sector), we have:

Effect of new tasks in i on aggregate labor demand = Productivity effect
+ Reinstatement effect
+ Composition effect,

where the new feature is again the composition effect.

The mechanization of agriculture in the United States illustrates how these forces jointly determine the behavior of aggregate labor demand. Data from Budd (1960) show that between 1850 and 1910, the replacement of manual labor by horse-powered reapers and harvesters in agriculture coincided with a sharp decline in the labor share of value in this sector, from 33 to 17 percent—a telltale sign of the displacement effect created by mechanization. Meanwhile, despite rapid mechanization of agriculture, at the time making up one-third of the US economy, two forces combined to generate an increase in aggregate labor demand. First, and in part as a consequence of mechanization, value-added and employment were reallocated from agriculture to the industrial sector. This created a powerful composition effect, as industry was (and still continues to be) much more labor intensive than agriculture. In addition, the labor share within the industrial sector rose further during this process, from 47 percent in 1850 to 55 percent by 1890. This change in industry labor share signals the presence of a powerful reinstatement effect created by the introduction of new labor-intensive jobs in this sector. This interpretation is consistent with significant growth in new factory jobs in farm equipment (Olmstead and Rhode 2001), cotton milling (Rasmussen 1982), and subsequently clerical occupations in trade and manufacturing industries (Goldin and Katz 2008; Michaels 2007).

Finally, the effects of factor-augmenting technologies in a multi-industry context can be analyzed similarly. Although they too generate composition effects and may affect aggregate labor demand via this channel, factor-augmenting technologies still have no impact on the task content of production. Absent powerful composition effects, they continue to affect labor demand mostly via their productivity effect.

Sources of Labor Demand Growth in the United States

We now use our framework to shed light on the factors that have shaped the evolution of US labor demand since World War II. To do this, we develop a decomposition of observed changes in the total wage bill in the economy. Our decomposition requires data on industry value added, factor payments, and labor shares. The change in aggregate wage bill between two periods can be decomposed (as we show in the online Appendix) as:

Change in aggregate wage bill = Productivity effect + Composition effect
+ Substitution effect + Change in task content.

The productivity effect is the sum of the contributions from various sources of technology to value added and thus GDP. Correspondingly, in our empirical exercise we measure this effect using changes in (log) GDP per capita.

The composition effect captures changes in labor demand resulting from reallocation of value added across sectors. As discussed in the previous section, this is related to the gap between the labor share of contracting and expanding sectors. In our empirical exercise, we measure it as the sum of the change in the value-added share of an industry weighted by its labor share (if all sectors had the same labor share, this term would be equal to zero). The composition effect includes not only the sectoral reallocation brought by new technologies but also changes in value added across sectors resulting from structural transformations and sectoral reallocation due to preferences (for example, Herrendorf, Rogerson, and Valentinyi 2013; Hubmer 2018; Aghion, Jones, and Jones 2017), differences in factor intensities (for example, Acemoglu and Guerrieri 2008), differential sectoral productivity growth (for example, Ngai and Pissarides 2007), or international trade in final goods (for example, Autor, Dorn, and Hanson 2013).

The substitution effect is an employment-weighted sum of the substitution effects of industries, and thus depends on industry-level changes in effective factor prices and the elasticity of substitution σ (as shown in the earlier expression for the labor share). To estimate the substitution effect in an industry, we choose as our baseline Oberfield and Raval's (2014) estimate of the elasticity of substitution between capital and labor, $\sigma = 0.8$.[7] In addition, we utilize information on sectoral factor prices from the Bureau of Economic Analysis, Bureau of Labor Statistics, and the national income and product accounts. To convert observed factor prices into effective ones, we start with a benchmark where A_i^L/A_i^K grows at a common rate equal to average labor productivity, which we take to be 2 percent a year between

[7] We show in the online Appendix that the results are very similar for reasonable variations in σ. Note also that the relevant σ is the elasticity of substitution between capital and labor at the industry level. This is greater than the firm-level elasticity, estimated to be between 0.4 and 0.7 (for example, Chirinko, Fazzari, and Meyer 2011) because of output substitution between firms. Note also that our framework, in particular the central role of changes in the task content of production, makes it clear that this elasticity of substitution cannot be estimated from aggregate data.

1947 and 1987 and 1.46 percent a year between 1987 and 2017. The motivation for this choice is that, if all technological progress were labor-augmenting, this would be the rate of growth in A_i^L required to match the behavior of labor productivity.[8]

The change in task content is given by an employment-weighted sum of the changes in task content of production of industries. We estimate industry-level change in task content as the residual change in labor share (observed directly in the data) that cannot be explained by the substitution effect. Namely,

$$\text{Change in task content in } i = \text{Percent change in labor share in } i - \text{Substitution effect in } i.$$

Intuitively, with competitive factor and product markets, the change in task content of production and the substitution effect are the only forces affecting the labor share of an industry. Hence, changes in task content can be inferred once we have estimates of the substitution effect.

Under additional assumptions, we can also separate the change in task content into its two components: the displacement and reinstatement effects. Assume that an industry will not simultaneously undertake automation and introduce new tasks (this is implied, for example, by the directed technological change reasoning in Acemoglu and Restrepo 2018a, where depending on factor prices, an industry will engage in one type of innovation or the other). Then, when the labor share of an industry declines beyond what one would expect based on factor prices, we estimate a positive displacement effect resulting from automation in that industry. Conversely, when the labor share in an industry rises beyond what one would expect based on factor prices, we estimate a positive reinstatement effect, attributed in our model to the introduction of new tasks. Motivated by this reasoning, we compute the displacement effect as the five-year moving average of the change in task content for industries with a negative change, and the reinstatement effect as the five-year moving average of the change in task content for industries with a positive change. The five-year time window is chosen to minimize the influence of measurement error in industry labor shares. To the extent that there are simultaneous introduction of new automation technologies and new tasks in a given industry within a five-year period, our estimates will be lower bounds both for the displacement and reinstatement effects.

Sources of Labor Demand: 1947–1987

We first apply this decomposition to data from the four decades following World War II, from 1947 to 1987. For this period, we have data from the Bureau of

[8] Our estimates for the growth rate of A_i^L/A_i^K should be interpreted as upper bounds, since in general growth in GDP per worker will be driven not just by labor-augmenting technological changes. Because in our main exercise $\sigma < 1$, this implies that we are also understating the importance of displacement effects in reducing the task content of production. Nevertheless, reasonable variations on the growth rate of A_i^L/A_i^K have small impacts on our decomposition results, as we show in the online Appendix.

Economic Analysis for 58 industries on value added and labor shares.[9] We combine these with data from the national income and product accounts on quantities of capital and labor in each industry to obtain measures of factor prices. We consolidate the data into 43 industries that covered the private sector and can be tracked consistently over time and across sources.

Figure 2 presents the evolution of the labor share for six broad sectors: construction, services, transportation, manufacturing, agriculture, and mining. Except for mining and transportation—two small sectors accounting for 10 percent of GDP—there are no significant declines in labor shares in these broad sectors in this time period. In fact, the labor share in manufacturing and services increased modestly during this period. The bottom panel of the figure shows the evolution of the share of value added of these sectors and confirms the secular reallocation from manufacturing towards services starting in the late 1950s.

Figure 3 presents our decomposition using the 43 industries in our sample. We have divided the wage bill by population, so that changes in population do not confound the effects we are focusing on. The top panel in Figure 3 shows that wage bill per capita grew at 2.5 percent per year during this period. The rapid and steady growth of the wage bill during this period is largely explained by the productivity effect (2.4 percent per year). The substitution and composition effects are small, and during this period changes in the task content of production are small as well.

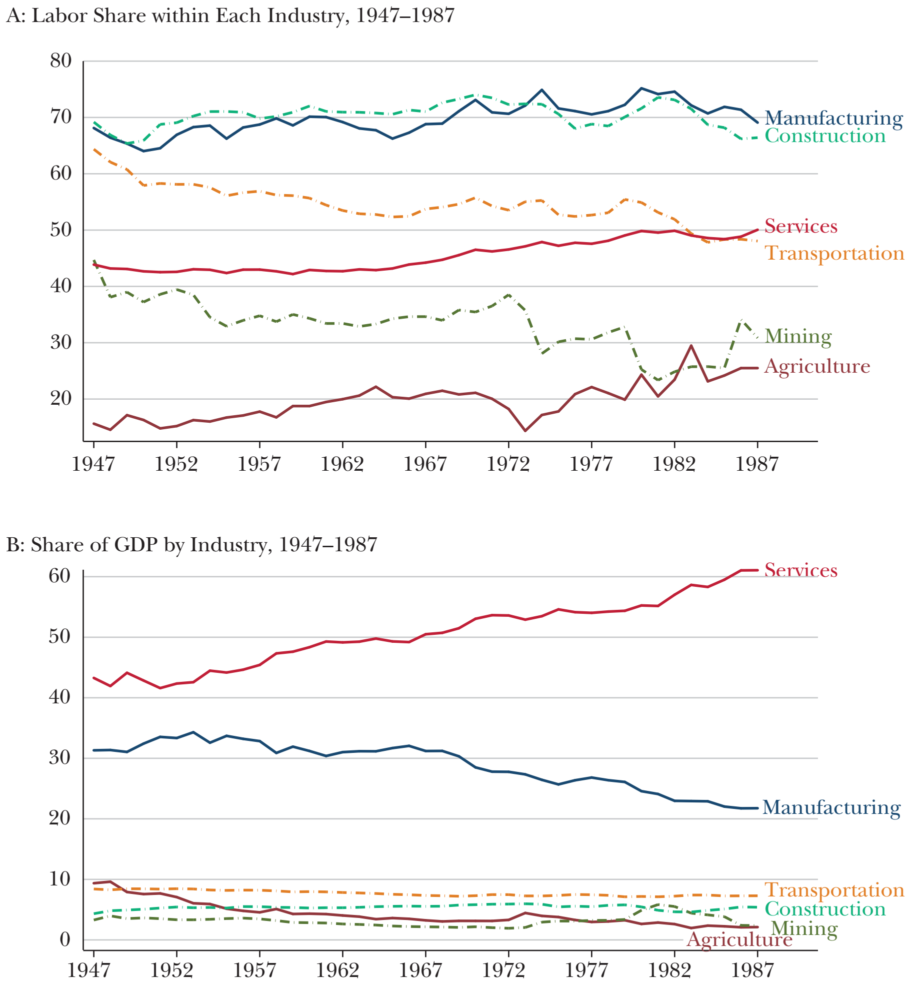
The middle panel of Figure 3 shows that, even though the overall change in the task content of production during this period is small, there is considerable displacement and reinstatement. Between 1947 and 1987, the displacement effect reduced labor demand at about 0.48 percent per year, but simultaneously, there was an equally strong reinstatement effect, equivalent to an increase in labor demand of 0.47 percent per year. The bottom panel of Figure 3 depicts a similar pattern in manufacturing, where the overall change in task content was also small, while displacement and reinstatement effects were substantial. In sum, our findings suggest that during the four decades following World War II there was plenty of automation, but this was accompanied by the introduction of new tasks (or other changes increasing the task content of production in favor of labor) in both manufacturing and the rest of the economy that counterbalanced the adverse labor demand consequences of automation.

Sources of Labor Demand: 1987–2017

For the 1987–2017 period, we use data from the Bureau of Economic Analysis for 61 industries covering the private sector and complement them with data from

[9] Our measure of labor demand is given by the wage bill in the private sector and thus excludes self-employment income. This avoids the need for apportioning self-employment income between labor and capital. Elsby, Hobijn, and Sahin (2013) explore this issue in detail and conclude that labor income from self-employment has either declined or remained constant as a share of total labor income over this period. This implies that labor share inclusive of self-employment income likely declined by even more, and thus, if anything, focusing on the labor share in the private sector understates the overall decline in labor demand.

Figure 2
The Labor Share and Sectoral Evolutions, 1947–1987



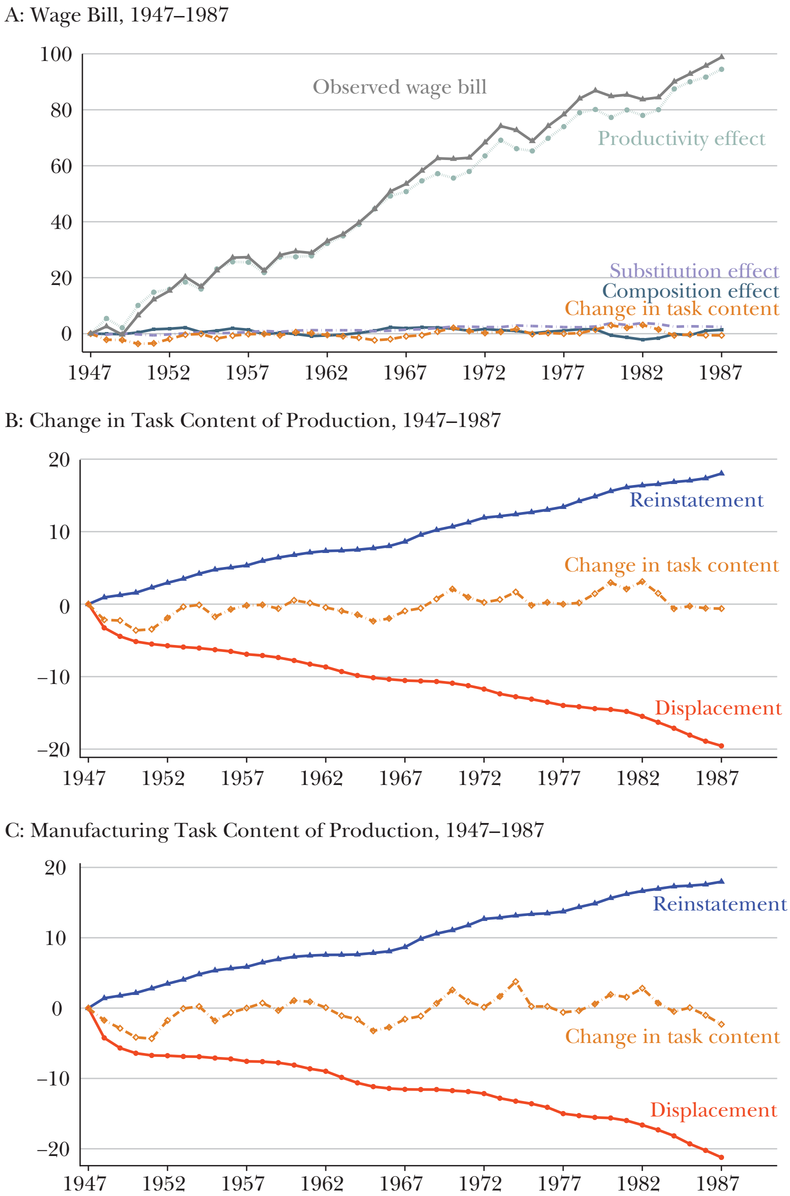
Source: Authors using data from the US Bureau of Economic Analysis industry accounts.
Note: The top panel shows the labor share in value added in services, manufacturing, construction, transportation, mining and agriculture between 1947 and 1987, while the bottom panel shows the share of value added in these sectors relative to GDP.

the Bureau of Labor Statistics on factor prices. The top panel of Figure 4 presents the evolution of the labor share for the same six broad sectors used above. In contrast to the 1947–1987 period, there is a sizable decline in the labor share in manufacturing and construction. The drop in the labor share for mining continues at a similar pace. The bottom panel of the figure shows the continued reallocation of economic activity from manufacturing to services.

The top panel of Figure 5 shows a striking slowdown in the growth of labor demand between 1987 and 2017. The wage bill per capita grew at a modest

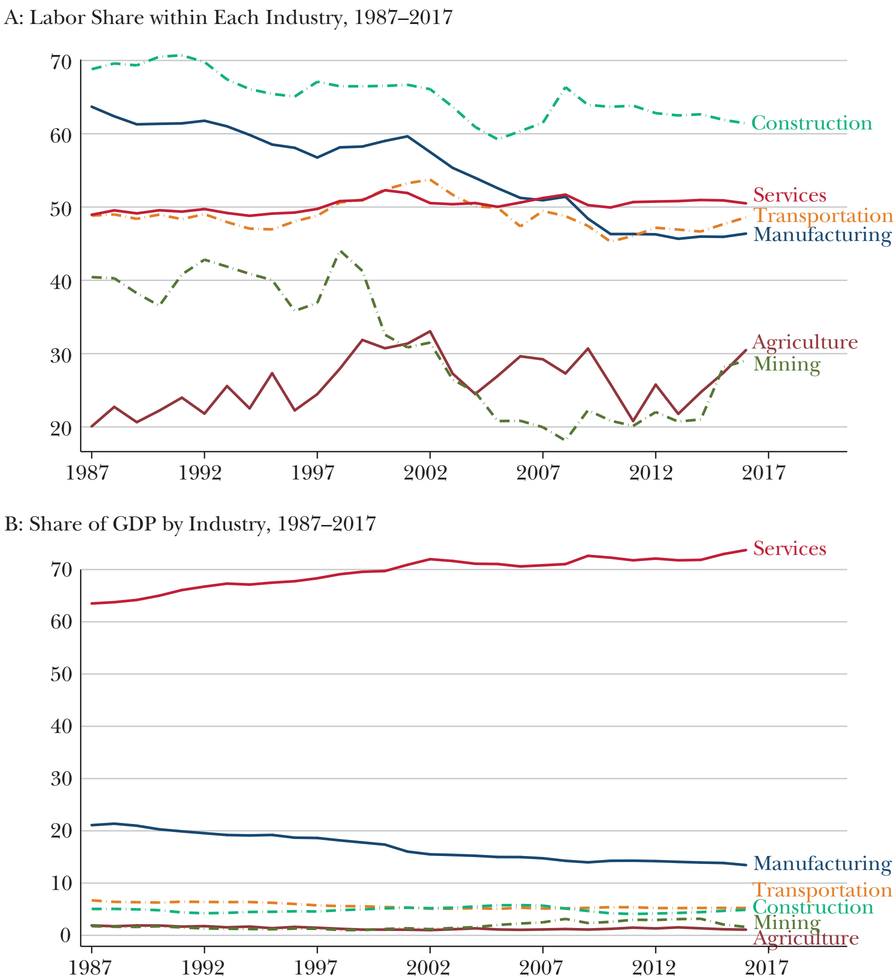
Figure 3

Sources of Changes in Labor Demand, 1947–1987



Source: Authors' calculations.
Note: The top panel presents the decomposition of the wage bill divided by population between 1947 and 1987. The middle and bottom panels present our estimates of the displacement and reinstatement effects for the entire economy and the manufacturing sector, respectively. See text for the details of the estimation of the changes in task content and displacement and reinstatement effects.

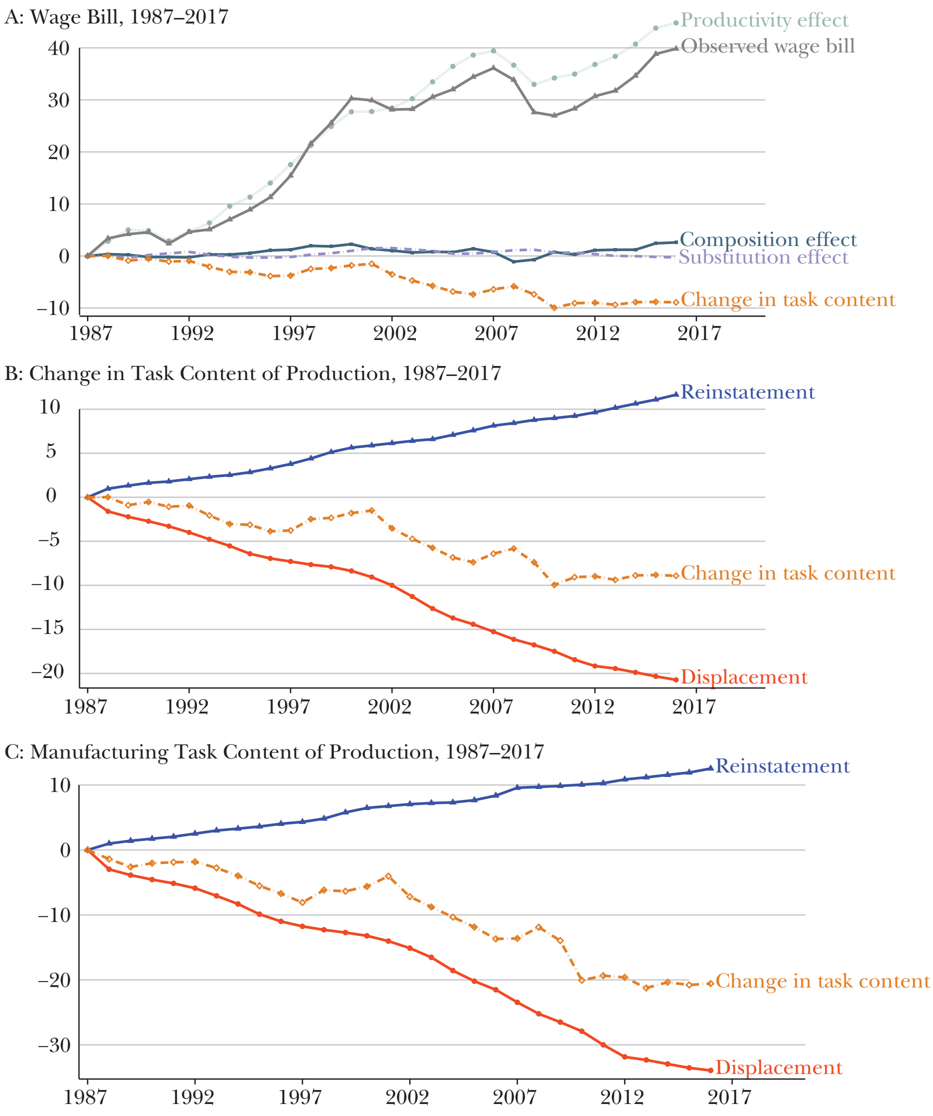
Figure 4
The Labor Share and Sectoral Evolutions, 1987–2017



Source: Authors using data from the US Bureau of Economic Analysis industry accounts and the Bureau of Labor Statistics.
Note: The top panel shows the labor share in value added in services, manufacturing, construction, transportation, mining, and agriculture between 1987 and 2017, while the bottom panel shows the share of value added in these sectors relative to GDP.

1.33 percent per year during the entire period and essentially stagnated since 2000. The first factor accounting for the deceleration of labor demand during this period is the slowdown of productivity growth (1.54 percent per year compared to 2.4 percent in 1947–1987). The second factor contributing to slower wage bill growth, especially after the late 1990s, is a significant negative shift in the task content of production against labor (of 0.35 percent per year), which caused labor

Figure 5
Sources of Changes in Labor Demand, 1987–2017



Source: Authors' calculations.
Note: The top panel presents the decomposition of wage bill divided by population between 1987 and 2017. The middle and bottom panels present our estimates of the displacement and reinstatement effects for the entire economy and the manufacturing sector, respectively. See text for the details of the estimation of the changes in task content and displacement and reinstatement effects.

demand to decouple from productivity. Cumulatively, changes in the task content of production reduced labor demand by 10 percent during this period.

The middle and bottom panels of Figure 5 show that, relative to the earlier period, the change in task content is driven by a deceleration in the introduction of technologies reinstating labor (reinstatement increased labor demand only by 0.35 percent per year compared to 0.47 percent in 1947–1987) and an acceleration of displacement (displacement reduced labor demand by 0.7 percent per year compared to 0.48 percent in 1947–1987). This pattern is particularly pronounced in manufacturing, where the displacement effect reduced labor demand at about 1.1 percent per year or about 30 percent cumulatively. These results are consistent with Elsby, Hobijn, and Sahin (2013), who document the important role of within-industry changes that are uncorrelated with factor prices in accounting for the aggregate behavior of the labor share. The change in the balance between displacement and reinstatement also corroborates the findings of Autor and Salomons (2018), who find that technological improvements after 1980 have been associated with declines in the labor share while those in the previous decades have not been.

Finally, the top panel also shows that the composition and substitution effects had a very limited impact on the wage bill. Although there is a sizable shift away from manufacturing, which is itself not unrelated to automation in this sector as well as to import competition, the resulting composition effects are small because the labor share in manufacturing is similar to that in the expanding service industries (see the top panel of Figure 4). These findings highlight that unlike the 19th-century mechanization of agriculture, there are no powerful composition effects contributing to labor demand. Even more importantly, there appears to be no equivalent of the powerful reinstatement effects that accompanied the mechanization of agriculture.

In summary, the deceleration of labor demand growth over the last 30 years is due to a combination of anemic productivity growth and adverse shifts in the task contents of production owing to rapid automation that is not being counterbalanced by the creation of new tasks.[10]

What Does the Change in Task Content Capture?

A natural concern is that our estimates of the change in task content capture something different than what might commonly be understood as displacement effects from automation technologies and reinstatement effects of new tasks. Here, we provide additional evidence that our estimates are informative about changes in the task content of production. We focus on the 1987–2017 period where we have measures of automation and can compute proxies for new tasks at the industry

[10] In the online Appendix, we verify that this pattern is robust to different values of the elasticity of substitution and to reasonable variations in the rates of factor-augmenting technological changes. Furthermore, we computed the changes in factor-augmenting technologies at the industry level that would be necessary to explain changes in industry labor shares without any change in task content of production. We found that this would require gargantuan changes in factor-augmenting technologies and productivity increases—several multiples larger than the observed increases in total factor productivity during the last seven decades. This exercise underscores the need for major changes in the task content of production to account for the evolution of sectoral labor shares and the wage bill. We also demonstrate in the online Appendix that the order in which the decomposition is carried out (composition effects first and within-industry changes next) does not matter for the results.

level, and then document the correlation between these measures and our estimates of the change in the task content of production.

We have three measures of industry-level automation technologies. The proxies are: 1) the adjusted penetration of robots measure from Acemoglu and Restrepo (2018b) for 19 industries, which are then mapped to our 61 industries; 2) the share of routine jobs in an industry in 1990, where we define routine jobs in an occupation as in Acemoglu and Autor (2011) and then project these across industries according to the share of the relevant occupation in the employment of the industry in 1990 (see also vom Lehn 2018); and 3) the share of firms (weighted by employment) across 148 detailed manufacturing industries using automation technologies, which include automatic guided vehicles, automatic storage and retrieval systems, sensors on machinery, computer-controlled machinery, programmable controllers, and industrial robots.[11]

Table 1 reports the estimates of the relationship between the change in task content of production between 1987 and 2017 and the proxies for automation technologies and new tasks; each row and column corresponds to a different regression model. The table shows that with all these proxies there is the expected negative relationship between higher levels of automation and our measure of changes in the task content of production in favor of labor (see also visual representations of these relationships in the online Appendix). These negative relationships remain very similar when we add various control variables, including, in column 1, a dummy for the manufacturing sector and, in column 2, imports from China (the growth of final goods imports from China as in Autor, Dorn, and Hanson 2013; Acemoglu, Autor, Dorn, Hanson, and Price 2016) and a measure of offshoring of intermediate goods (Feenstra and Hanson 1999; Wright 2014). Consistent with our conceptual framework, changes in task content are unrelated to imports of final goods from China, but are correlated with offshoring, which often involves the offshoring of labor-intensive tasks (Elsby, Hobijn, and Sahin 2013). Controlling for offshoring does not change the relationship we report in Table 1 because offshoring is affecting a different set of industries than our measures of automation (see the online Appendix).

We also looked at a series of proxies for the introduction of new tasks across industries, and how they are correlated with our measure of the change in task content for 1987–2017. Our four proxies for new tasks are: 1) the 1990 share of employment in occupations with a large fraction of new job titles, according to the 1991 *Dictionary of Occupational Titles* compiled by Lin (2011); 2) the 1990 share of employment in occupations with a large number of "emerging tasks" according to O*NET, which correspond to tasks that workers identify as becoming increasingly

[11] These data are from the Survey of Manufacturing Technologies, and are available in 1988 and 1993 for 148 four-digit SIC industries which are all part of the following three-digit manufacturing sectors: fabricated metal products; nonelectrical machinery, electric and electronic equipment; transportation equipment; and instruments and related products (Doms, Dunne, and Troske 1997). For this exercise, we computed measures for the change in task content of these four-digit manufacturing industries using detailed data from the Bureau of Economic Analysis input-output tables for 1987 to 2007.

Table 1

Relationship between Change in Task Content of Production and Proxies for Automation and New Tasks

	Raw data (1)	*Controlling for manufacturing* (2)	*Controlling for Chinese import and offshoring* (3)
Proxies for automation technologies:			
Adjusted penetration of robots, 1993–2014	–1.404	–0.985	–1.129
	(0.377)	(0.369)	(0.362)
Observations	61	61	61
R^2	0.18	0.21	0.27
Share of routine jobs in industry, 1990	–0.394	–0.241	–0.321
	(0.122)	(0.159)	(0.164)
Observations	61	61	61
R^2	0.14	0.19	0.27
Share of firms using automation technologies, 1988–1993 (SMT data)	–0.390		–0.397
	(0.165)		(0.166)
Observations	148		148
R^2	0.08		0.09
Proxies for new tasks:			
Share of new job titles, based on 1991 DOT* and 1990 employment by occupation	1.609	1.336	1.602
	(0.523)	(0.530)	(0.541)
Observations	61	61	61
R^2	0.12	0.23	0.32
Number of emerging tasks, based on 1990 employment by occupation	8.423	7.108	7.728
	(2.261)	(2.366)	(2.418)
Observations	61	61	61
R^2	0.14	0.25	0.33
Share of employment growth between 1990 and 2016 in new occupations	2.121	1.638	1.646
	(0.723)	(0.669)	(0.679)
Observations	61	61	61
R^2	0.08	0.20	0.26
Percent increase in number of occupations represented in industry	0.585	0.368	0.351
	(0.156)	(0.207)	(0.215)
Observations	61	61	61
R^2	0.14	0.19	0.25

Source: Authors.
Note: The table reports estimates of the relationship between the change in task content of production between 1987 and 2017 and proxies for automation technologies and new tasks. Each row and column corresponds to a different regression model. Column 1 reports estimates of the bivariate relationship between change in task content of production and the indicated proxy at the industry level. Column 2 includes a dummy for manufacturing industries as a control. In addition, Column 3 controls for the increase in Chinese imports (defined as the increase in imports relative to US consumption between 1991 and 2011, as in Acemoglu et al. 2016) and the increase in offshoring (defined as the increase in the share of imported intermediates between 1993 and 2007, as in Feenstra and Hanson 1999). Except for the third row, which uses the Survey of Manufacturing Technologies (SMT), all regressions are for the 61 industries used in our analysis of the 1987–2017 period. When using the SMT, the regressions are for 148 detailed manufacturing industries. Standard errors robust against heteroskedasticity are in parenthesis. When using the measure of robot penetration, we cluster standard errors at the 19 industries for which this measure is available.
*The DOT is the *Dictionary of Occupational Titles.*

important in their jobs; 3) the share of employment growth in an industry accounted for by "new occupations," defined as occupations that were not present in that industry in 1990 but are present in 2016; and 4) the percent increase in the number of occupations in an industry between 1990 and 2016. The first two measures are projected onto industries using the share of these occupations in industry employment in 1990. All four of these measures are meant to capture major changes in the types of activities performed in occupations (then mapped to industries) or the introduction of certain new activities into an industry. We thus expect the correlations between these proxies for new tasks and our measure of changes in task content in favor of labor to be positive and significant, and they are. These results hold regardless of whether or not we include additional controls in columns 2 and 3 of Table 1.

These correlations bolster the interpretation that our estimates of changes in task content of production contain valuable information on displacement from automation technologies and reinstatement from the introduction of new tasks.

Confounding Factors

Our approach has been predicated on competitive markets and has also abstracted from various other changes potentially affecting US labor markets. We now briefly discuss these issues.

First, as we have already noted, trade in final goods should have no impact on our estimates of the change in the task content of production (because they will affect prices and sales, which are captured by our productivity effect, and they induce sectoral reallocations, which are part of our composition effects). This is confirmed by our results in Table 1. Offshoring, on the other hand, will directly change the task content of production because it involves the replacement of some labor-intensive tasks by services from abroad (Grossman and Rossi-Hansberg 2008). Our estimates in Table 1 are consistent with this, but also show that offshoring does not change the quantitative or qualitative relationship between various measures of automation and our estimates of the change in the task content of production.

Second, as also noted above, sectoral reallocations resulting from structural transformation do not affect the task content of production either and are part of our composition effects. The fact that these composition effects are small suggests that these sectoral reallocations have not been a major factor in the slowdown in labor demand and changes in labor share in national income.

Third, we have abstracted from the presence of workers with different skills, and thus a potential question is whether changes in the skill composition of the workforce would affect our estimates of the change in the task content of production. The answer is "no," provided that industry-level factor payments are well-measured. Hence, as long as the increase in the wage bill caused by skill upgrading in a sector is factored in, this compositional change does not cause a shift in the task content of production. An implication is that secular changes such as population aging and increased female labor force participation, though they will affect the composition of the workforce and factor prices, should not confound our estimates of changes in task content of production.

Fourth, changes in factor supplies should also have no impact provided that our estimates of the substitution effect (which form the basis of our estimates of the change in the task content of production) remain accurate.

In contrast to these factors, deviations from competitive labor or product markets would potentially confound our estimates of task content. Particularly worth noting are deviations from competitive labor markets. If the supply side of the market is determined by bargaining or other rent-sharing arrangements, then our approach still remains valid provided that firms are on their labor demand curve (for overall labor or for different types of labor in the presence of heterogeneity). This is because our analysis only uses information from the labor demand side, so whether workers are along a well-defined labor supply curve is not important. On the other hand, changes in the extent of monopsony and bilateral bargaining and holdup problems forcing firms off their labor demand curve would potentially confound our estimates. A similar confounding would result if there are changing product market markups. Though these issues are important, they are beyond the scope of the current paper and are some of the issues we are investigating in ongoing work.

What Explains the Changing Nature of Technology and Slow Productivity Growth Since 1987?

Our results suggest that it is the combination of adverse shifts in the task content of production—driven by accelerated automation and decelerating reinstatement—and weak productivity growth that accounts for the sluggish growth of labor demand over the last three decades and especially since 2000. Why has the balance between automation and new tasks changed recently? Why has productivity growth been so disappointing despite the acceleration in automation technologies? Though we do not have complete answers to these questions, our conceptual framework points to a number of ideas worth considering.

There are two basic reasons why the balance between automation and new tasks may have changed. First, the innovation possibilities frontier linking these two types of technological change may have shifted, facilitating further automation and making the creation of new tasks more difficult (for a formal analysis, see Acemoglu and Restrepo 2018a). For example, new general-purpose technologies based on advances in hardware and software may have made further automation cheaper, or we may have run out of ideas for generating new high-productivity (labor-intensive) tasks. We find a second reason for a change in this balance more plausible: that is, the US economy may have moved along a given innovation possibilities frontier because incentives for automation have increased and those for creating new tasks have declined. Several factors may push in this direction. The US tax code aggressively subsidizes the use of equipment (for example, via various tax credits and accelerated amortization) and taxes the employment of labor (for example, via payroll taxes). A tendency towards further (and potentially excessive) automation may have been reinforced by the growing focus on automation and use of artificial intelligence for removing the human element from most of the production process. This focus has recently been boosted both by the central role that large tech companies have come to play in innovation with their business model based on automation and small workforces, and by the vision of many of the luminaries of the tech world (think of the efforts of Tesla to automate production extensively, which turned out to be very costly). Finally, the declining government support for innovation may have also contributed by discouraging research with longer horizons, which likely further disadvantaged the creation of new tasks (which bear fruit more slowly) relative to automation.

This list of factors may contribute not just to the changing balance between automation and new tasks, but also to the slowdown in productivity growth. First, because new tasks contribute to productivity, slower reinstatement will be associated with slower productivity growth. Therefore, factors tilting the balance against new tasks likely translate into lost opportunities for improved productivity. In addition, slower wage growth resulting from a weak reinstatement effect indirectly makes automation less productive—because productivity gains from automation are increasing in the effective wage in tasks being replaced, and lower wages thus reduce these productivity gains. Second, if innovations in both automation and new tasks are subject to diminishing returns (within a given period of time or over time), a significant change in the balance between these two types of new technologies will push us towards more marginal developments and cause slower productivity growth. Third, as we emphasized earlier, productivity gains from automation could be quite small for so-so technologies—when automation substitutes for tasks in which labor was already productive and capital is not yet very effective. In this light, further automation, especially when it is induced by tax distortions or excessive enthusiasm about automating everything, would take the form of such so-so technologies and would not bring much in productivity gains. Finally, in Acemoglu and Restrepo (2018d), we suggest there may be a mismatch between the available skills of the workforce and the needs for new technologies. This could further reduce productivity gains from automation and hamper the introduction of new tasks, because the lack of requisite skills reduces the efficiency with which new tasks can be utilized.

If the balance between automation and new tasks has shifted inefficiently and if indeed this is contributing to rapid automation, the absence of powerful reinstatement effects, and the slowdown of productivity growth, then there may be room for policy interventions to improve both job creation and productivity growth. These interventions might include removing incentives for excessive automation (such as the preferential treatment of capital equipment) and implementing new policies designed to rebalance the direction of technological change (for a more detailed discussion in the context of artificial intelligence, see Acemoglu and Restrepo 2019).

Concluding Remarks

This paper develops a task-based model to study the effects of different technologies on labor demand. At the center of our framework is the task content of production—measuring the allocation of tasks to factors of production. Automation, by creating a displacement effect, shifts the task content of production against labor, while the introduction of new tasks in which labor has a comparative advantage improves it via the reinstatement effect. These technologies are qualitatively different from factor-augmenting ones, which do not impact the task content of production. For example, automation always reduces the labor share and may reduce labor demand, and new tasks always increase the labor share.

We then show how changes in the task content of production and other contributors to labor demand can be inferred from data on labor shares, value added, and factor prices at the industry level. The main implication of our empirical exercise using this methodology is that the recent stagnation of labor demand is explained by an acceleration of automation, particularly in manufacturing, and a deceleration in the creation of new tasks. In addition, and perhaps reflecting this shift in the composition of technological advances, the economy also experienced a marked slowdown in productivity growth, contributing to sluggish labor demand.

Our framework has clear implications for the future of work, too. Our evidence and conceptual approach support neither the claims that the end of human work is imminent nor the presumption that technological change will always and everywhere be favorable to labor. Rather, they suggest that if the origin of productivity growth in the future continues to be automation, the relative standing of labor, together with the task content of production, will decline. The creation of new tasks and other technologies raising the labor intensity of production and the labor share are vital for continued wage growth commensurate with productivity growth. Whether such technologies will be forthcoming depends not just on our innovation capabilities but also on the supply of different skills, demographic changes, labor market institutions, government policies including taxes and research and development spending, market competition, corporate strategies, and the ecosystem of innovative clusters. We have pointed out some reasons why the balance between automation and new tasks may have become inefficiently tilted in favor of the former—with potentially adverse implications for jobs and productivity—and some directions for policy interventions to redress this imbalance.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Chap. 12 in *Handbook of Labor Economics,* vol. 4, Part B, edited by David Card and Orley Ashenfelter. Elsevier.

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2016. "Import Competition and the Great US Employment Sag of the 2000s." *Journal of Labor Economics* 34(S1): S141–S198.

Acemoglu, Daron, and Veronica Guerrieri. 2008. "Capital Deepening and Nonbalanced Economic Growth." *Journal of Political Economy* 116(3): 476–98.

Acemoglu, Daron, and Pascual Restrepo. 2018a. "The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108(6): 1488–1542.

Acemoglu, Daron, and Pascual Restrepo. 2018b. "Robots and Jobs: Evidence from US Labor Markets." NBER Working Paper 23285.

Acemoglu, Daron, and Pascual Restrepo. 2018c. "Modeling Automation." *AEA Papers and Proceedings,* 108: 48–53.

Acemoglu, Daron, and Pascual Restrepo. 2018d. "Artificial Intelligence, Automation and Work." NBER Working Paper 24196.

Acemoglu, Daron, and Pascual Restrepo. 2018e. "Demographics and Automation." NBER Working Paper 24421.

Acemoglu, Daron, and Pascual Restrepo. 2019. "The Wrong Kind of AI? Artificial Intelligence and the Future of Labor Demand." NBER Working Paper 25682.

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones. 2017. "Artificial Intelligence and Economic Growth." NBER Working Paper 23928.

Allen, Robert C. 2009. *The British Industrial Revolution in Global Perspective.* Cambridge University Press.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118(4): 1279–1333.

Autor, David, and Anna Salomons. 2018. "Is Automation Labor Share–Displacing? Productivity Growth, Employment, and the Labor Share." *Brooking Papers on Economic Activity*, no. 1, pp. 1–87.

Ayres, Robert U., and Steven M. Miller. 1983. *Robotics: Applications and Social Implications.* Ballinger Publishing Company.

Budd, Edward C. 1960. "Factor Shares, 1850–1910." Chap. 9 in *Trends in the American Economy in the Nineteenth Century*. Princeton University Press.

Chandler, Alfred D. 1977. *The Visible Hand: The Managerial Revolution in American Business.* Harvard University Press.

Chirinko, Robert C., Steven M. Fazzari, and Andrew P. Meyer. 2011. "A New Approach to Estimating Production Function Parameters: The Elusive Capital–Labor Substitution Elasticity." *Journal of Business and Economic Statistics* 29(4): 587–94.

Dauth, Wolfgang, Sebastian Findeisen, Jens Suedekum, and Nicole Woessner. 2018. "Adjusting to Robots: Worker-Level Evidence." Institute Working Paper 13, Opportunity and Inclusive Growth Institute.

Doms, Mark, Timothy Dunne, and Kenneth R. Troske. 1997. "Workers, Wages, and Technology." *Quarterly Journal of Economics* 112(1): 253–90.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Sahin. 2013. "The Decline of the U.S. Labor Share." *Brooking Papers on Economic Activity* no. 2, pp. 1–63.

Feenstra, Robert, and Gordon Hanson. 1999. "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979–1990." *Quarterly Journal of Economics* 114(3): 907–40.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology.* Belknap Press.

Graetz, Georg, and Guy Michaels. 2018. "Robots at Work." *Review of Economics and Statistics* 100(5): 753–68.

Groover, Mikell P., Mitchell Weiss, Roger N. Nagel, and Nicholas G. Odrey. 1986. *Industrial Robotics: Technology, Programming and Applications.* McGraw-Hill Inc.

Grossman, Gene M., and Esteban Rossi-Hansberg. 2008. "Trading Tasks: A Simple Theory of Offshoring." *American Economic Review* 98(5): 1978–97.

Habakkuk, John H. 1962. *American and British Technology in the Nineteenth Century: The Search for Labor-Saving Inventions.* Cambridge University Press.

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi. 2013. "Two Perspectives on Preferences and Structural Transformation." *American*

Economic Review 103(7): 2752–89.

Hubmer, Joachim. 2018. "The Race Between Preferences and Technology." Unpublished paper, Yale University.

Lin, Jeffrey. 2011. "Technological Adaptation, Cities, and New Work." *Review of Economics and Statistics* 93(2): 554–74.

Mantoux, Paul. 1928. *The Industrial Revolution in the Eighteenth Century: An Outline of the Beginnings of the Modern Factory System in England*. Jonathan Cape, Ltd.

Michaels, Guy. 2007. "The Division of Labor, Coordination, and the Demand for Information Processing." CEPR Discussion Paper 6358.

Mokyr, Joel. 1990. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press: New York.

Ngai, L. Rachel, and Christopher A. Pissarides. 2007. "Structural Change in a Multisector Model of Growth." *American Economic Review* 97(1): 429–43.

Oberfield, Ezra, and Devesh Raval. 2014. "Micro Data and Macro Technology." NBER Working Paper 20452

Olmstead, Alan, and Paul Rhode. 2001. "Reshaping the Landscape: The Impact and Diffusion of the Tractor in American Agriculture, 1910–1960." *Journal of Economic History* 61(3): 663–98.

Rasmussen, Wayne D. 1982. "The Mechanization of Agriculture." *Scientific American* 247(3): 76–89.

vom Lehn, Christian. 2018. "Understanding the Decline in the U.S. Labor Share: Evidence from Occupational Tasks." *European Economic Review* 108: 191–220.

Wright, Greg. 2014. "Revisiting the Employment Impact of Offshoring." *European Economic Review* 66: 63–83.

Zeira, Joseph. 1998. "Workers, Machines, and Economic Growth." *Quarterly Journal of Economics* 113(4): 1091–1117.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 31–50

Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction

Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb

Much of the public attention paid to artificial intelligence concerns its impact on jobs. Understanding this impact requires comprehending the capabilities of this technology. The majority of recent achievements in artificial intelligence are the result of advances in machine learning, a branch of computational statistics. Most of the concepts in standard machine learning textbooks (like Alpaydin 2010; and Hastie, Tibshirani, and Friedman 2009) are familiar to economists, like regression, maximum likelihood estimation, clustering, and nonparametric regression. Other techniques are just entering the econometrician's toolkit: regression trees, neural networks, and reinforcement learning (for discussions in this journal, see Varian 2014; Mullainathan and Spiess 2017; Athey and Imbens 2017). Over the past decade or so, advances in computer speed, data collection, data storage, and algorithms have led to substantial improvements in these techniques, such that their use for commercial applications is proceeding rapidly.

Machine learning does not represent an increase in artificial *general* intelligence of the kind that could substitute machines for all aspects of human cognition,

■ *Ajay Agrawal is the Geoffrey Taber Chair in Entrepreneurship and Innovation and Professor of Strategic Management, Joshua S. Gans is the Jeffrey Skoll Chair in Technical Innovation and Entrepreneurship and Professor of Strategic Management, and Avi Goldfarb is the Rotman Chair in Artificial Intelligence and Healthcare and Professor of Marketing, all at the Rotman School of Management, University of Toronto, Toronto, Canada. All three authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are ajay@agrawal.ca, joshua.gans@utoronto.ca, and agoldfarb@rotman.utoronto.ca.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.31 doi=10.1257/jep.33.2.31

but rather one particular aspect of intelligence: *prediction* (Agrawal, Gans, and Goldfarb 2018). We define prediction in the statistical sense of using existing data to fill in missing information. As deep-learning pioneer Geoffrey Hinton (2016) said, “Take any old problem where you have to predict something and you have a lot of data, and deep learning is probably going to make it work better than the existing techniques.”

Prediction is useful because it is an input into decision-making. Prediction has no value in the absence of a decision. In this sense, each prediction task is a perfect complement to a decision task. A prediction specifies the confidence of a probability associated with an outcome under conditions of uncertainty. As an input into decision-making under uncertainty, prediction is essential to many occupations, including service industries: teachers decide how to educate students, managers decide who to recruit and reward, and janitors decide how to deal with a given mess. This wide breadth of application means that developments in artificial intelligence represent what Bresnahan and Trajtenberg (1995) called a “general purpose technology.”

Prediction, however, is not the only element of a decision. Effective decision-making also requires collection and organization of data, the ability to take an action based on a decision, and the judgment to evaluate the payoffs associated with different outcomes. We characterize the decision task as distinct from the prediction task (Agrawal, Gans, and Goldfarb 2018, 2019).

We examine four direct effects through which advances in prediction technology may affect labor in a task-based framework: 1) substituting capital for labor in prediction tasks, 2) automating decision tasks when automating prediction increases the relative returns to capital versus labor, 3) enhancing labor when automating the prediction task increases labor productivity in related decision tasks and thereby increases the relative returns to labor versus capital in those tasks, and 4) creating new decision tasks when automating prediction sufficiently reduces uncertainty as to enable new decisions that were not feasible before.

First, artificial intelligence may directly substitute capital for labor in prediction tasks. Some tasks, like demand forecasting, are already prediction tasks. Where humans currently perform these prediction tasks, they are increasingly replaced by artificial intelligence. At the same time, other tasks that were not historically viewed as prediction tasks are being transformed into prediction-oriented tasks as machine learning improves and the quality-adjusted cost of prediction decreases. Many parts of the workflow in human resources are being broken down into prediction tasks so that they can then be performed by artificial intelligence tools. For example, in the broad area of human resources, recruiting is the task of predicting—based on resumes, cover letters, LinkedIn profiles, and interview transcripts—which subset of applicants will perform best in the job. Promotion is the task of predicting which existing employees will perform best in a higher-level position. And retention is the task of predicting which star employees are most likely to leave and which of the available incentive options could most effectively be employed to encourage them to stay.

Second, when automated prediction can increase the relative returns to capital versus labor in complementary decision tasks, it can lead to the complete automation of a complementary decision task. For example, human reaction times are slower than those for machines. The returns to a machine predicting a potential car accident a few seconds or even a fraction of a second before a human would predict the accident is higher when the response time of the machine is faster. Thus, automating the prediction task increases the returns to also automating certain decision tasks associated with vehicle control. Sometimes, the artificial intelligence is able to make better predictions than a human could because it has access to different data, such as feeds from cameras, RADAR, and LIDAR around a car.[1] Once the prediction task is automated, it increases the returns to automating some of the complementary tasks, such as those associated with vehicle control.

Third, automating the prediction task, in some cases, may have no impact on the productivity of capital performing a complementary task but may increase the productivity of labor. For example, ODS Medical developed a way of transforming brain surgery for cancer patients. Previously, a surgeon would remove a tumor and surrounding tissue based on previous imaging (say, an MRI scan). However, to be certain all cancerous tissue is removed, surgeons frequently end up removing more brain matter than necessary. The ODS Medical device, which resembles a connected pen-like camera, uses artificial intelligence to predict whether an area of brain tissue has cancer cells or not. Thus, while the operation is taking place, the surgeon can obtain an immediate recommendation as to whether a particular area should be removed. By predicting with more than 90 percent accuracy whether a cell is cancerous, the device enables the surgeon to reduce both type I errors (removing noncancerous tissue) and type II errors (leaving cancerous tissue). The effect is to augment the labor of brain surgeons. Put simply, given a prediction, human decisionmakers can in some cases make more nuanced and improved choices.

The fourth and final type of direct impact of artificial intelligence on labor happens when automated prediction sufficiently reduces uncertainty as to enable new decision tasks that did not exist before. The new tasks can be performed by capital or labor, depending on the relative costs of each (Agrawal, Gans, and Goldfarb 2019). Some tasks that are not economically viable when uncertainty is high become viable as prediction technology reduces the level of uncertainty. This relates to the reinstatement force in Acemoglu and Restrepo (in this issue) where a freeing up of labor as a result of automation increases the returns to technologies that use labor for new tasks. At this early stage in the development and use of machine learning, there are few tangible examples of new tasks that have already arisen because of recent advances in prediction technology.

The interaction of these four forces determines the net *direct* effect of cheaper quality-adjusted predictions on labor demand. There are also *indirect* effects: as some tasks become more efficient, demand for upstream and downstream tasks

[1] For an example of artificial intelligence detecting a crash two cars ahead via RADAR, before it was humanly possible to predict, see https://www.youtube.com/watch?v=FadR7ETT_1k.

might change. For example, an artificial intelligence that automates translation on an online trading platform significantly enhances international trade (Brynjolfsson, Hui, and Liu 2018). The application of this technology not only affects translators, but also the labor involved upstream and downstream on both sides of the trade.

For individual workers, the relative importance of these forces will depend on the degree to which the core skill they bring to their job is predicated on prediction. Workers whose core skill is something other than prediction, such as the brain surgeon described above, may find that automated prediction enhances the value of their occupation. On the other hand, workers whose core skill is prediction, such as human resource workers who screen resumes, may find the value of their occupation diminished.

In our work with the Creative Destruction Lab at the University of Toronto, looking systematically at several hundred artificial intelligence startups in the last few years, we have found that these firms often discuss how their technology will affect labor markets in specific occupations through substitution, complementarity, *and* demand expansion. We have seen very few companies building unambiguously labor-replacing technologies.

Overall, we cannot assess the net effect of artificial intelligence on labor as a whole, even in the short run. Instead, most applications of artificial intelligence have multiple forces that impact jobs, both increasing and decreasing the demand for labor. The net effect is an empirical question and will vary across applications and industries.

Automating Prediction Tasks

In this section, we describe examples that highlight substitution of human prediction by machine prediction in real-world applications. The subsequent two sections describe how such substitutions could increase either the relative returns to capital or to labor in the decision task and therefore result in automating or increasing the demand for labor in the decision task, respectively.

Prediction in Legal Services

A number of artificial intelligence applications substitute capital for labor by automating prediction tasks in legal work, while still leaving the decision tasks to the human lawyer. We describe two examples.

Kira Systems uses artificial intelligence technology to scan contracts and summarize relevant content. This may involve predicting which party in a particular lease agreement is liable for what actions or expenses, or it may involve scanning all of the contracts signed by a firm to predict which ones would be impacted if that firm were involved in a merger or acquisition. It is still up to human lawyers to make the decisions (as regulation requires), but Kira's technology predicts the relevance of clauses and information in a fraction of the time it would take a lawyer or paralegal.

In addition, artificial intelligence technology is being used to predict likely judicial outcomes based on earlier legal judgments. Blue J Legal's artificial intelligence scans tax law and decisions to provide firms with predictions of their tax liability. As one example, tax law is often ambiguous on how income should be classified (as discussed in Alarie 2018). At one extreme, if someone trades securities multiple times per day and holds securities for a short time period, then the profits are likely to be classified as business income. In contrast, if trades are rare and assets are held for decades, then profits are likely to be classified by the courts as capital gains. Currently, a lawyer who takes on a case collects the specific facts, conducts research on past judicial decisions in similar cases, and makes predictions about the case at hand. Blue J Legal uses machine learning to predict the outcome of new fact scenarios in tax and employment law cases. In addition to a prediction, the software provides a "case finder" that identifies the most relevant cases that help generate the prediction.

The end result of this process is not certainty. In the securities trading example above, the artificial intelligence predicts the likelihood of particular case facts being classified as business income or capital gains. As Blue J Legal founder Benjamin Alarie (2018) describes it, judges take the input of facts found at trial and output a judgment, using legal reasoning as a mapping function from inputs to outputs. In contrast, the Blue J Legal artificial intelligence utilizes test-facts that are assumed and entered into the system by the user, rather than the case facts found by trial. Instead of legal reasoning, the mapping function is a prediction generated by machine learning that is based on training data of past cases. Blue J Legal claims 90 percent accuracy. Given the uncertainty, a lawyer still makes the ultimate decision.

These examples show how a machine can save time and improve accuracy in generating predictions. In legal work, lawyers still make the ultimate decision. Thus, it is hard to forecast how the effect of these artificial intelligence technologies will show up in the aggregate labor statistics for legal work. They substitute for lawyers' prediction tasks but may create opportunities at the decision-task level because better prediction might affect prices and quantities in a way that increases demand for legal decision-making overall.

Prediction in Driving

The potential for mass adoption of fully autonomous vehicles generates headlines, but prediction technology is already changing driving in a number of ways that do not replace human drivers with machines.

For example, vehicle manufacturers use artificial intelligence to warn drivers about imminent risks like "there is probably a car in your blind spot" or "there is likely a pedestrian behind your car" in the form of a beep or blinking light. The machine provides the prediction, but the driver is still responsible for the decision of whether to stop, turn, or proceed.

Vehicle maintenance scheduling is another a prediction problem. Decades ago, Rust (1987) developed an empirical model of Harold Zurcher, who was the superintendent of maintenance at the Madison (Wisconsin) Metropolitan Bus

Company. Using statistical predictions of Zurcher's decisions, the model could be used to substitute for his predictions about when buses would break down. Today, advances in sensors and prediction algorithms have led to many new products that predict when a vehicle will break down and thus inform the decision of whether to bring a vehicle in for maintenance.

Finally, prediction is changing commercial driving by providing effective predictions of the most efficient route between two locations at any given time. Perhaps the most dramatic example is the case of London taxicabs. For decades, earning a taxi license in London meant acquiring "The Knowledge," which involved learning the location of every address in London as well as the shortest route between any two addresses. To pass the resulting test took two to four years of study with the help of specialist training schools. But now, best-route prediction apps like Waze deliver "The Knowledge" to any driver with a smartphone, which is part of what enables ride-sharing services such as Uber to compete with London taxis. Although the skill of London cabbies did not diminish, their competitive advantage was seriously eroded by artificial intelligence.

The end result on employment is unclear. While it is surely negative for the incomes of London cabbies, overall it may be positive if more drivers (are allowed by regulators to) enter the market. This provides some insight into the types of jobs likely to be most negatively affected by artificial intelligence: jobs in which the core skill involves a prediction task.

Predictions in Email Responses

Composing an email response can be formulated as a prediction problem. Google developed Smart Reply for its email service, Gmail, using artificial intelligence to scan incoming emails and predict possible responses. Smart Reply doesn't automate sending the email response but rather predicts possible responses and provides the user with three suggestions. In 2018, within weeks of Google rolling out Smart Reply as a default setting for all of its 1.4 billion active Gmail accounts, 10 percent of all Gmail responses sent were generated by Smart Reply (as reported by Marcelis and MacMillan 2018). This saves the user the time of composing a response in cases where one of the three predicted replies are sufficient. However, the user must still decide whether to send a predicted response or to compose one directly.

In some cases, this kind of artificial intelligence implementation might lead to a setting where a worker must still apply judgment about the benefits and costs of a particular decision before deciding or taking an action; in others, it might automate the full decision.

To understand how drafting email might affect different types of jobs differently, we turn to the O*NET database. Sponsored by the US Department of Labor through a grant to the North Carolina Department of Commerce, O*NET offers detailed descriptions of the tasks involved in almost 1,000 occupations (https://www.onetcenter.org).

This data includes a task described as "Prepare responses to correspondence containing routine inquiries." The job of Executive Assistants includes this task, along with eight other occupations: Correspondence Clerks, Tellers, Receptionists and Information Clerks, License Clerks, Legal Secretaries, Insurance Policy Processing Clerks, Medical Secretaries, and Loan Interviewers and Clerks. Executive Assistants would typically draft possible responses for someone else to decide whether or not to send, and so a system like Gmail's Smart Reply fully automates the Executive Assistant's decision. In the other jobs, the worker might make use of this technology but still retain the decision task of what to ultimately send. So in the former case, the artificial intelligence replaces labor, while in the latter case it enhances labor.

In this section, we provided examples in which machine prediction displaces human prediction. However, we can say little about the overall effect on jobs, which depends on the impact of better prediction on the decision tasks that they inform.

Automating Decision Tasks

Under certain conditions, automating the prediction task increases the relative returns to automating the decision task compared to performing the decision task with human labor. In other words, when artificial intelligence is used to automate prediction, it can enhance the usefulness of implementing other technologies to automate the decision. If both the prediction and the decision are automated, it must be possible to specify the desired action to be taken for each realization of uncertainty (that is, for each realization of a prediction). For reasons of simplicity, the most common type of machine-based decision is binary—say, to reject or accept a credit application or to recommend or reject a candidate for a job interview. As artificial intelligence improves, it will provide better predictions in more complex environments.

We have started to see this type of automation in environments where machine-learning techniques are applied to mimic human decision-making. For example, a machine fitted with sensors is trained by observing the choices made by a human operator. With sufficient observations, the machine learns to predict what action a human would take given different sensory inputs. As another example, the autonomous operation of vehicles on public roads has been advanced by humans driving millions of kilometers in vehicles that are able to collect both the perception data regarding environmental conditions on the road (input) and the action data regarding the decisions made by human drivers behind the wheel (output) in response to the perception data. In many cases, the response time of an automated control system is sufficiently faster than that of a human, so machines are better able to take advantage of the higher fidelity predictions generated by artificial intelligence compared to predictions generated by humans. Thus, the returns to automating the action decision (control of the vehicle) increase upon automating the prediction task. In this way, the application of artificial intelligence to automate the prediction task leads to automating the entire driving task resulting in a full substitution of labor for capital.

Autonomous driving gets a great deal of attention because so many people spend so much time driving. The labor-saving time from automation is therefore potentially large. However, this is also an area where removing human operation completely involves substantial risks, because the cost of failure can be so high. At present, a measure of human supervision is still required due to the probability of edge cases arising for which the machine has not been appropriately trained.

Commercial cleaning illustrates a more pedestrian attempt at automation. A&K Robotics takes existing, human-operated cleaning devices, retrofits them with sensors and a motor, and then trains a machine learning-based model using human operator data so the machine can eventually be operated autonomously. Artificial intelligence enables prediction of the correct path for the cleaning robot to take and also can adjust for unexpected surprises that appear in that path. Given these predictions, it is possible to prespecify what the cleaning robot should do in a wide range of predicted scenarios, and so the decisions and actions can be automated. If successful, the human operators will no longer be necessary. The company emphasizes how this will increase workplace productivity, reduce workplace injuries, and reduce costs.

Artificial intelligence also has enabled the automation of vehicles which move items from the storage part of a warehouse to the packing and shipping department. Much of this automation occurred without machine learning, by simply using dedicated tracks for delivery vehicles. However, recent applications of artificial intelligence enable swarms of robots to predict optimal routes and avoid collisions, eliminating the need for human controllers to decide on route planning. Under these conditions, warehouse vehicles can be fully automated.

Similarly, vehicle automation is growing in the mining industry, in particular for remote operations. In Australia's Pilbara region, the iron-ore mining sites are over 1,000 miles from the nearest major city. Given the remoteness of the region and the extremely hot temperatures, human truck drivers are unusually expensive. Mining giant Rio Tinto initially addressed this problem by driving the trucks remotely from the offices in Perth, but in 2016 the company went a step further and deployed dozens of self-driving trucks. Artificial intelligence made this automation of the steering decision task possible by predicting hazards in the roads and by coordinating the trucks with each other. As with robot cleaning and robots in warehouses, with better prediction other technologies made automation of the decision tasks possible. Automated prediction was the last step in removing humans from the decisions involved.

Unlike autonomous vehicles on public roadways, in these controlled environments with far fewer "edge" cases, cheap prediction has already led to widespread automation of the decisions. In this way, jobs in which the core bottleneck to automation is prediction become more likely candidates for elimination.

Augmenting Labor on Decision Tasks

Discussions of artificial intelligence often envision a future of full automation. Although this may happen in some situations, in this section we discuss examples where the automation of prediction through artificial intelligence can improve decision-making by humans and consequently the productivity of labor, specifically by allowing workers to make state-contingent decisions that reduce errors, enhancing payoffs.

Bail Decisions

Judges make decisions about whether to grant bail and thus to allow the temporary release of an accused person awaiting trial, sometimes on the condition that a sum of money is lodged to guarantee their appearance in court. Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018) study the predictions that inform this decision:

> Soon after arrest, a judge *decides* where defendants will await trial, at home or in jail. By law, this decision should be based solely on a prediction: What will the defendant do if released? Will they flee or commit a new crime? … Currently the predictions on which these decisions are based are, in most jurisdictions, formed by some judge processing available case information in their head. … A judge must trade off these risks [flee or commit a new crime] against the cost of incarceration. This is a consequential decision for defendants since jail spells typically last several months (or longer); recent research documents large costs of detention even over the long term. It is also costly to society: at any point in time the US has over 750,000 people in jail, disproportionately drawn from disadvantaged and minority populations.

Bail risk-prediction software will not replace people. Judges will continue to weigh the relative costs of errors, and in fact the US legal system requires human judges to decide. But artificial intelligence could enhance the productivity of judges. The main social gains here may not be in hours saved for judges as a group, but rather from the improvement in prediction accuracy. Police arrest more than 10 million people per year in the United States. Based on AIs trained on a large historical dataset to predict decisions and outcomes, the authors report simulations that show enhanced prediction quality could enable crime reductions up to 24.7 percent with no change in jailing rates or jailing rate reductions up to 41.9 percent with no increase in crime rates. In other words, if judicial output were measured in a quality-adjusted way, output and hence labor productivity could rise significantly.

Emergency Medicine

Prediction technology additionally has the potential to make medicine more efficient and effective through the personalization of treatment. In particular, machine learning can identify those patients for which a given treatment will be most effective.

In the United States in particular, many prominent policymakers, economists, and medical researchers have argued that doctors test too much. For example, doctors appear to test too much for heart attacks among patients who arrive at the emergency department of hospitals, in the sense that the average return to testing appears to be less than the average cost. Mullainathan and Obermeyer (2018) emphasize that prediction tools enable doctors to use the theoretically relevant object: marginal benefits and marginal costs. Using machine learning, Mullainathan and Obermeyer demonstrate that not only does the system suffer from overtesting (many low-risk patients are tested), but it also suffers from *under* testing (many high-risk patients go untested). Using better prediction models when patients arrive at the emergency department could substantially increase health outcomes for the same spending (or substantially reduce spending for equal health outcomes).

Automation of the prediction of who to test does not change the workflow. The machine prediction saves only a little time for the doctors, nurses, and staff working in the emergency department, who would otherwise make the testing decision quickly, but it improves efficiency and outcomes, both reducing the number of unnecessary tests and increasing necessary ones, thus saving lives. In other words, automating the prediction task improves the productivity of emergency medicine in the context of heart attacks without much substitution for human work. Furthermore, by helping to overcome problems of under-testing, it increases demand for labor in complementary downstream tasks such as surgery, the subject of the next section.

In this section, we highlighted how automation of the prediction task can enhance labor in the decision task. The prediction tasks in these examples do not require much human labor, and so the net effect could be labor-enhancing with better and sometimes more decision tasks done by labor.

Indirect Effects: Augmenting Labor on Other Tasks

When automated prediction leads to better decisions, labor can also be augmented through tasks that are upstream or downstream of the improved decision task.

Drug Discovery

A company called Atomwise uses artificial intelligence to enhance the drug discovery process. Traditionally, identifying molecules that could most efficiently bind with proteins for a given therapeutic target was largely based on educated guesses and, given the number of potential combinations, it was highly inefficient. Downstream experiments to test whether a molecule could be of use in a treatment often had to deal with a number of poor-quality candidate molecules.

Atomwise automates the task of predicting which molecules have the most potential for exploration. Their software classifies foundational building blocks of organic chemistry and predicts the outcomes of real-world physical experiments. This makes the decision of which molecules to test more efficient. This increased efficiency, specifically enabling lower cost and higher accuracy decisions on which molecules to test, increases the returns to the downstream lab testing procedure that is conducted by humans. As a consequence, the demand for labor to conduct such testing is likely to increase. Furthermore, higher yield due to better prediction of which chemicals might work increases the number of humans needed in the downstream tasks of bringing these chemicals to market. In other words, automated prediction in drug discovery is leading to increased use of already-existing complementary tasks, performed by humans in downstream occupations.

Language Translation

Machine language translation offers another example of machine prediction affecting a wide variety of downstream decisions. In this setting, artificial intelligence predicts how a human would translate a string of characters from one language into another. In one of the first attempts to estimate the economic impact of a commercial deployment of artificial intelligence, Brynjolfsson, Hui, and Liu (2018) measure the effect of an improvement in the quality of translation by an artificial intelligence on the volume of trade conducted on the online platform eBay. The authors find that moving to a translation using artificial intelligence resulted in a 17.5 percent increase in the volume of trade. This improvement in the prediction task results in a significant increase in downstream trade activity, much of which we can assume is performed by human labor. Of course, increased trade has many forms of impact on economic activity, and so we cannot draw any conclusions on the overall impact of this implementation of artificial intelligence on labor in equilibrium.

This section provided examples of the effects of improved prediction on workers in other parts of the production chain, beyond the focal prediction and decision tasks.

The Case of Radiology

Radiology offers an example of how artificial intelligence is leading to the automation of an occupation. At an artificial intelligence conference several years ago, deep learning pioneer Geoffrey Hinton (2016) publicly asserted, “We should stop training radiologists now,” comparing the profession to Wile E. Coyote from the Road Runner cartoon, who has run off the cliff but hasn’t yet looked down. However, the effect of artificial intelligence on the number of workers in radiology turns out, on closer examination, to be ambiguous and nuanced. This occupation offers a useful case study that employs several of the themes developed so far.

Hinton’s remark was motivated by the progress of artificial intelligence tools that are increasingly applied to identify abnormalities in medical images. IBM and

GE commercialized artificial intelligence tools that identify breast, lung, and other cancers from medical images. Smaller companies and startups have similar products. For example, Zebra Medical Vision received approval from the Food and Drug Administration to predict whether coronary heart disease is present in a CT scan. Zebra also develops tools to predict the presence of various medical issues, including bone, liver, and lung disease (as discussed at http://www.zebra-med.com).

A common practice is to embed image-recognition technology using artificial intelligence into the software that radiologists use to read scans. The software highlights areas predicted to be abnormalities. Radiologists examine the highlighted image when interpreting and reporting on the results. This approach uses artificial intelligence to augment the diagnosis decisions of humans rather than replace them altogether (Wang, Khosla, Gargeya, Irshad, and Beck 2016). In these cases, a human radiologist remains in the loop for each scan, but the readings become faster and more accurate. If the number of scans stays fixed, then the demand for radiologists declines. On the other hand, if readings are faster, more accurate, and cheaper, then the number of scans could increase enough to counteract the increased number of scans read per radiologist. This scenario belongs in the earlier section where artificial intelligence automates the prediction task but not the decision.

However, some recent research suggests that machine prediction can meet or even surpass human diagnostic accuracy in detecting some types of disease (for example, Lee et al. 2017). While the current level of technology suggests a human should remain in the loop, it is plausible that over time artificial intelligence will lead to full automation of the image interpretation task. In this scenario, if the "interpret imaging results" task is done by machine, and to the extent that this task takes up a significant fraction of the overall time, then automating this task could reduce the demand for radiologists.

But even in this situation, many tasks in the workflow of diagnostic radiologists would remain: choosing the exam, directing the technologists, reporting on the results, and deciding on an action given the probabilities reported by the machine. Many radiologists serve as the "doctor's doctor," communicating the meaning of images to other patient-facing doctors (Hall 2009). The interpretation of scans is often probabilistic, and radiologists have expertise in interpreting probabilities to help the patient-facing doctor recommend a course of action. Thus, reporting on the results may require a human intermediary between the machine prediction and the doctor who requested a test. For example, a human is needed to consider payoffs in order to recommend a course of action. What is the cost of conducting a biopsy if no disease is present? What is the cost of failing to conduct a biopsy if disease is present? In other words, what is the probability of disease threshold over which a biopsy (or some other further action) should be conducted? How does that vary based on patient characteristics, whether fully codifiable (such as age and medical history) or not (such as the doctor's sense of the patient's personality and preferences)? As the prediction task becomes better, faster, and cheaper, the demand for these related, complementary tasks may increase. In other words, it is plausible that automating the image prediction task, while reducing the demand for labor to perform that specific task, may increase the overall demand for labor due to an increased demand for complementary tasks.

In Table 1, we list the 29 different tasks that comprise the radiologists' workflow according to the occupational classification database O*NET. Only two of these tasks are directly affected by an image recognition AI: #3 and #25. Overall, the 29 tasks reveal that even if image interpretation becomes fully automated, plenty of tasks for humans remain. The key open question is whether those tasks are best conducted by a radiologist. Perhaps some of these tasks might be better performed by medical practitioners with different expertise? For example, judgment on the best course of action for a patient might be best decided by a primary care physician or perhaps even a social worker. The supervision of radiology technologists might be better managed by more experienced radiology technologists.

Technology using artificial intelligence will also affect radiology in a variety of other ways, apart from predicting abnormalities in scans. For example, radiologists often dictate their reports. Past practice was that the recorded reports were sent to a (human) transcription service (as in Task #4 in Table 1). But many radiology departments already use artificial-intelligence-based transcription services to automate the transcription task. While this step can reduce costs and reduce wait times for radiologists and patients, the direct effect is the elimination of transcription-related jobs (as discussed in Thrall et al. 2018), not radiologists.

The overall message here is that even when considering what may seem at first to be a clear-cut case—automation of the prediction related to reading medical image scans—the overall effects on jobs can be complex. Humans working in radiology who are not radiologists and do not work on scans—such as those providing transcription services—may have their jobs automated completely. Radiologists perform many other nonprediction tasks, and so artificial intelligence is unlikely to automate these tasks; however, it is not clear that radiologists will be the humans who perform these tasks if reading scans becomes automated. It is ultimately not obvious even whether the number of radiologists will rise or fall, since that will depend on whether radiologists perform the nonprediction tasks and whether overall demand for radiology services rises as radiology becomes more efficient.

New Tasks through New Decisions

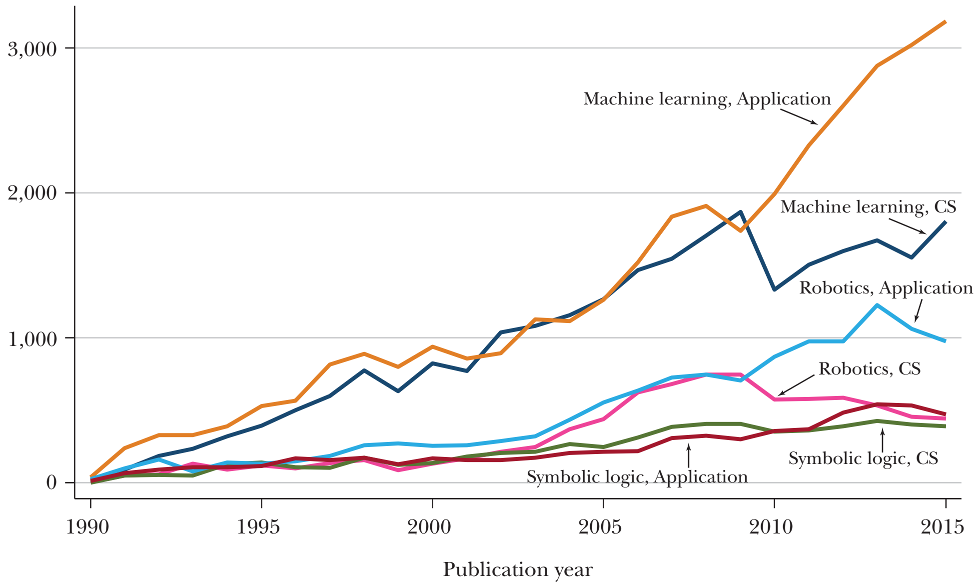
As artificial intelligence improves prediction, it may allow for new decisions to be made where previously it was impossible or too costly to do so. As Herbert Simon (1972) emphasized, when rationality is bounded—for example, in terms of being able to distinguish adequately between important outcomes in a complex environment—economic agents will instead resort to rules. Those rules can take various forms: they may be followed by individuals, be part of operational procedures in companies, or be embedded in machines. However, when uncertainty is reduced, generic rules may be replaced with probability-driven decisions. This is important because state-contingent choices can be consequential for companies.

Table 1
Twenty-nine Tasks Associated with the Occupation of Radiologist

1. Obtain patients' histories from electronic records, patient interviews, dictated reports, or by communicating with referring clinicians.
2. Prepare comprehensive interpretive reports of findings.
3. Perform or interpret the outcomes of diagnostic imaging procedures including magnetic resonance imaging (MRI), computer tomography (CT), positron emission tomography (PET), nuclear cardiology treadmill studies, mammography, or ultrasound.
4. Review or transmit images and information using picture archiving or communications systems.
5. Communicate examination results or diagnostic information to referring physicians, patients, or families.
6. Evaluate medical information to determine patients' risk factors, such as allergies to contrast agents, or to make decisions regarding the appropriateness of procedures.
7. Provide counseling to radiologic patients to explain the processes, risks, benefits, or alternative treatments.
8. Instruct radiologic staff in desired techniques, positions, or projections.
9. Confer with medical professionals regarding image-based diagnoses.
10. Coordinate radiological services with other medical activities.
11. Document the performance, interpretation, or outcomes of all procedures performed.
12. Establish or enforce standards for protection of patients or personnel.
13. Develop or monitor procedures to ensure adequate quality control of images.
14. Recognize or treat complications during and after procedures, including blood pressure problems, pain, oversedation, or bleeding.
15. Administer radiopaque substances by injection, orally, or as enemas to render internal structures and organs visible on x-ray films or fluoroscopic screens.
16. Participate in continuing education activities to maintain and develop expertise.
17. Participate in quality improvement activities including discussions of areas where risk of error is high.
18. Supervise and teach residents or medical students.
19. Implement protocols in areas such as drugs, resuscitation, emergencies, power failures, or infection control.
20. Schedule examinations and assign radiologic personnel.
21. Provide advice on types or quantities of radiology equipment needed to maintain facilities.
22. Participate in research projects involving radiology.
23. Perform interventional procedures such as image-guided biopsy, percutaneous transluminal angioplasty, transhepatic biliary drainage, or nephrostomy catheter placement.
24. Administer or maintain conscious sedation during and after procedures.
25. Interpret images using computer-aided detection or diagnosis systems.
26. Serve as an offsite teleradiologist for facilities that do not have on-site radiologists.
27. Develop treatment plans for radiology patients.
28. Treat malignant internal or external growths by exposure to radiation from radiographs (x-rays), high energy sources, or natural or synthetic radioisotopes.
29. Conduct physical examinations to inform decisions about appropriate procedures.

Source: O*NET, https://www.onetonline.org/link/summary/29-1069.10.

Figure 1
Publications in Computer Science (CS) versus Application Journals, by Artificial Intelligence Field



Source: Cockburn, Henderson, and Stern (2019, figure 4).
Note: The Figure shows the number of publications in computer science and applications journals by artificial intelligence field: machine learning, robotics, or symbolic logic.

For example, Google deployed artificial intelligence developed by its DeepMind unit to optimize the use of air conditioners in its data centers. The artificial intelligence enables new decisions on energy usage. The end result was a 40 percent reduction in energy used in a highly energy-intensive operation (Evans and Gao 2016).

New tasks may also be performed by humans. As highlighted above in the context of drug discovery, artificial intelligence is already having an impact on scientific research. Uncertainty is pervasive in many aspects of research, and so prediction technology is likely to have a large effect on the production of science. Cockburn, Henderson, and Stern (2019) show that machine learning is used by scientists in a wide variety of fields. Figure 1 shows the number of publications in computer science and applications journals by artificial intelligence field. The dark blue, pink, and green lines show publications in three different artificial intelligence subfields of computer science: machine learning, robotics, and symbolic logic. The results show a slow and steady increase in publications in all three, with the largest increase in machine learning. The most striking result in the figure, however, is the orange line. It shows the increase in publications outside of computer science that mention machine learning. In other words, it demonstrates that, since 2012, the biggest change in artificial intelligence publications did not occur in computer science.

It occurred in other fields of science that use machine learning. The same is not true for symbolic logic (green and red lines) and it is much weaker for robotics (pink and light blue). This use of machine learning as an input into innovation is apparent in our own field of economics, in which researchers are increasingly using machine learning to improve our statistical models and advance our knowledge of economics (Athey 2019).

In this way, the recent advances in machine learning can be seen as an invention in the method of inventing, as highlighted in Griliches (1957) for the example of hybrid corn. Cockburn, Henderson, and Stern (2019) describe this insight: "The challenge presented by advances in artificial intelligence is that they appear to be research tools that not only have the potential to change the method of innovation itself, but also have implications across a wide range of fields." Agrawal, McHale, and Oettl explain how artificial intelligence may influence the knowledge production function (2019a) and model the implications of using artificial intelligence to produce a map of the complex combinatorial search space of ideas for the purpose of reducing the cost of predicting which combinations of ideas offer the greatest promise (2019b). In the context of Atomwise, new tasks may arise if drug discovery becomes more efficient and drugs can be better-targeted to narrower populations. In other words, in addition to increasing demand for existing tasks, artificial intelligence is likely to create innovations that lead to new industries and new types of jobs with new tasks in those industries.

At this early stage in the diffusion of machine learning technology, examples of new tasks created by artificial intelligence are scarce and speculative. However, we provide three examples that apply artificial intelligence in unique areas that suggest the potential for new industries, jobs, and tasks.

At the University of Toronto, Alan Aspuru-Guzik's research group is developing a "self-driving chemistry lab" that enables the discovery of chemicals and materials at a fraction of the price of a current lab. Using advances in robotics and machine learning, the lab could be deployed in thousands of locations around the world, without the need for a local workforce with deep expertise in chemistry. This would enable industries in rural areas and developing countries to have access to a wide variety of materials. The inventors emphasize that this tool could "provide the scientific community with an easy-to-use package to facilitate novel discovery at a faster pace" (Roch et al. 2018, p. 1) and "democratize autonomous discovery" (p. 12). One can imagine many new tasks associated with the arrival of an autonomous chemistry lab with the capability of on-site discovery.

The commercialization of space offers another example of how machine learning could generate a new industry at a commercial scale. Uncertainty is a key challenge to operating assets in space. For example, the risk of destruction from debris is a well-established deterrent for deploying commercial satellites (Liou and Johnson 2006). The company Seer Tracking has built an artificial intelligence to predict the trajectory of space debris. Most directly, this could create a set of (human and machine) tasks focused on moving space assets out of the way of incoming debris. Perhaps more importantly, it could enable more commercial opportunities in space by reducing the risk that a space asset will be destroyed by debris. In other words, the uncertainty associated with space debris may mean that some decision tasks are never undertaken. Resolving this uncertainty could enable new commercial opportunities in space.

Another example is the management of chronic disease. One such disease, diabetes, leads to hundreds of thousands of deaths annually. Key to preventing hospital admissions and severe complications is the control of blood glucose levels; however, many diabetes patients have difficulty maintaining a relatively safe level of glucose control. Better prediction of current glucose levels could substantially reduce complications by enabling response (Ismail 2017). This improved prediction could generate a new set of tasks that benefit from lower cost and more accurate monitoring. Potential new tasks in managing chronic diseases like diabetes include managing the sensors, interpreting the data, and real-time advising on dietary and exercise habits. Inherent in these new tasks is the ability to tailor medical decisions and treatment to individual patients (Contreras and Vehí 2018). Improved management of chronic disease could arise because of the new individual-level decisions enabled by better prediction.

These examples are speculative. The technology is too early and the diffusion is too limited to offer definitive examples of new tasks arising from recently automated predictions. At the time of this writing, the most likely consequences of artificial intelligence on labor come from *existing* tasks that are affected from better, faster, and cheaper prediction. We already observe real examples of the reduction of labor due to automating existing prediction tasks, and we also see examples of increased demand for labor due to enhanced demand for certain existing tasks that are complements to prediction. Our broader theme is that uncertainty can render certain activities economically infeasible and so reduced uncertainty can enable new opportunities and new tasks to be implemented by some mixture of capital and labor.

Conclusion

Our contribution to the task-based model of technology and labor (as discussed in Acemoglu and Restrepo in this issue; Autor and Acemoglu 2011; Acemoglu and Restrepo 2019) is to highlight the usefulness of thinking in terms of prediction tasks and decision tasks, where decision tasks are perfect complements to prediction tasks, in the sense that prediction has no value without a decision. This structure describes how artificial intelligence directly substitutes capital for labor in the case of prediction tasks and may indirectly effect decision tasks by increasing or decreasing the relative returns to labor versus capital for decision tasks. It may also lead to increases in labor tasks upstream or downstream.

For any given worker, a key predictor of whether artificial intelligence will substitute for their job is the degree to which the core skill they bring to the job involves prediction. Transcription jobs are being automated as the core skill of that labor is predicting which words to type upon hearing a recording. For London taxi drivers, when artificial intelligence was employed to predict the optimal route through the city's streets, their jobs were put at risk (though other drivers' labor became augmented).

Artificial intelligence does not fit easily into existing analyses of the effect of automation on labor markets. The reasons are threefold. First, prediction is always strictly complementary to other tasks—namely decision-related tasks. Those tasks can be existing or newly possible because of better prediction. Second, better prediction improves decisions—whether taken by labor or capital—by enabling more nuanced decisions through the reduction of uncertainty. Finally, it is not yet possible to say whether the net impact on decision tasks—whether existing or new—is likely to favor labor or capital. We have found important examples of both, and there is no obvious reason for a particular bias to emerge. Thus, we caution on drawing broad inferences from the research on factory automation (for example, Acemoglu and Restrepo 2017; Autor and Salomons 2018) in forecasting the net near-term consequences of artificial intelligence for labor markets.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Chap. 12 in *Handbook of Labor Economics,* vol. 4, Part B, edited by David Card and Orley Ashenfelter, 1043–1171. Elsevier.

Acemoglu, Daron, and Pascual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets." NBER Working Paper 23285.

Acemoglu, Daron, and Pascual Restrepo. 2019. "Artificial Intelligence, Automation, and Work." Chap. 8 in *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb. University of Chicago Press.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence.* Harvard Business Review Press.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2019. "Prediction, Judgment, and Complexity: A Theory of Decision Making and Artificial Intelligence." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Agrawal, Ajay, John McHale, and Alexander Oettl. 2019a. "Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Agrawal, Ajay, John McHale, and Alexander Oettl. 2019b. "A Multi-task Combinatorial Model of AI-Aided Scientific Discovery and Innovation." Working paper, University of Toronto.

Alarie, Benjamin. 2018. "Predicting Judicial Decision-Making with Machine Learning." *Blue J Legal* webinar series, February 26, https://www.youtube.com/watch?v=-pgbQejGVPc&submissionGuid=c9852990-2f9e-46e6-8cd2-b7507bebcfe9 (accessed February 20, 2019).

Alpaydin, Ethem. 2010. *Introduction to Machine Learning*, 2nd edition. Cambridge, MA: MIT Press.

Athey, Susan. 2019. "The Impact of Machine Learning on Economics." Chap. 21 in *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31(2): 3–32.

Autor, David, and Ana Salomons. 2018. "Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share." *Brookings Papers on Economic Activity*, Spring.

Bresnahan, Timothy F., and Manuel Trajtenberg. 1995. "General Purpose Technologies 'Engines of Growth'?" *Journal of Econometrics* 65(1): 83–108.

Brynjolfsson, Erik, Xiang Hui, and Meng Liu. 2018. "Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform." NBER Working Paper 24917.

Cockburn, Iain M., Rebecca Henderson, and Scott Stern. 2019. "The Impact of Artificial Intelligence on Innovation." Chap. 4 in *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Contreras, Iván, and Josep Vehí. 2018. "Artificial Intelligence for Diabetes Management and Decision Support: Literature Review." *Journal of Medical Internet Research* 20(5): e10775.

Evans, Richard, and Jim Gao. 2016. "DeepMind AI Reduces Google Data Centre Cooling Bill by 40%." DeepMind blog, July 20, https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25(4): 501–522.

Hall, Ferris M. 2009. "The Radiology Report of the Future." *Radiology* 251(2): 313–16.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. New York: Springer.

Hinton, Geoff. 2016. "On Radiology." YouTube of comments at the 2016 Machine Learning and Market for Intelligence Conference in Toronto. Creative Destruction Lab: Machine Learning and the Market for Intelligence. https://www.youtube.com/watch?v=2HMPRXstSvQ (accessed February 20, 2019).

Ismail, Nick. 2017. "How AI Can Improve Patient Outcomes for Type 1 and 2 Diabetes." *Information Age,* July 25. https://www.information-age.com/ai-can-improve-patient-outcomes-type-1-2-diabetes-123467505/ (accessed February 20, 2019).

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics* 133(1): 237–93.

Lee, Hyunkwang, Shahein Tajmir, Jenny Lee, Maurice Zissen, Bethel Ayele Yeshiwas, Tarik K. Alkasab, Garry Choy, and Synho Do. 2017. "Fully Automated Deep Learning System for Bone Age Assessment." *Journal of Digital Imaging* 30(4): 427–41.

Liou, J.-C., and N. L. Johnson. 2006. "Risks in Space from Orbiting Debris." *Science,* January 20, 311(5759): 340–41.

Marcelis, David, and Douglas MacMillan. 2018. "Is This Article Worth Reading? Gmail's Suggested Reply: 'Haha, Thanks!'" *Wall Street Journal,* September 18, https://www.wsj.com/articles/very-interesting-awesome-love-it-gmail-users-confront-chipper-smart-reply-1537282569 (accessed February 20, 2019).

Mullainathan, Sendhil, and Ziad Obermeyer. 2018. "Using Machine Learning to Understand Human Decision-Making: Application to Health Care." Working paper, University of Chicago.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2): 87–106.

Roch, Loíc, Florian Häse, Christoph Kreisbeck, Teresa Tomayo-Mendoza, Lars Yunker, Jason Hein, and Alán Aspuru-Guzik. 2018. "ChemOS: An Orchestration Software to Democratize Autonomous Discover." ChemRxiv. https://chemrxiv.org/articles/ChemOS_An_Orchestration_Software_to_Democratize_Autonomous_Discovery/5953606.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica 55*(5): 999–1033.

Simon, Herbert. 1972. "Theories of Bounded Rationality." Chap. 8 in *Decision and Organization,* edited by C. B. McGuire and Roy Radner. North-Holland.

Thrall, James H., Xiang Li, Quanzheng Li, Cinthia Cruz, Synho Do, Keith Dreyer, and James Brink. 2018. "Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success." *Journal of the American College of Radiology* 15(3, Part B): 504–508.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2): 3–28.

Wang, Dayong, Aditya Khosla, Rishab Gargeya, Humayn Irshad, and Andrew H. Beck. 2016. "Deep Learning for Identifying Metastatic Breast Cancer." arXiv:1606.05718.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 51–70

"Automation" of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study

Jeremy Atack, Robert A. Margo, and Paul W. Rhode

Over the course of the nineteenth century, the United States experienced its first "industrial revolution." A central feature of this revolution was the mechanization of production, first through water power and later steam power. By the late nineteenth century, the process was well advanced, fostering serious concerns about its effects on labor (Giedion 1948; Hounshell 1984). For example, David A. Wells (1889, p. 68), a prominent US economist of the time, wrote that "the increasing frequency of strikes and industrial revolts ... have been largely prompted by changes in the conditions of production resulting from prior labor-saving inventions and discoveries" and he opined "the depression of industry in recent years has been experienced with greatest severity in those countries where machinery has been most extensively adopted." Indeed, the historical process was so disruptive that it inspired Edward Bellamy's *Looking Backward, 2000–1887* (1888), a utopian science fiction novel, which quickly became the era's third-largest best-seller and provoked extensive political and social discussion.

In the first annual report to Congress, Commissioner of Labor Carroll D. Wright (US Bureau of Labor 1886) drew attention to the problem of the "temporary displacement of labor and to conditions of industry and of society which would

■ *Jeremy Atack is Professor of Economics Emeritus, Vanderbilt University, Nashville, Tennessee. Robert A. Margo is Professor of Economics, Boston University, Boston, Massachusetts. Paul W. Rhode is Professor of Economics, University of Michigan, Ann Arbor, Michigan. All three authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are jeremy.atack@vanderbilt.edu, margora@bu.edu, and pwrhode@umich.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.51 doi=10.1257/jep.33.2.51

exist without the presence of power machinery," illustrating with several examples. In small arms production, one worker using conventional hand tools turned and fitted one musket stock per ten-hour day, whereas using specialized machines and dividing the tasks between them, three workers could turn and fit between 125 and 150 musket stocks per day, a 40- to 50-fold gain in labor productivity. Similarly, data from boot and shoe manufacturers suggested an 80 percent savings in labor for machine over handicraft production (US Bureau of Labor 1886, p. 81).

In 1894, Congress requested a fuller investigation, noting "there are works now in existence where the very best and highest grade of machinery is used that formerly employed cruder methods, and the men in charge have knowledge of the old methods as compared with the new; but these men are fast passing away, and the difficulty increases each year of securing the information sought …" (US Congress, House of Representatives 1894). To this end, it directed the Commissioner of Labor to "investigate and report upon the effect of the use of machinery upon labor and the cost of production, the relative productive power of hand and machine labor … and whether changes in the creative cost of products are due to a lack or surplus of labor or to the introduction of power machinery."

The resulting "Hand and Machine Labor" (HML) study took five years to complete, finally appearing as the thirteenth annual report of the Commissioner of Labor (US Department of Labor 1899). [1] The HML study presents its information at a level of detail that was highly unusual not only for its time but even ours, by analyzing the production of highly specific goods (for example, production of circular saw blades with a given number of teeth) at the task level for a matched pair of establishments, one of which produced the product by "hand" (or traditional artisanal) methods and the other using "machine" methods. Among other data, the report specifies the amount of time each task took, the sequence in which these were performed, the characteristics of the workers employed, the tool(s) used, and notably, the source of inanimate power, if any, including steam power, which was the key "general-purpose technology" of that historical period. Brynjolfsson and McAfee (2014, p. 6), for example, describe steam power as the first machine age's "most important" technological development, "overcoming the limitations of

[1] The US Bureau of Labor was established in the Department of the Interior by the Bureau of Labor Act (23 Stat. 60) on June 27, 1884. The Bureau's mission was to collect information about employment and labor. The Act also created the post of US Commissioner of Labor to direct the Bureau. Carroll Wright served as the first US Commissioner of Labor. The Bureau of Labor became an independent (sub-Cabinet) department through the Department of Labor Act (25 Stat. 182) on June 13, 1888. As indicated by the title of the legislation, the Bureau of Labor was renamed the "US Department of Labor" in 1888. The cabinet-level Department of Commerce and Labor was created in 1903 by the Department of Commerce Act (32 Stat. 827) on February 14, 1903. The Act authorized a new "Bureau of Labor" within the Department of Commerce and Labor, which took over the activities of the preceding "Department of Labor." Finally, in 1913, Congress created a separate cabinet-level Department of Labor, within which the "Bureau of Labor" was renamed the Bureau of Labor Statistics. As is clear from this timeline, the 1890s "Department of Labor" is a direct predecessor of the modern Bureau of Labor Statistics, which is why the National Archives stores the extant records of the 1890s department in its Record Group 257 ("Records of the Bureau of Labor Statistics"). The timeline above expands on Rockoff's (2019, pp.147–51) discussion.

muscle power, human and animal" and propelling a "sudden, sharp, and sustained jump in human progress."

The enormously complex Hand and Machine Labor data were published in two large, very dense volumes. We have digitized these data, coding and restructuring them to be tractable to modern econometric techniques. Our analysis here focuses on transitions at the task level from hand to machine production, and on the impact of inanimate power on labor productivity in machine production.

By "transitions at the task level" we mean whether particular tasks in hand production were no longer present under machine production; whether the task content remained the same, even if inanimate power was used under machine production; whether task reorganization occurred in the move from hand to machine labor; or whether entirely new tasks were present under machine labor. Transitions in which the task content remained the same except for the possibility of mechanization—we call these 1:1 transitions—were the most common. However, highly complex task reorganization did occur, and new task creation substantially dominated the abandonment of obsolete hand tasks. Overall, the transition to machine labor brought very large gains in productivity. We show in a regression analysis of the 1:1 transitions that use of steam power explains a large fraction of the productivity gain. Economic historians have been studying the diffusion and impact of steam power for a very long time; but as far as we know, our regressions are the first to show the productivity effects of steam power at the level of individual production tasks in an historical context.

We consider the Hand and Machine Labor data and our findings in the context of the modern "task-based approach" to production (Acemoglu and Autor 2011; Autor 2013; Zeira 1998). This literature develops models allowing technological change to reduce returns to specific factors, which is not possible in standard models of factor-augmenting technological change. We will focus in particular on Acemoglu and Restrepo's (2018) recent model of automation (also discussed in their paper for this symposium). Their model is quite useful in drawing out inferences as to how, in response to technical progress, some tasks are abandoned; others automated, and new, non-automated tasks created. Substituting "mechanized" for "automated" in their framework, we find a similar pattern in the data from the HML study. However, we will also argue that our historical example clearly parts company with Acemoglu and Restrepo in that their model abstracts from the division of labor. Indeed, there is considerable evidence that the diffusion of steam power enhanced the division of labor (Atack, Bateman, and Margo 2008), as Thomson (1989) also shows in the transformation of US boot and shoe production during this time. The underlying issue is the degree to which workers are specialized or not in the tasks they perform, and how this may feed back into human capital investment. Indeed, we will suggest that one of the meaningful differences between nineteenth-century mechanization and the current technological revolution based in robotics and artificial intelligence is that they seem to have quite different implications for the division of labor and thus for human capital investment.

The Hand and Machine Labor Study

Although the title of the 1899 study was "Hand and Machine Labor," Commissioner of Labor Wright cautioned in his introductory remarks that the words were not used in their strictest sense, but rather to characterize two different methods of production. "Machines" were used in "hand" production although these were usually simple hand tools—saws, hammers, chisels, files, knitting needles, screwdrivers, and the like—what he called "the primitive method of production which was in vogue before the general use of automatic or power machines" (US Department of Labor 1899, vol. 1, p. 11). Similarly, some tasks in machine production continued to be performed by hand using these same simple tools, including adjusting the machinery. For Wright, however, a crucial distinction was that, in machine production, "every workman has his particular work to perform, generally but a very small portion of that which goes to the completion of the article"—that is, division of labor was central (p. 11).

The basic unit of observation in the Hand and Machine Labor study was a matched pair of production units: one using hand methods, the other using machine methods to make a particular quantity of product. The products chosen were highly specific—for example, the output of "Unit 71" was described in the report as "SHOES:—100 pairs of men's medium grade, calf, welt, lace shoes, with single soles and soft box toes" (US Department of Labor 1899). Where necessary, production was scaled to industry norms by adjusting the time (and thus the cost) spent on tasks by the appropriate factor, keeping the number of workers unchanged (as we will explain further below). Overall, there are 672 paired units in the HML study: 27 in agriculture, 10 in mining and quarrying, and 9 in transportation, leaving 626 paired units producing manufactures. We focus on these manufactures.

As mentioned, the data were reported in two parts (and volumes). In Part 1, the following was reported for each unit (matched pair of plants producing a highly specific product): an industry classification, an exact description of the product, the standardized quantity of that product, the year in which the production under each method took place, the number of separate tasks of production, the number of different workers employed, and the total number of hours of work to produce the given quantity, the total labor costs, and the average daily hours of operation of the unit. In Part 2, the following information was reported for each mode of producing the product: a brief description of the task in the order in which it was performed; a list of capital goods or machines used in the task; the type of motive power if used; the number of workers assigned to that task; the number, age, gender, and occupational titles of the workers employed in the task; the hours of work by each employee engaged in the task; and the labor cost of each employee engaged in the task along with any miscellaneous comments.

The raw data were collected by trained agents either through direct observation or from written records, following up (sometimes repeatedly) when necessary to resolve inconsistencies and ambiguities. For machine production, the vast majority of the observations pertain to activities conducted in the mid-to-late 1890s (1894–98).

For a few products, the study was unable to find matching hand production from the same year that occurred nearby, presumably because the relevant establishments were no longer in existence. In such cases, the agents assiduously sought out historical records or, in 13 instances, located hand production establishments overseas that they deemed similar to those that no longer survived in the United States. All machine production data, however, was taken from US establishments. Moreover, in the majority of cases, two reports on hand and machine production were secured for establishments/manufacturers from different, widely separated, localities to help spot errors and omissions with "the better and more complete one then selected for presentation" (US Department of Labor 1899, vol. 1, p. 13).

A concrete example illustrates the exceptional (indeed, stupefying) detail in the published study. In making men's medium grade, laced shoes (Unit 71), the study compared production by a bespoke shoemaker producing a single pair of shoes with that of a factory producing 1,500 pairs, scaling the time (and cost) as if each in fact produced 100 pairs of shoes. [2] The shoe size is not specified but is (implicitly) assumed to be different for each pair. The data were tabulated, verso and recto, across several pages, with task identifiers aligning the rows across the left- and right-hand pages and with the numbering sequenced according to the order in which the tasks were performed in machine production.

Hand production of medium grade, laced shoes involved 72 tasks. Selecting and sorting the leather was one task in hand production—presumably so that the uppers for one pair of shoes could come from the same hide—compared with eight separate operations in machine production, for uppers, vamps, quarters, outsoles, insoles, lifts, and counters (machine-molded heel reinforcements), all of which had to be both sorted and matched. In hand production, the individual shoemaker traced each foot to create a cutting pattern and subsequently hand-carved a "last" (a wooden form around which each shoe was molded). These steps were crucial for the fit of the shoe and would be repeated for each individual customer served by the shoemaker. Producing lasts by hand was time-consuming, taking 54 minutes 24 seconds per pair—almost 92 hours for the production run of 100 different pairs of shoes. By contrast, under machine production, the factory skipped these steps, instead purchasing lathe-turned lasts for left and right feet in standard sizes from outside specialist suppliers, which would be used in the fabrication of thousands of pairs of shoes—an example of the subsidiary industries predicted to emerge to meet special needs once a certain scale of operation was achieved (Marshall and Marshall 1881, p. 52).

In the machine production of these shoes, the Hand and Machine Labor study identified 173 separate tasks. These include not only tasks directly related to the manufacture of shoes, like sorting leather, cutting out the vamps (the main part of the shoe between the toe and the laces), quarters (the heel portions), toes, soles, insoles,

[2] Exhibits 1 and 2 in the online Appendix available with this paper at the journal's website reproduce sections of the tables for Unit 71 detailing the tasks in the hand and machine production of men's medium grade, laced shoes from the Hand and Machine Labor report (US Department of Labor 1899).

and heels and sewing these together around the last to form the shoe and punching holes for the laces. Tasks also included finishing the shoes for market by smoothing the welts, waxing and polishing, matching pairs, stamping with the maker's name and size, and boxing for shipment. Moreover, other tasks involved keeping the shoemaking machinery in good order, and maintaining and firing the steam engine that powered the various machines—tasks not directly involved with production but vital to that production. Some of the tasks, like sorting, required nothing more than a good eye. Others, like cutting out the parts, still used basic hand tools (scissors and knives) rather than steam-powered die presses. Eighty of the tasks, however, including trimming, making eyelets, nailing heels, polishing and buffing, made use of steam power driving specialized machines (US Department of Labor 1899, vol. 2, pp. 544–51).

The study investigators carefully linked each operation in hand production to the corresponding operation in machine production via the machine task number. Machine tasks that were a part of several hand tasks had lowercase letters appended to the machine task number. The data showing the connections from hand tasks to machine tasks can be displayed as a "slope chart"[3] relating each of the various hand tasks for shoe-making on the left to the (far more numerous) machine tasks for shoe-making on the right, as in Figure 1. Tasks are numbered in sequence.

Some hand tasks link to multiple machine tasks. Some are performed in quite different sequences between hand and machine production—these lines cross over. A few hand tasks like "selecting and sorting stock" vanish in machine production (we have connected these to "Task 0" in Machine Production on the right hand side). Moreover, the white space on the right-hand axis to which no hand production tasks connect represent new tasks created by mechanization for which there was no hand production analog. In the next section, we discuss these task "transitions."

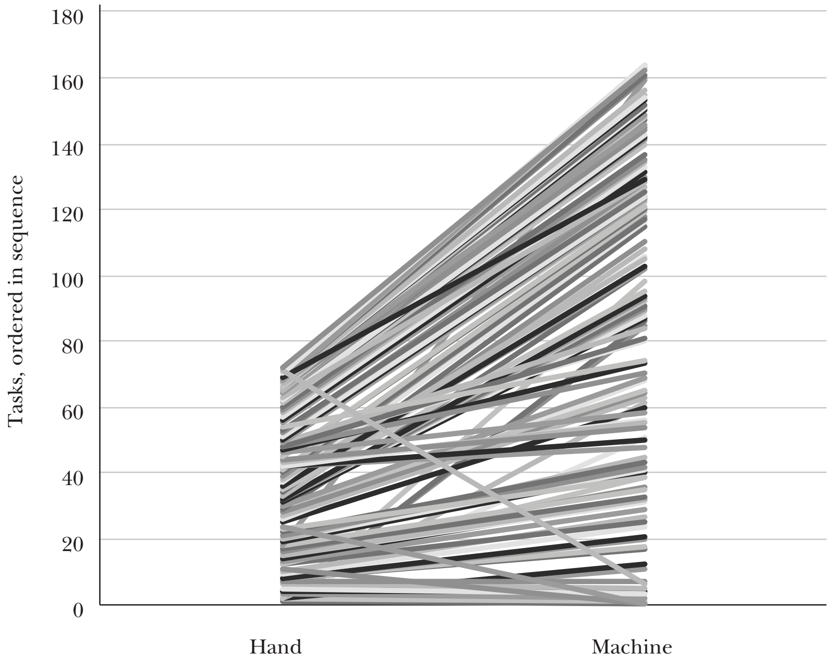
The complexity of the Hand and Machine Labor data overwhelmed statisticians at the time. As Carroll Wright (1900) would later remark: "This report answers in a measure the many demands for information … but no aggregation can be made because it is impossible to carry out calculations through the innumerable ramifications of production under hand and machine methods … although such a summary would be of the greatest possible value in the study of the question of machinery." Its complexity has also largely prevented analysis by modern economic historians until very recently.[4]

Before turning to our findings, we highlight four limitations of the Hand and Machine Labor data. First, although a wide range of goods and industries are covered, the establishments that were included are in no sense a random sample either within or across industries. Second, no information was collected on output prices, revenues, or costs, except those pertaining to the labor involved directly in

[3] For more on slope charts, see Tufte (1983, p. 143), or https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0003nk.

[4] Stanley Engerman has informed us (via personal communication) that he and Robert Fogel included the Hand and Machine Labor study on an unpublished list of key data sources in US economic history that the two prepared in the early 1970s. However, in their view the data were far too complex to digitize and analyze at that time, which was a reasonable judgment until recent advances in information technologies.

Figure 1
Slope Chart Linking Hand to Machine Tasks for Unit 71



Source: Authors.
Note: Figure 1 relates each of the various hand tasks for shoe-making on the left to the (far more numerous) machine tasks for shoe-making on the right. Tasks are numbered in sequence. Some hand tasks link to multiple machine tasks. Some are performed in quite different sequences between hand and machine production—these lines cross over. A few hand tasks like "selecting and sorting stock" vanish in machine production (we have connected these to "Task 0" in machine production on the right-hand side). Moreover, the white space on the right-hand axis to which no hand production tasks connect represent new tasks created by mechanization for which there was no hand production analog.

the production of the product (and its supervision). Consequently, any analysis of productivity, including ours, must rely on the measure provided by the study—the amount of time that it took to complete a task—rather than a measure that would be more conventional for economists like value added per worker. Third, while the agents recorded additional information on the survey form that would have been very useful to have for some analyses—for example, the names of the individual workers, and the address of the establishment—this information was not included in the published study. Moreover, as far as we can determine, the completed survey forms have not survived and so this additional information has been lost.[5] Finally,

[5]We tracked down copies of the original survey instrument which are now stored in Record Group (RG) 257 in the US National Archives (US Bureau of Labor Statistics 1890–1905). The forms asked for additional information that was not published, such as the name and location of the establishment, and the names of the workers employed in the production of the various articles.

as previously noted, the study reported the labor requirements for a standardized scale of production, which enhances comparability. But the number of workers employed and the organization of work may not reflect how producers, especially hand producers, would have operated at that specific scale under realistic time-cost considerations.

Task Transitions and the Role of Steam Power

We focus on three broad features of the Hand and Machine Labor data: transitions of tasks from hand to machine labor; the overall productivity gains associated with machine labor; and the impact of steam power on productivity in machine production for the subset of tasks that were common to both hand and machine labor (tasks in the 1:1 transition category as discussed below).

Task Transitions: From Hand to Machine Production

The data from the Hand and Machine Labor study allow us to see the transition from hand to machine labor at the task level. The agents collecting the data listed tasks in production order under both hand and machine manufacture, adding a column linking hand to machine tasks. This allows us to draw a slope chart as in Figure 1 and to distinguish six types of transitions from hand to machine tasks:

a) Hand tasks that were no longer performed under machine labor, or old tasks, which we label as 1:0 transitions;
b) Tasks whose content was deemed to be essentially the same in hand and machine production, except that the machine task might be mechanized, which we label as 1:1 transitions;
c) A single hand task that was subdivided into *M* machine tasks, which we label as 1:*M* transitions;
d) *N* hand tasks that were combined into a single machine task, which we label as *N*:1 transitions;
e) *N* hand tasks were mapped into *M* machine tasks, with both *N* and *M* greater than one, which we label as *N*:*M* transitions; and, lastly,
f) Tasks present under machine production but not hand production, or "new" tasks, which we label as 0:1 transitions.

Table 1 presents summary statistics on each different kind of task transition across production units in the study. Table 1A presents these statistics from the point of view of the origin, hand labor, and Table 1B from the point of view of the destination, machine labor. Instead of counts of transitions, we focus on the share of tasks for each type of transition. We normalize within production units either by the total number of tasks (equal weights) or by weighting each task by its share of total production time (time weights); in either case, our estimates of average shares are equally weighted across units. Both panels also show the proportion of tasks that

Table 1
Tasks Transitions, Hand to Machine Labor

Transition	*Share of tasks, equal weights*	*Share of tasks, time weights*	*Share using steam power, equal weights*	*Share using water power, equal weights*	*Share using steam power, time weights*	*Share using water power, time weights*
A: Hand Labor						
1:0	0.044	0.030	0.003	0.002	0.002	< 0.001
1:1	0.673	0.604	0.014	0.017	0.009	0.020
1:*M*	0.134	0.192	0.023	0.005	0.008	0.006
N:*M*, *N* > 1, *M* > 1	0.040	0.054	< 0.001	0.010	< 0.001	0.003
N:1	0.108	0.121	0.010	0.018	0.031	0.019
Total	1.000	1.000	0.014	0.014	0.011	0.016
B: Machine Labor						
1:1	0.458	0.563	0.436	0.029	0.461	0.033
1:*M*	0.146	0.172	0.558	0.058	0.538	0.068
N:*M*, *N*>1, *M*>1	0.024	0.038	0.593	0.070	0.518	0.068
N:1	0.037	0.070	0.757	0.051	0.764	0.060
0:1	0.334	0.158	0.360	0.014	0.333	0.020
Total	1.000	1.000	0.444	0.034	0.477	0.040

Source: Computed from a digitized version of the Hand and Machine Labor study, see text and US Department of Labor (1899).
Notes: The unit of observation is a task as described by the staff of the Hand and Machine Labor study. The basic sample size in Panel A is 7,152 hand tasks from 610 production units. The basic sample size in Panel B is 12,473 machine labor tasks from 610 production units. 1:0 transitions are the hand labor tasks that disappeared in the transition to machine labor. 1:1 transitions are hand labor tasks that have a unique counterpart in machine production. In a 1:*M* transition, a single hand labor task subdivides into *M* machine tasks. In an *N*:*M* transition, *N* hand labor tasks transition into *M* machine labor tasks, with *N* > 1 and *M* > 1. In an *N*:1 transition *N* hand tasks combine to a single machine task. 0:1 indicates new tasks under machine labor. For equal weights, observations count equally in determining the average task shares within units. For time weights, observations are weighted by completion time in determining the average task shares within units.

were mechanized, whether by steam or water, similarly weighted. The sample used to compute Table 1 covers 610 of the original 626 manufacturing units.[6] For the most part, our discussion of Table 1 focuses on the equally weighted (rather than the time-weighted) statistics in Table 1.

Although some hand tasks were abandoned in the transition to machine labor, these comprised a small share of hand tasks and of the time spent in hand labor. The largest category of transitions by far was 1:1—that is, the agents were able to match a singleton task in hand production with a singleton task in machine production whose content was deemed to be the same, except that in machine production of the product, the task was far more likely to be mechanized. As can be seen from

[6] We excluded units from foreign countries, those that used horses, and which were otherwise missing data necessary for our analyses.

Table 1B, about 46.5 percent of 1:1 tasks in machine production used inanimate power; of these, 94 percent (0.436/0.465) were steam-powered.

As examples of mechanization in 1:1 transitions, consider the relatively large number of tasks involving trimming excess leather at various stages (for example, operations 33, 117, 133, and 135) or "skiving" leather where it overlapped to reducing bulk (operations 17–21, 54, 63, 71, 82, 88–93, and 135) in the production of boots and shoes, such as in Unit 71. Many patents were issued for tools and machines to facilitate these activities (see, for example, patents issued for leather trimming and skiving prior to 1874 (US Patent Office 1874). Under hand production, these activities were accomplished with a sharp (sometimes specialized) knife, guided by hand and eye. Under machine production, however, the knives were built into powered machines (for example, see US Patent 609868A granted 1898; also see the YouTube https://www.youtube.com/watch?v=Jv-nE1ixiB0 for the mechanics of a belt-driven skiving machine that is more or less contemporaneous with the Hand and Machine Labor study). These operated at high speed, allowing little chance of recovery if the product was wrongly placed, and with considerable risk to the operator. The operation of the trimming machine (and trimmer), for example, is described as follows (in Goldmark 1912, p. 65):

> [It] consists of a sharp knife edge, operating constantly against a sharp edged revolving top. The man who works the machine stands, holding upside down somewhat below the level of his eyes, the partly made, still unsoled shoe. He turns it skillfully and rapidly on the revolving top, against whose sharp edge the second knife-blade operates, cutting off all the surplus crimpled leather. The work is extremely rapid and absolutely uniform. But it takes skill and close attention. The machine could easily cut off too much, or could cut into the upper, if the swift handling of the shoe were not absolutely correct.

The more complex transitions that involved subdividing hand tasks (1:*M*) or consolidating (*N*:1) them (or possibly both (*N*:*M*)), were less common but by no means unusual. Keeping with the example of producing medium grade, laced shoes (Unit 71), machine production subdivided the selection of stock into several tasks (1, 23, and 35) because the various parts of a pair of shoes no longer came from a single hide as they would for hand-made shoes. Selecting various parts by look and feel so that the pattern and texture of the leather was similar improved the appearance of the finished product, making machine-made shoes look more like their hand-made counterparts. Mechanization, however, sometimes led to the consolidation of separate tasks. For example, the development of a super heavy-duty sewing machine (such as patent US 502873, granted to J. E. Bertrand in 1893) allowed the outsoles to be attached to the welts directly while locking the shoe shank in place. This reduced what had been two separate tasks (124a and 124b) in handicraft production (where the shoemaker used two variations on a single tool—an awl, round for locking the shank in place and square to attach the outsole) to a single machine operation (that is, a *N*:1 transition, or consolidation). From the perspective of hand labor, about 28 percent of hand tasks, on average, fell into the 1:*M*, *N*:*M*, or *N*:1 transitions; the corresponding figure from the perspective of machine labor was smaller, about 21 percent. As Panel B shows, the more complex transitions, especially consolidations, were also more likely to be mechanized by steam.

An important finding in Panel B is that new tasks were one-third of all tasks in machine labor, a much higher fraction than the share of hand labor tasks that were abandoned. Compared with the other tasks performed by machine labor, these new tasks were considerably less likely to use steam power, although the overall rate of steam use in new tasks, about 36 percent, was still substantial in an absolute sense. Many of these tasks were themselves directly related to that power source: engineers and firemen, for example, represented 15 percent of these new tasks. However, the more important group of new nonpowered tasks in machine production were those related to monitoring of the workplace activities (for example, the task of foreman/supervisor) and inspection of the finished product (for example, inspector, examiner, packer, and finisher). These activities made up about 20 percent of the 0:1 tasks and were essential to the smooth flow of the production line and the quality of the final product given that no single worker or group thereof assumed responsibility for the outcome of the production process.

The relative importance of new tasks declines when the data are weighted by time, indicating that many of the new tasks were relatively brief in duration. Even allowing for this, however, new tasks performed by machine labor accounted for a larger share of total production time than the share of time accounted for by old tasks in hand labor. We return to this point later in the paper when we consider our findings in light of Acemoglu and Restrepo's (2018) model of automation.

Although we have focused on task shares in Table 1, it is important to acknowledge that the absolute number of tasks increased from hand to machine labor. This increase is a direct manifestation of the increased division of labor that accompanied mechanization. As we described in detail in Atack, Margo, and Rhode (2017), it is possible to use the information in the Hand and Machine Labor study to compute a summary statistic of the division of labor—specifically, the proportion of tasks performed by the average worker. Multiplying by the number of tasks transforms the statistic into the average number of tasks per worker. The median number of tasks per worker in hand production was two, whereas in machine production, it was just one. In other words, the division of labor in machine production was virtually complete—if the HML study delineated a task, one or more workers were assigned to it and, on average, that was pretty much all they did, as far as the production of the specific good was concerned. As Marshall and Marshall (1881, p. 49) would note, "when the division of [labor] is carried very far a man's whole attention is concentrated on one operation …[and] such operations are performed … with a rapidity and an unerring accuracy …" We return to the division of labor below.

The Productivity Effects of Machine Labor and the Role of Steam Power

The standard way to measure labor productivity is by the flow of output over some period of time (for example, annually) divided by total labor hours over the same period. The Hand and Machine Labor study did not do this. Rather, for a standardized quantity of the specific good, the HML staff computed the amount of time each task took and then summed to get the total amount of time. Because a specific good is held constant (insofar as this is ever possible) while looking at hand and machine production, as is the standardized quantity, the overall productivity gain is simply the difference in total time between machine and hand labor. Given that the range of products considered were so very different, we do not compute the productivity gain in absolute units of time (say, hours) but rather calculate the logarithm of the ratio of machine to labor time, which is then averaged (equally weighted) across units. This average is -1.96. If we take the exponent, it is 0.14 [$e^{-1.96}$]—that is, on average, machine labor reduced total production time by a factor of seven (≈ 1/.14).

What accounts for these remarkable gains in productivity? In Atack, Margo, and Rhode (2017), we concentrated on the role of division of labor (to which we return in the discussion below). In this paper, we shift our attention to mechanization—that is, the use of an inanimate power source, in particular, steam power.

Economic historians have long had a keen interest in the diffusion of the steam engine and its attendant microeconomic and aggregate effects. These include the geography of steam adoption, changes in relative power costs in the face of technological innovation, externalities of steam power such as its role in fostering urbanization, and its impact on aggregate total factor productivity growth (Atack 1979; Atack, Bateman, and Weiss 1980; Kim 2005; Temin 1966). More recently, there have also been studies of how mechanization, whether steam power or electrification, affected the relative demand for different occupations (de Pleijt, Nuvolari, and Weisdorf 2018; Franck and Galor 2017; Gray 2013; Ojala, Pehkonen, and Eloranta 2016). By comparison, the Hand and Machine Labor study allows us to narrow the focus down to the task level for highly specific goods, comparing hand to machine production. This is straightforward to accomplish for the 1:1 overlap tasks; for these, we can difference the data at the task level within production units.

Table 2 reports our productivity regressions. Recall that "productivity" in the Hand and Machine Labor study is measured by the amount of time that it takes to complete a particular task in sequence in the making of a given amount of a specific good. If it is possible to change something to complete the given task more quickly than before—for example, use inanimate power—productivity has increased. We derive the regression specification from the following equation:

$$\ln T(i, j, k) = \alpha(i, k) + \beta\,(j, k) + \gamma \times (\text{Steam} = 1|i, j, k) + \delta \times (\text{Water} = 1|i, j, k) + \varepsilon(i, j, k)$$

The index i refers to the task; the index j, to the type of labor (j = hand or machine); and the index k, to the specific product or equivalently, what the HMLS staff called the "unit." Ln $T(i, j, k)$ is the log of the amount of time that it takes to complete task

Table 2
The Productivity Effects of Steam and Water Power Use in Machine versus Hand Production: 1:1 Task Transitions

Independent variable	*Dependent variable* ln (Time spent in machine task) – ln (Time spent in hand task) (1)	ln (Time spent in machine task) – ln (Time spent in hand task) (2)
ln (Time spent in hand task)		−0.36 (12.29)
Δ (Steam = 1)	−1.13 (19.29)	−0.84 (15.67)
Δ (Water = 1)	−0.35 (2.86)	−0.28 (2.42)
Adjusted R^2	0.52	0.61

Source: Authors.
Note: The sample consists of tasks in the 1:1 transition category for which there was complete information on the regression variables (N = 4,257). The dependent variable is the difference between machine and hand labor in the log of the amount of time that it took to complete the task. The mean value of the dependent variable is −1.74. See the text for the derivation of the regression equation. Both regressions in the table include unit fixed effects. Standard errors are clustered at the unit level. Absolute value of *t*-statistics shown in parentheses.

i for labor type j in unit k; the greater is T, the longer it took to complete the given task. The parameter α is a task–unit fixed effect; that is, it is indexed for task i and unit k but not for labor type j. In making α dependent on i and k but not j, we are assuming that, while some tasks might take proportionately longer than others for a given product, these relative differences are the same under both machine and hand labor. The parameter β is a labor type–unit fixed effect; it is indexed by j and k but not by i. This allows for the possibility that machine labor was more productive in general and that the productivity gain differed across specific products. Our main interest is in the parameters γ and δ, which are the log effects of steam and water power use and which we assume have the same values under machine and hand labor. If steam or water power use proportionately reduces the amount of time to complete a task, then $\gamma < 0$ and $\delta < 0$.

To estimate this regression, we difference between machine and hand labor within units for all variables measured at the task level. We can do this directly because for every 1:1 task under machine labor there is an exact match to a counterpart task under hand labor. After differencing, we have:

$$\Delta \ln T = \Delta\beta + \gamma \times \Delta\ (\text{Steam} = 1) + \delta \times \Delta\ (\text{Water} = 1) + \Delta\ \varepsilon$$

For ease of reading, we suppress the indexes but keep in mind that the unit of observation is the task. The dependent variable is the difference between machine and hand labor in the log of the amount of time it took to complete a task,

(Δln T). The right-hand side variables are product fixed effects ($\Delta\beta$), the differences in the steam and water power dummies between machine and hand labor at the task level, and the difference in the error terms ($\Delta\varepsilon$). The mean value of the dependent variable is –1.74. If we take the exponent of this mean value, it is 0.18 [= exp (–1.74)] or approximately 18 percent. That is, on average in the set of 1:1 transitions, a task under machine labor took 18 percent of the time to complete as the counterpart task under hand labor, indicating that labor was much more productive in completing the machine task than the equivalent hand task. Note that this mean value, –1.74, is smaller in absolute value than the analogous difference overall between hand and machine labor, -1.96, implying that the more complex transitions in Table 1 were, in an accounting sense, more important in generating overall productivity gains than were the 1:1 transitions.

As shown in column 1 of Table 2, the estimates of γ and δ are negative and highly significant, indicating the use of steam or water is associated with a reduction in time to complete a task. Relative to the mean value of the dependent variable (–1.74), the magnitude of the steam power coefficient (–1.13) in column 1 is quite large, suggesting a very large impact of steam use. By contrast, the coefficient for water (–0.35) is much smaller, although still statistically significant. The impact of water was more modest than steam, probably because of water's seasonality and storage constraints that limited its sustained flow.

Of course we cannot claim that these coefficients are causal; in particular, there may be omitted variables that are correlated with Δ Steam or Δ Water. One way to explore this possibility is to include the log of the amount of time the task took under hand labor as a right-hand side variable, as shown in column 2 of Table 2. If use of steam or water became more likely in machine labor for tasks for which hand labor was particularly unskillful (which we cannot observe directly), we would expect the absolute values of the steam and water power coefficients to be smaller in column 2 than in column 1. While this is the case, the coefficients of steam and water power use remain quite large and highly significant, suggesting that inanimate power use was, indeed, a major factor contributing to higher productivity under machine labor.

Discussion

We discuss our results in light of the recent paper by Daron Acemoglu and Pascual Restrepo (2018, and see also their paper in this symposium), which provides a formal task-based model for analyzing the effects of automation. In Acemoglu and Restrepo's model, tasks are ordered on a continuum along the unit interval from $N - 1$ to N in terms of the productivity of labor relative to capital. At date $t = 0$, capital costs are assumed to be lower than labor costs. Some tasks can be performed by either capital or labor, so if capital is sufficiently cheap, these will already be "automated" at date $t = 0$. However, other tasks might still be performed just by labor, simply because the technology is not sufficiently advanced for automation to occur. An improvement in technology, then, will induce additional automation to the new level of technical feasibility in the unit interval, or to the point where the firm is indifferent, on cost grounds, between capital and labor.

Their model also allows for new tasks to be created that are superior to existing tasks. The process by which this occurs is independent from changes in automation. The assumptions in the model ensure that new tasks will appear at N^*, the new right endpoint of the unit interval, while abandoned tasks will come from the former left endpoint, up to $N^* - 1$. The entire unit interval, therefore, moves to the right.

The key implications of the Acemoglu and Restrepo framework concern the net impacts of automation and new task creation on labor demand. If automation occurs, there is a displacement effect—capital replaces labor in some tasks below the threshold. There is also a productivity effect. If overall output increases sufficiently, demand for labor in non-automated tasks will increase on net. If, on net, new tasks use more labor, labor demand will further increase. However, if new tasks use less labor compared with abandoned tasks, the net impact of task replacement is negative, reducing any positive net effect that automation might have otherwise through productivity gains.

We cannot use the Hand and Machine Labor data to "test" the Acemoglu and Restrepo model literally for three reasons. First, their model orders tasks in terms of labor's comparative advantage at performing them. This is not the same as the order that tasks are actually performed in production. Second, we cannot re-order the HML tasks in terms of labor's comparative advantage because this is not observed in the HML data. Third, tasks in the Acemoglu and Restrepo model are on a continuum, whereas the task descriptions in the HML study are written summaries of discrete activities—in effect, subsets of the tasks in the unit interval of tasks in the Acemoglu and Restrepo model. Even if the HML staff had somehow channeled the logic of an economic model from 125-odd years into the future and managed to collect information on labor's comparative advantage, this would refer to the discrete activity, not to points (tasks) on a unit interval.

Nevertheless, the Acemoglu and Restrepo model is still highly valuable as an interpretive framework. First, their displacement effect is obviously present in the Hand and Machine Labor data; inanimately powered machines did things that were previously done by hand using simple tools. In some cases, the machine task was a sped-up version of what hand labor did—a machine-powered sander or polisher, for example. But as shown by the N:1 transitions, multiple hand tasks were also consolidated into a single machine task, a complicated transition that cannot simply be described as a faster version of a single hand labor activity. As Marshall (1890, p. 112) observed in his *Principles:* "[M]achinery constantly supplants and renders unnecessary that purely manual skill, the attainment of which was, even up to Adam Smith's time, the chief advantage of division of [labor]. But this influence is more than countervailed by its tendency to increase the scale of manufactures and to make them more complex; and therefore to increase the opportunities for division of [labor] of all kinds." Moreover, the displacement effect must have been largest for the *N*:1 transitions, because the *N* hand tasks took nearly twice as long to complete (as a share of total time) than one machine task, a far larger amount of "labor-saving" than is evident in the other transitions. The *N*:1 transitions, as we noted earlier, were the most mechanized—the share of machine tasks using steam or water power—of all the transitions from hand to machine production.

Second, the productivity effect was enormous. While detailed data are lacking, there is little doubt the average annual hours of operation per establishment in manufacturing increased over the nineteenth century (Atack and Bateman 1992; Whaples 1990). Yet, as the Hand and Machine Labor study shows, the amount of time it took machine labor to complete a product was a mere fraction of the time it took machine labor. On an average annual basis, therefore, the increase in total output was an order of magnitude larger than the displacement effect per unit of output, implying a very large positive impact on labor demand.

Third, the net effect of the introduction of new tasks on labor demand appears to have been positive. This is because the share of time taken up by new tasks in machine labor was larger than the share of time associated with hand tasks that were abandoned—indeed, five times larger. Among other activities, these new tasks included maintenance of steam engines, a foreman supervising large numbers of workers (discussed further below), and workers packaging products for distant markets.

The upshot is that the transition from hand to machine labor led to a vast expansion in the size of the manufacturing labor force, both in absolute number and as a proportion of the national aggregate. This was because, not in spite, of an equally vast increase in productivity, such that by the end of the nineteenth century, output per worker in US manufacturing was twice the level in Britain or Germany (Broadberry 1998). As we have noted, a long literature in economic history and economics asserts that the diffusion of steam power was a major factor behind the increase in productivity, but never, until the regression analysis in this paper, has this been demonstrated for individual production tasks.

However, our analysis also shows that steam power was not the full story. In our earlier paper, Atack, Margo, and Rhode (2017), we studied the overall difference in productivity between machine and hand labor at the unit-level, rather than task-level. Because we were analyzing differences across units rather than across tasks within units, we could include measures of the overall division of labor in the relevant regressions. We found that direct measures of the division of labor—specifically, the fraction of total tasks performed by the average worker and the number of tasks—fully account for the positive effects of overall scale, as measured by the number of workers. Unlike the regressions in Table 2 of this paper, those in our earlier paper do not control directly for steam (or water) power, but instead have a dummy variable for hand production. The coefficient of this dummy variable is positive and significant, implying that, once we control for the division of labor, other factors associated with machine labor compared with hand labor, such as greater use of steam power, contributed to overall productivity gains. Our results in Table 2 here are fully consistent with this interpretation.

The point we wish to make here is that, as useful as it is as an overall framing device, the Acemoglu and Restrepo model omits a fundamental feature of historical industrialization—namely, its extensive division of labor. As far as that model is concerned, the individual workers who perform tasks before and after automation could be the same people.

In point of fact, however, they were not the same people. In the tiniest shops that are iconic depictions of hand production in early manufacturing, the artisan was highly skilled in the sense of performing most or all of the production tasks from start to finish, as well as "nonproduction" tasks associated with managing the business. In the transition to machine labor, the artisan shop was displaced by the factory, which was different in many ways that could perhaps be summarized as "more" of everything—more capital, more labor, and more output. Establishments grew in size and complexity, an evolution that spawned the rise of a white collar labor force to oversee it—a "visible hand" in Alfred Chandler's (1977) memorable phase.

Our concern here is not so much the rise of the modern corporation *a la* Chandler, but rather what labor historians call "deskilling." Examples of deskilling are everywhere to be found in the data from the Hand and Machine Labor study. We have already cited the example of shoemaking; another example is blacksmithing—previously, this involved making rakes (Unit 30), most of the assorted carriage and wagon products (units 140–185), tools, and various other metal goods. The "village smithy," fashioning metal objects like pots, pans, plows, and numerous other objects from iron, could be found in small towns and in the countryside all over the United States as late as 1850. Atack and Margo (2019) use census data to study the relative decline of blacksmithing as a "hand trade" over the second half of the nineteenth century. Machine production led to establishments specializing in, for example, agricultural implements. These establishments were much larger in terms of employment than blacksmith shops, and far more productive in making plows, rakes and hoes, and related tools. Faced with such competition, blacksmith shops either shifted away from making objects to fixing them by offering repair services, or simply disappeared. The job of blacksmithing was once considered sufficiently numerous to warrant its own industry classification, but by the very end of the nineteenth century it was dropped from the manufacturing census as no longer worth the trouble to enumerate.

The point we are emphasizing, however, is not deskilling per se, but rather that the extent to which individual workers might be specialized in allocating their labor across tasks has important implications. The massive division of labor documented front and center in the Hand and Machine Labor study dramatically affected the nature of the human capital investment decision facing successive cohorts of American workers contemplating whether to enter the manufacturing sector. Earlier in the nineteenth century, the human capital investment problem such workers faced was mastering the diverse set of skills associated with most or all of the tasks involved in making a product, along with managing the affairs of a (very) small business, an artisan shop. The human capital investment problem facing the prospective manufacturing worker in the 1890s was quite different. There was little or no need to learn how to fashion a product from start to finish; mastery of one or two tasks would do, and such mastery might be gained quickly on the job. The more able or ambitious might gravitate to learning new skills, such as designing, maintaining, or repairing steam engines, or clerical/managerial tasks, the demand for which had grown sharply as average establishment size increased over the century (Katz and Margo 2014).

For many decades in the twentieth century, specialization was economically beneficial to workers—the costs of learning skills were relatively modest and the return on the investment—a relatively secure, highly paid job in manufacturing—made that investment worthwhile. The prospect of widespread automation has arguably changed this calculus. No single "job" is safe and the optimal investment strategy may be very different—a suite of diverse, relatively uncorrelated skills as insurance against displacement by robotics and artificial intelligence. This is perhaps the sense in which the history of how technology affects jobs is not repeating itself, and "this time" really is different.

Concluding Remarks

To understand the effects of automation on jobs, a number of labor economists have turned away from traditional "black box" models of production and their assumptions of relative complementarity or substitutability between capital and different types of labor. Instead, production is modeled as a collection of tasks, some of which might be performed by labor or automated with capital. Empirical assessments of these models have generally been indirect, in part because the data demands are so formidable. Even in today's world awash in "big data," information on production is rarely recorded at the task level. In the absence of such data, analysts must infer the task content of jobs indirectly through the use of, for example, the *Dictionary of Occupational Titles* (US Employment Service 1991).

This paper has reported on some preliminary but ongoing analyses of the US Department of Labor's Hand and Machine Labor (US Department of Labor 1899) study. The study has been long known by economic historians—but almost never used because the data were, until recent advances in information technologies, too complex to analyze. Our analysis of the HML data confirms the modern view that the "machine age" was transformative. It also reveals, however, that current task-based models of automation need elaboration to take into account certain effects of mechanization on labor that were historically relevant, like the division of labor.

The modern debate over automation and labor frequently invokes historical antecedents, most notably the steam engine during the early industrialization. Typically, historical evidence serves as anecdote to provide a context against which qualitative predictions can be made. For example, the steam engine was revolutionary in its time, and in retrospect it is clear that it "destroyed" some jobs but created many others. However, the extent to which the disruptive effects of the mechanization of the past serves as a prologue to the technologies of the present or future, or whether the modern technologies of robotics and artificial intelligence are fundamentally different in some way, remains an open question. It is intriguing to imagine how artificial intelligence might reduce the cost of reassigning and reorganizing tasks, allowing for more efficient dynamic optimization of production in response to changing conditions. Models that allow for such shifts of tasks and alterations in the division of labor may pay a useful role in understanding the technological shifts to come.

■ *We thank Daron Acemoglu, Hoyt Bleakley, Zorina Khan, Pascual Restrepo, Elyce Rotella, the JEP editors, and seminar participants at New York Stern, University of Michigan, and Yale University for helpful comments.*

References

Acemoglu, Daron, and David H. Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Chap. 12 in *Handbook of Labor Economics,* vol. 4B, edited by Orley Ashenfelter and David Card, 1043–171. Amsterdam: Elsevier-North Holland.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108(6): 1488–542.

Atack, Jeremy. 1979. "Fact or Fiction? The Relative Costs of Steam and Water Power: A Simulation Approach." *Explorations in Economic History* 16(4): 409–37.

Atack, Jeremy, and Fred Bateman. 1992. "How Long Was the Workday in 1880?" *Journal of Economic History* 52(1): 129–60.

Atack, Jeremy, Fred Bateman, and Robert A. Margo. 2008. "Steam Power, Establishment Size, and Labor Productivity in Nineteenth Century American Manufacturing." *Explorations in Economic History* 45(2): 185–98.

Atack, Jeremy, Fred Bateman, and Thomas Weiss. 1980. "The Regional Diffusion and Adoption of the Steam Engine in American Manufacturing." *Journal of Economic History* 40(2): 281–308.

Atack, Jeremy, and Robert A Margo. 2019. "Gallman Revisited: Blacksmithing and American Manufacturing, 1850–1870." *Cliometrica* 13(1): 1–23.

Atack, Jeremy, Robert A. Margo, and Paul W. Rhode. 2017. "The Division of Labor and Economies of Scale in Late Nineteenth Century American Manufacturing." Paper presented at the NBER Development of the American Economy Summer Institute, July 2017.

Autor, David H. 2013. "The 'Task Approach' to Labor Markets: An Overview." *Journal for Labour Market Research* 46(3): 185–99.

Bellamy, Edward. 1888. *Looking Backward, 2000–1887.* Boston: Ticknor and Co.

Broadberry, Stephen N. 1998. "How Did the United States and Germany Overtake Britain? A Sectoral Analysis of Comparative Productivity Levels, 1870–1990." *Journal of Economic History* 58(2): 375–407.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies.* W. W.

Norton.

Chandler, Alfred D., Jr. 1977. *The Visible Hand: The Managerial Revolution in American Business.* Cambridge, MA: Belknap Press.

de Pleijt, Alexandra, Alessandro Nuvolari, and Jacob Weisdorf. 2018. "Human Capital Formation during the First Industry Revolution: Evidence from the Use of Steam Engines." CEPR Working Paper 12987, Centre for Economic Policy Research.

Franck, Raphaël, and Oded Galor. 2017. "Technology-Skill Complementarity in Early Phases of Industrialization." NBER Working Paper 23197.

Giedion, Sigfried. 1948. *Mechanization Takes Command: A Contribution to Anonymous History.* New York: Oxford University Press.

Goldmark, Josephine. 1912. *Fatigue and Efficiency: A Study in Industry.* New York: Russell Sage Foundation.

Gray, Rowena. 2013. "Taking Technology to Task: The Skill Content of Technological Change in Early Twentieth Century United States." *Explorations in Economic History* 50(3): 351–67.

Hounshell, David A. 1984. *From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States.* Baltimore: Johns Hopkins University Press.

Katz, Lawrence F., and Robert A Margo. 2014. "Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective." Chap. 1 in *Human Capital in History: The American Record,* edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo. University of Chicago Press.

Kim, Sukkoo. 2005. "Industrialization and Urbanization: Did the Steam Engine Contribute to the Growth of Cities in the United States?" *Explorations in Economic History* 42(4): 586–98.

Marshall, Alfred. 1890. *Principles of Economics.* London and New York: Macmillan.

Marshall, Alfred, and Mary Paley Marshall. 1881. *The Economics of Industry.* London: Macmillan.

Ojala, Jari, Jaakko Pehkonen, and Jari Eloranta. 2016. "Deskilling and Decline in Skill Premium during the Age of Sail: Swedish and Finnish Seamen, 1751–1913." *Explorations in Economic History* 61: 85–94.

Rockoff, Hugh. 2019. "On the Controversies behind the Origins of the Federal Economic Statistics." *Journal of Economic Perspectives* 33(1): 147–64.

Temin, Peter. 1966. "Steam and Waterpower in the Early Nineteenth Century." *Journal of Economic History* 26(2): 187–205.

Thomson, Ross. 1989. *The Path to Mechanized Shoe Production in the United States.* Chapel Hill: University of North Carolina.

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

US Bureau of Labor. 1886. *First Annual Report of the Commissioner of Labor, March 1886: Industrial Depressions.* Washington DC: US Government Printing Office.

US Bureau of Labor Statistics. 1890–1905. "RG 257 Records of the Bureau of Labor Statistics, Records of the Department of Commerce and Labor, Reports of Bureau Activities, 1890–1905 Box No. 1, Scrap Book: Memo and Compilations." National Archives and Records Administration.

US Congress, House of Representatives. 1894. *Effects of Machinery on Labor.* Washington DC: US Government Printing Office.

US Department of Labor. 1899. *Thirteenth Annual Report of the Commissioner of Labor, 1898: Hand and Machine Labor.* Washington DC: GPO. Available from HathiTrust.org: http://hdl.handle.net/2027/nnc1.cu08593957.

US Employment Service. 1991. *Dictionary of Occupational Titles.* Washington DC: US Government Printing Office.

US Patent Office. 1874. *Subject-Matter Index of Patents for Inventions.* Washington DC: US Government Printing Office.

Wells, David Ames. 1889. *Recent Economic Changes, and their Effect on the Production and Distribution of Wealth and the Well-Being of Society.* New York: D. Appleton.

Whaples, Robert. 1990. "The Shortening of the American Work Week: An Economic and Historical Analysis of Its Context, Causes, and Consequences." Thesis (PhD in Economics), Graduate School of Arts and Sciences, University of Pennsylvania.

Wright, Carroll D. 1900. "Hand and Machine Labor." In *Gunton's Magazine*, 209–17. New York: Political Science Publishing Co.

Zeira, Joseph. 1998. "Workers, Machines, and Economic Growth." *Quarterly Journal of Economics* 113(4): 1091–117.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 71–88

The Rise of Robots in China

Hong Cheng, Ruixue Jia, Dandan Li, and Hongbin Li

China is the world's largest user of industrial robots. In 2016, sales of industrial robots in China reached 87,000 units, accounting for around 30 percent of the global market. To put this number in perspective, robot sales in all of Europe and the Americas in 2016 reached 97,300 units (according to data from the International Federation of Robotics). Between 2005 and 2016, the operational stock of industrial robots in China increased at an annual average rate of 38 percent.

In this paper, we describe the adoption of robots by China's manufacturers using both aggregate industry-level and firm-level data, and we provide possible explanations from both the supply and demand sides for why robot use has risen so quickly in China. Our focus is on the manufacturing sector, which is responsible for over 80 percent of China's industrial robot use.

We begin by documenting the rising importance of China in the global robot market. We show that the industrial composition of robot adoption in China emphasizes the same industries as other major robot markets like Japan, the United States, South Korea, and Germany: automotive and electronics. Also, using

■ *Hong Cheng is Professor of Economics and Management at Wuhan University, Wuhan, China. Ruixue Jia is Assistant Professor of Economics at the School of Global Policy and Strategy, University of California-San Diego, La Jolla, California, and Associate Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, Canada. Dandan Li is Assistant Professor at the Institute of Quality Development Strategy, Wuhan University, Wuhan, China. Hongbin Li is the James Liang Director of the China Program, Stanford Center on Global Poverty and Development, Stanford University, Stanford, California. Their email addresses are 78034431@qq.com, rxjia@ucsd.edu, lidandanwhu@yahoo.com, and hongbinli@stanford.edu. The corresponding author is Dandan Li.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

https://doi.org/10.1257/jep.33.2.71 doi=10.1257/jep.33.2.71

administrative data from China, we examine signs of the sharp rise in production of robots in China, including some preliminary evidence from robot production and research firms and robotics-related innovation patents. We discuss how robot use in China's manufacturing sector is growing against the backdrop of two other factors: labor costs and government policies. The rise of robots in China coincides with the declining growth of the working-age population and rising wages. In this respect, the robot revolution in China reminds us of how high labor costs accompanied the Industrial Revolution in 18th-century Britain (as discussed in Allen 2009). At the same time, China's government has identified the robotics industry as a strategically important sector (along with artificial intelligence and automation), and has initiated various programs and subsidies to encourage the use of robots as a way of transforming and upgrading China's manufacturing industries.

Research on firm adoption of robots is often hindered by the lack of firm-level data. Several recent studies have investigated the link between robots and jobs using data aggregated at the country or industry level (for example, Graetz and Michaels 2018; Acemoglu and Restrepo 2017, 2018). The lack of firm-level information has often precluded more in-depth analysis (Seamans and Raj 2018). In fact, we have found no prior research on firms' robot adoption behaviors in any country. Thus, a key contribution of this paper is that we have collected some of the world's first data on firms' robot adoption behaviors with our China Employer-Employee Survey (CEES), which contains the first firm-level data that is representative of the entire Chinese manufacturing sector. After a brief introduction and overview of this data, we then discuss some of the firm-level patterns in robot adoption.

We find wide variations in China's adoption of industrial robots both across and within industries. We look at correlations between firms' decisions to adopt robot technology and a selection of variables: government connections that might encourage robot purchases; market factors that could influence robot adoption, such as labor costs, concern over product quality, and expanding production; and whether robot adoption is associated with firms in which the employees are more likely to be doing certain tasks. We find that several market and government factors are associated with robot adoption, and that firms requiring more manual tasks have a greater likelihood of robot adoption. We also investigate whether factors that influence a firm's robot adoption are different from those that influence a firm's general machinery usage. These findings suggest that it may be valuable for future research to study how different dimensions of labor costs and job task characteristics affect the use of robots versus general machinery.

Given the aggressive promotion of robot adoption and production via industrial policies, it seems that the Chinese government does not fear the consequences of this disruptive technology. Similarly, in our interviews with employers and employees, we do not find that they are nervous about job replacement. In light of these perhaps surprising findings, we offer a few possible explanations for why China embraces robots, from the perspectives of the government, the employers, and the workers.

Table 1
Annual Robot Sales in China and the World

Year	World (1,000 units)	China (1,000 units)	China's share in the world (%)
1995	69.3	0.0	0.0
2000	98.7	0.4	0.4
2005	120.1	4.5	3.7
2010	120.6	15.0	12.4
2011	166.0	22.6	13.6
2012	159.3	23.0	14.4
2013	178.1	36.6	20.5
2014	220.6	57.1	25.9
2015	253.7	68.6	27.0
2016	294.3	87.0	29.6

Source: International Federation of Robotics (2017).
Notes: This table shows the rise of China in the world robot market, especially after 2013.

Robot Adoption and Production in China

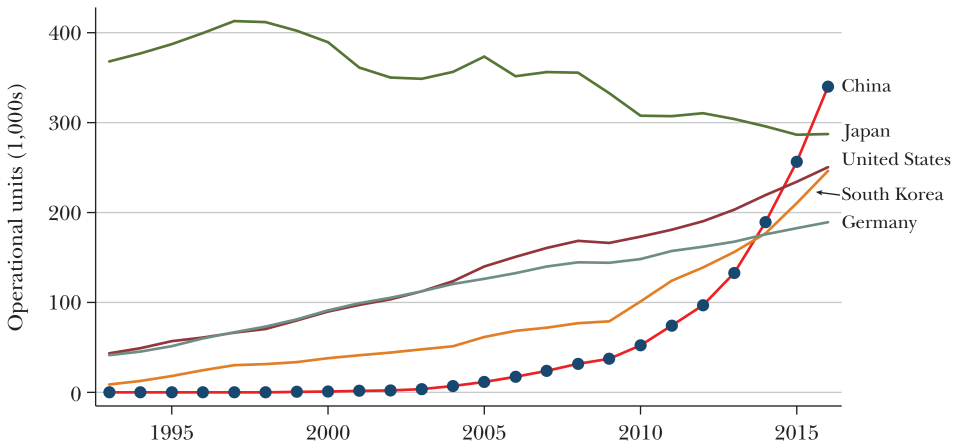
Robot Adoption

Annual sales of robots in China have risen dramatically (International Federation of Robotics 2017), as shown in Table 1. In 2000, a mere 380 units were sold in China, accounting for only 0.4 percent of the world total; China's share rose to 3.7 percent of annual global sales in 2005 and 12.4 percent in 2010. In 2016, sales further rose to 87,000 units, accounting for about 30 percent of the global market of 294,000 units.

Figure 1 shows the stock of operational robots for the top five robot markets in the world—China, Japan, the United States, South Korea, and Germany—which accounted for 72 percent of the world's operational robot units in 2016. China became the country with the largest operational robot stock in 2016, with 339,970 operational units—accounting for 19 percent of the total worldwide stock.

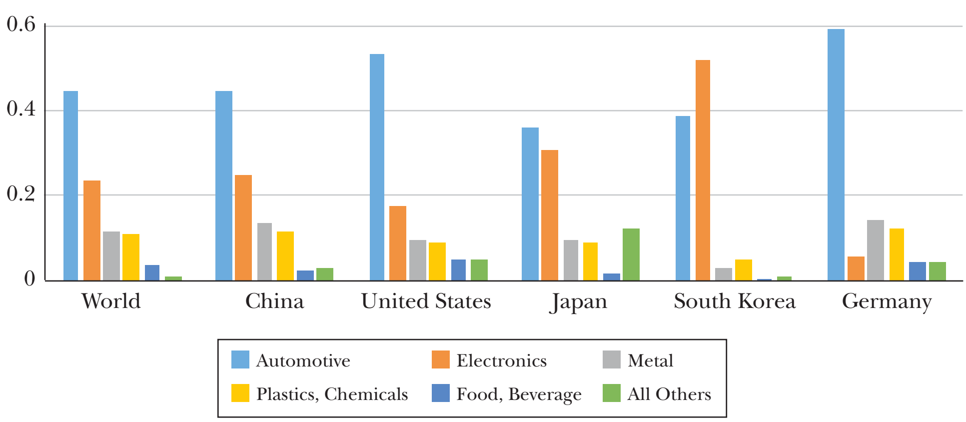
The distribution of robot usage across industries in China is similar to those of the other major markets for robots. In Figure 2, we plot the share of robots across industries in the manufacturing sector by major countries. Globally, the leaders in robot usage are the automotive and electronics industries (which accounted for 44.7 and 23.6 percent of all manufacturing robot usage in the world in 2015), followed by the metal (11.5 percent), plastic and chemical (10.8 percent), and food and beverage industries (3.7 percent). Textiles, wood and furniture, paper production, and glass and ceramics are among the industries in which robot adoption is still rare, and these industries are grouped in the "other" category, altogether accounting for only 1.3 percent of all manufacturing robots. There is some variation across countries. For instance, the share of robots used in South Korea for electronics and in Germany for auto production appears higher than in the other four countries. Nevertheless, China does not look systematically different in terms of industry-level robot adoption. In particular, the top industries in China for robot adoption are also automotive (accounting for 44.5 percent of

Figure 1
Stock of Operational Robots in Major Countries 2016



Source: Data is from International Federation of Robotics (2017).
Notes: This figure plots the operational stock of robots in the five major markets. China exceeded Japan and became the country with the largest operational robot stock in 2016.

Figure 2
Industrial Composition of Operational Robot Stock in Major Countries 2016



Source: Data is from International Federation of Robotics (2017).
Notes: This figure plots the share of robots across industries in the manufacturing sector by countries. China is not dramatically different from the other countries, suggesting that the supply of the technology matters in explaining which sectors use robots more.

all manufacturing robots), electronics (24.7 percent), metals (13.9 percent), plastics and chemicals (11.5 percent), and food and beverages (2.9 percent).

The higher rate of robot adoption in the automotive and electronics industries has implications for the future of robots in China. China has been the largest national producer of automobile units since 2008: indeed, since 2009, annual production of automobiles in China has exceeded that of the United States and Japan combined. China also clearly dominates the global electronics industry: over 70 percent of the world's computers and electronics are made in China. These industries in China seem likely to keep expanding, which implies that China will become an even more significant user of robots.

The variety of robots is also increasing in the Chinese market. Using data on 38 types of applications from the International Federation of Robotics, we construct a Herfindahl–Hirschman index to capture the variety of robots by their applications.[1] For the entire world, this index remained relatively stable at the level of 0.10 to 0.11 between 2005 and 2015. In contrast, the index for China decreased from 0.16 in 2005 to 0.10 in 2015, implying that applications of robots broadened within a decade. In 2005, the top four applications (handling operations and machine tending, plastic molding, welding and soldering, and arc welding) accounted for 75 percent of the market; in 2015, the share of the top applications (handling operations and machine tending, welding and soldering, spot welding, and fixing and press fitting) dropped to 54 percent of the market. Once again, this change shows that the variety of robots in China has increased.

Robot Production and Innovation by Firms

The rise of robot production in China is no less striking than that of robot adoption, although this increase is more recent. In 2012, only about 5,800 robots were produced in China (based on our reading of various government reports). By 2017, however, the number of robots produced in China annually had risen more than 20-fold to 131,000, among which 29 percent (37,800 units) were made by local (nonforeign) firms.

Unsurprisingly, the number of firms that produce robots, or do robotics research, has been rising fast. To our knowledge, little has been written on China's robot manufacturers and their technology. As a starting point, we examine the number of firms with "robotics" in their names by year, using firm registration data provided by the State Administration for Industry and Commerce (SAIC). In 2005, China only had 221 registered robotics firms, but by the end of 2015, the number had risen to 6,478. The year 2013 appears to have been a turning point for robotics manufacturers: the number of registered firms doubled each year during 2013–2015. It is unclear how profitable these firms are: after all, most of them are newly established. But based on public reports, government subsidies are a major driver of the rise of these manufacturers. In 2016, 40 percent of the net profits of the four publicly listed robotics firms—SIASUN Robot & Automation, Estun

[1] We calculate the Herfindahl–Hirschman index by squaring the market share of each application and then summing the resulting numbers.

Automation, Guangdong Topstar Tech, and Shanghai Step Electric Corporation—derived from government subsidies (as reported by Lin 2018).

China is also advancing rapidly in robotic technology. As a first effort of gauging China's progress in robot production technology, we examine innovation patents with "robotics" in their titles granted by China's State Intellectual Property Office (SIPO). In 2000, SIPO granted only 54 innovation patents with "robotics" in their titles, but the number rose to 319 in 2010 and 1,145 in 2015. The annual growth of the number of robotics-related innovation patents was around 40 percent during 2005–2015.

Labor Force and Government Subsidy

Although China's original success as the "world's factory" was built upon the cheap labor of hundreds of millions of manufacturing workers, China has now been experiencing the combination of a shrinking labor force and rapidly rising labor costs over the last decade (as discussed in this journal by Li, Li, Wu, and Xiong 2012). In turn, this has led to an increase in both economic pressure and political support for growth of the Chinese robotics industry.

China's past economic growth was in part driven by a "demographic dividend"—a rise in the working-age population as a share of China's total population. However, China is rapidly approaching a demographic deficit. China's working-age population (age 15–64) is declining both in absolute size and as a share of China's overall population. The annual increase in China's working-age population peaked in 2003 at around 17.7 million, but it then started declining and turned negative in 2015. Interestingly, the timing for the rise of robots roughly corresponds to that for the declining labor force; that is, the rise of robots started in 2003, when the gains in the working-age population started to decline, and accelerated since 2010, rising even faster since 2015, when the size of the working-age population declined outright.

During 2005–2016, the importance of manufacturing employment has been gradually increasing. In 2005, among the 746 million individuals in the labor force, 62 million (8.3 percent) were employed by the manufacturing sector; in 2016, among the 776 million workers in the labor force, 103 million (13.3 percent) were employed by the manufacturing sector. An important factor underlying this increase is that rural workers have moved to the manufacturing sector in urban areas.

Besides the size of the labor force, the skill composition of the labor force is also changing, especially due to the large-scale expansion of college enrollments from 1999 to 2009, which increased the number of college students by an average of 18 percent each year (Li, Loyalka, Rozelle, and Wu 2017). In 2005, only 6.6 percent of the labor force and 7.6 percent of manufacturing workers had a college education. These numbers rose to 18.1 percent for the whole labor force and 15.8 percent for manufacturing workers in 2016. Although it is impressive to see that the college education share doubled for the manufacturing sector within a decade, this change is actually smaller than that for the whole labor force, likely reflecting the difficulties that China's manufacturers have in attracting workers with college education.

Accompanied by the change in labor force, the wages of urban workers are also rising. During 2005–2016, China's average annual growth rate in real wage was 10 percent (deflated by China GDP deflator) for those employed by urban units, and the annual wage growth rate for the manufacturing sector was 9.7 percent. Manufacturing labor costs per hour in China were estimated to be $3.30 (in US dollars) in 2015, which is higher than those in Malaysia, India, Thailand, Indonesia, and Vietnam (Giffi, Rodriguez, Gangual, Roth, and Hanley 2016). Thus, to deal with the challenges of a labor shortage and rising labor costs, China's manufacturers have experienced pressures to automate, use machinery, and adopt robots.

In addition, the Chinese government has aggressively promoted the production and use of industrial robots in recent years. In 2013, for example, the Ministry of Industry and Information Technology (MIIT) released its "Guidance on the Promotion and Development of the Robot Industry." Some goals outlined in the report included developing 3–5 world-leading robot companies and 8–10 supporting industrial clusters; increasing China's global market share of high-end robot products to more than 45 percent; and promoting the use of robots in factories with the aim of a density of 100 robots per 10,000 workers. These initiatives were further bolstered by the launch of the "Made in China 2025" program in the year 2015, which set national goals of producing 100,000 industrial robots per year and achieving a density of 150 robots per 10,000 workers by 2020, which would triple the robot density in the manufacturing sector reported for 2015 (State Council 2015). In addition, in 2016, the MIIT, the National Development and Reform Commission (NDRC), and the Minister of Finance jointly launched the Robotics Industry Development Plan (2016–2020) to promote robot applications to a wider range of fields including the service sector. This plan sets several targets by 2020, including 100,000 industrial robots annually produced by domestic technology and annual sales of ¥30 billion (about $4.4 billion in US dollars) for service robots.

Like Chinese industrial policies implemented in other areas (for example, the electric car and solar industries), the most common form of government support is subsidies, which appear to be effective (but not necessarily efficient) at steering firms into industries they might otherwise ignore.[2] To the best of our knowledge, no systematic data exist on subsidies from the Chinese government to finance the production and use of industrial robots, but numerous media reports have commented on the scale and size of these subsidies. At the local level, governments have set up some investment capital and allocated funds to support robot usage and innovation. Like any other type of policy in China, there is regional variation in policies supporting the adoption of robot technology related to factors like differences in fiscal capacity, regional industrial structure, and priorities of local leaders. As one example, in 2015 the government of Guangdong Province put together a fund of $150 billion (in US dollars) to encourage firms to invest in automation technology and promote robotics innovation (Yang 2017).

Contrasting with the negative sentiment about robots in many countries due to their potential to replace jobs, the overall perception of robots in China has always been positive. The threat of job replacement is rarely mentioned in the government documents promoting robot adoption and production. Instead of

[2] As usual, political economy is likely to matter. Providing subsidies creates more opportunities for rent-seeking than alternatives like providing training.

worrying about job replacement, the government emphasizes robot adoption as a way to deal with challenges in the labor force. One reason that Chinese see robotics (and automation) as a positive phenomenon is that advances in science and technology are believed by many to be essential for China's rise as a world power. This perception partly originates from China's painful early encounters with Western powers. Since the Opium War in the 1840s, China has endured numerous foreign invasions, which many have attributed to the inferiority of technology in China. The following discussion of the importance of technology in a 2016 national plan by the State Council is revealing:

> One important reason why China fell into backwardness and took beatings in the modern era is that the previous industrial revolutions slipped through our fingers, leaving us with weak technology and a weak state. To realize the great rejuvenation of the Chinese nationhood that is the Chinese Dream, we must make genuine use of science and technology, this revolutionary force and lever of power in the highest sense.

During the decade between January 2009 and January 2019, *People's Daily*—the flagship newspaper of the central government—published 346 reports related to industrial robots. In these reports, "industrial/technological/robot revolution" was mentioned 206 times, and "job replacement/unemployment" was mentioned 85 times. When examining the reports mentioning job replacement, we find that the sentiment is still generally positive. Below is an example from that paper (He 2016, translated from Chinese), which illustrates the reasoning of government officials:

> Since 2014, Dongguan City (a city in Guangdong Province) has implemented the "replacement of workers with robots" incentive (funding) policy to promote the transformation and upgrading of the manufacturing industry. As of the end of last year, the number of enterprises applying for these funds reached 1,262, with a total investment of over 10 billion yuan and a reduction of 71,000 jobs. But the practice has proved that it is a big misunderstanding that the robots will steal people's jobs and cause unemployment. He Yu, deputy mayor of Dongguan City, said that the city has made a serious analysis of employment. More than 75 percent of the enterprises that implement "replacement of workers with robots" either have not changed or have increased the number of workers. Like Zheng Zhangteng (a Chinese worker), affected by the "replacement of workers with robots" policy, a large number of front-line operators are liberated from the heavy and dirty working environment. After training, they are transferred to technical personnel positions, and they have upgraded their careers while upgrading their industries. Even if there were a small number of people who left their positions, they were immediately absorbed by other companies.

Given this background, China seems likely to lead the world in the volume and sales of robot adoption and production in the future. Current robot technology is most suitable in auto and electronics industries, and since China dominates global sales in both areas, more and more robots will be used in China. With China's declining labor force and rising wages, more of China's manufacturers will find it profitable to adopt robots. Furthermore, government industrial policies can induce additional demand. In the next sections, we employ micro-level data to further examine the patterns of robot adoption at the firm level.

The China Employer-Employee Survey

The China Employer-Employee Survey (CEES) is a new longitudinal study of manufacturing firms and workers in China. CEES was initiated by two of the authors (Hong Cheng and Hongbin Li) together with Yang Du at the Chinese Academy of Social Sciences and Albert Park at the Hong Kong University of Science and Technology. The survey is administered by the China Enterprise Survey and Data Center at Wuhan University, which is directed by Cheng and Li. It began in 2015 with a survey of firms and workers in the coastal province of Guangdong, which borders Hong Kong, and expanded to the interior province of Hubei in 2016. Guangdong has been China's most important industrial province in the past few decades and accounted for 13.4 percent of all manufacturing firms and 19.4 percent of all manufacturing workers in China in 2015. In 1980, when the central government initiated the Special Economic Zones policy, three of the four Special Economic Zones were located in Guangdong. In recent years, the manufacturing sector has been expanding to the interior provinces like Hubei. In 2015, Hubei accounted for 4 percent of all manufacturing firms and 6.6 percent of all manufacturing workers.

In this paper, we focus on the 2016 data (covering information on firm behavior in 2015), in which we began to include questions on robots in the survey instrument. For the most recent round of the survey, conducted in the summer of 2018, we followed up with the previously surveyed firms in Guangdong and Hubei and expanded the survey to include three additional provinces: Jiangsu, Liaoning, and Sichuan. This data is in the process of being entered and cleaned. We plan to make the CEES data available to researchers step by step. The existing data has also been used to study the performance of state-owned enterprises (Cheng, Li, and Li forthcoming) and management practice (Bloom, Cheng, Duggan, Li, and Qian 2018), where the authors provide a detailed description of other variables in the data.

Sampling

In 2016, the China Employer-Employee Survey was conducted in Guangdong and Hubei. We used the third National Economic Census, which was conducted in early 2014, as our sampling frame. Sampling was conducted in two stages, each using probability proportionate-to-size sampling, with size defined as manufacturing employment. In the first stage, 20 county-level districts were randomly sampled in each province, with probabilities proportionate to manufacturing employment in each district. In the second stage, 50 firms were sampled in each district as a target sample, again with probabilities proportionate to manufacturing employment in each firm. Enumerators then visited the 50 firms and attempted to survey the first 36 eligible firms (that have production activities in the sampled district). With this approach, the firm sample can be viewed as reasonably representative of manufacturing firms in China.

Employees were also randomly selected using stratification. We first asked each firm to provide a list of all employees enrolled at the end of the previous year, with middle and high-level managers listed separately. Then, we randomly selected ten employees in each firm (six to nine for smaller firms), three (two for smaller firms) of whom were middle and senior managers. If selected employees could not participate (for example, because they were not working on-site during the survey period), they were replaced with the closest name on the list of workers. This process continued until the targeted number of sampled employees was reached.

After excluding firms that were no longer in operation, there were 1,326 firms across 26 prefectures in Guangdong and Hubei that were eligible to be surveyed. In 2016, we managed to survey 1,115 firms and achieved a response rate of 84 percent. The median asset value of surveyed firms was ¥55.7 million (roughly $9 million in US dollars). The median number of workers across these firms was 160, with a 25th percentile of 55 employees and a 75th percentile of 520. About 90 percent of the initially sampled workers participated in the employee surveys. This provides us with information on 8,848 workers, among which 3,691 are production-line workers.

Robot Adoption across Industry and Region

We asked two sets of questions on robot adoption in the survey. First, we asked whether a firm utilized robots in its production processes in 2015. According to our data, 8.6 percent of the 1,115 firms used robots in 2015. Second, we asked questions related to the purchase of robots, namely, how much robots cost, and whether the government had subsidized the firm's purchase.

The responses reveal considerable variation in the adoption of robot technology across industries. Indeed, the share of robot units across industries in the International Federation of Robotics data versus the probability of using robots by industries in our CEES data have a correlation coefficient of 0.97, which provides a useful validity check of the quality of CEES data.[3] Such differences across industries also lead to differences in the use of robots by Chinese firms across regions. For example, we find that in Guangdong's Huizhou prefecture, where electronics manufacturing is the dominant industry, over 20 percent of sampled firms use robots. In contrast, no sampled firms in Hubei's Qianjiang prefecture use robots, likely because firms in this prefecture are generally involved in the garment and leather product industries where robot adoption is still rare.

However, our results also show substantial variation in robot usage across firms within a given industry. Indeed, we find that province-by-industry fixed effects

[3] This correlation is visualized in an online Appendix available with this paper at the journal website, where we also present more background information about the China Employer-Employee Survey, including summary statistics of the data for 2015.

(where "industry" refers to the 12 industries defined by the International Federation of Robotics) can only explain 9 percent of the variation in firm-level robot adoption. Therefore, it appears useful to investigate firm-level correlates of robot adoption.

Patterns in the Adoption of Robots by Chinese Firms

In this section, we use a series of regressions to describe the patterns of robot adoption by Chinese firms. This evidence is primarily cross-sectional and descriptive. However, we believe it nonetheless sheds light on the rise of robots in China.

As a starting point, we have already noted that firms in the automotive and electronics sectors are more likely to use robots. Moreover, firm size and capital-labor ratio are also correlated with a greater probability of robot adoption. As shown by the binned scatter plots in Figure 3A, an increase in log number of workers (x-axis) by one standard deviation is associated with the rise in the probability of robot adoption (y-axis) by 8.3 percentage points, after controlling for province and industrial fixed effects. Similarly, Figure 3B shows that an increase in log capital-labor ratio by one standard deviation is associated with the rise in the probability of robot adoption by 3.7 percentage points. These correlations still hold when we run a horse race test between firm size and capital-labor ratio, suggesting that both factors are relevant.

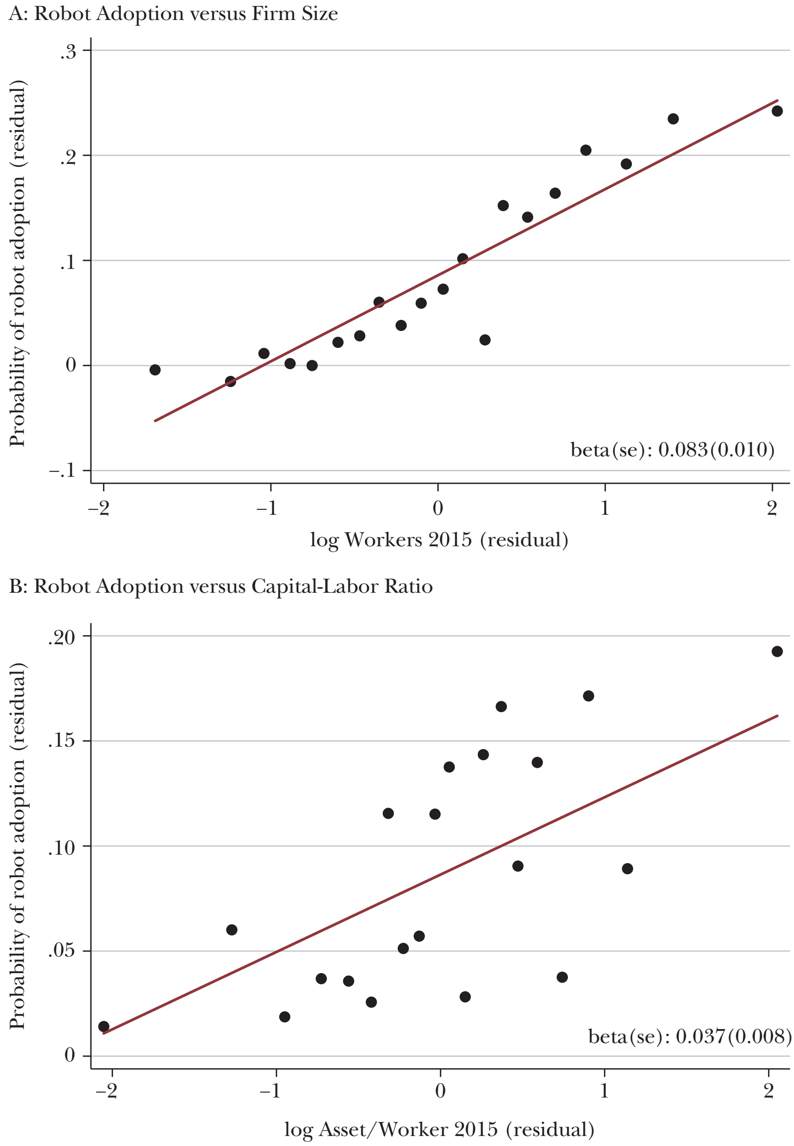
Next, we examine patterns in firms' robot adoption that reflect factors other than industry, firm size, and running a capital-intensive plant. In particular, in Table 2, we summarize the correlations between robot adoption and factors we care about in regressions that control for province and industry fixed effects, log number of workers, and log capital-labor ratio. In essence, we are comparing firms with similar size and capital-labor ratio within the same industry and province. We look at 12 different factors, some involving government, some involving market factors, and some involving the mix of tasks at the firm. The coefficient in each cell of Table 2 is generated by a separate regression.

Government Policy

One possibility is that China's firms adopt robots because of government policies that facilitate or subsidize robot purchases. In the survey, we asked whether firms receive subsidies specific to robot adoption. Among all robot-using firms, 15 percent answered "yes," suggesting that government industrial policies may have contributed to their adoption decisions.

It is also important to consider whether politically connected firms might be more likely to adopt robots either because of their better access to government funding or because of their tendency to comply with government policies. To examine the potential influence of political connections on firms' robot adoption behavior, we first examine the impact of firm ownership. In our sample, 12 percent of firms are state-owned enterprises (which include those with collective ownership). The coefficient reported in cell 1 of Table 2 does not suggest that state-owned enterprises are more likely to adopt robots than other firms. If anything, they are

Figure 3
Robot Adoption: Firm Size and Capital-Labor Ratio



Source: Authors.
Notes: This figure plots the correlations between firm size, capital-labor ratio and robot adoption, after controlling for province fixed effects and industry fixed effects. Log workers and log capital-labor ratio are standardized so that 1 (and –1) means one standard deviation above (and below) average. Each dot indicates a bin of firm observations. These correlations also hold when we run a horse race between these two factors.

Table 2
The Correlation between Robot Adoption and Firm Characteristics
(Dependent Variable: Robot Adoption = 1 if yes, 0 if no)

	(1) SOE	*(2)* CEO: Party member	*(3)* ln (Wage cost) std	*(4)* Union	*(5)* ln (Workers leaving voluntarily) std	*(6)* Quality control
coefficient	–0.042	0.024	0.031	–0.020	0.027	0.006
s.e.	(0.025)	(0.018)	(0.014)	(0.017)	(0.013)	(0.011)

	(7) Defect rate std	*(8)* Exporter	*(9)* ln (Sales change) std	*(10)* Manual task std	*(11)* Routine task std	*(12)* Abstract task std
coefficient	0.003	0.023	0.011	0.015	0.006	-0.005
s.e.	(0.009)	(0.019)	(0.006)	(0.007)	(0.007)	(0.008)

Notes: This table summarizes the correlations between each factor and robot adoption, after controlling for province fixed effects, industry fixed effects, log number of workers, and log capital-labor ratio. The factors are: being a state-owned enterprise (1); the chief executive officer (CEO) is a member of the Communist Party (2); a one-standard-deviation increase in log wage (per worker) (3); labor union presence (4); a one-standard-deviation increase in the log of workers leaving voluntarily (5); having quality control (6); a one-standard-deviation in the increase of defect rate (7); being an exporter (8); a one-standard-deviation increase in sales growth (9); a one-standard-deviation increase in manual task (10); a one-standard-deviation increase in routine task (11); and a one-standard-deviation increase in abstract task (12). The coefficient in each of the 12 cells is generated by one independent regression. Robust standard errors are reported in the parentheses.

less likely to do so, as we find a negative correlation between robot adoption and being a state-owned enterprise. One possible reason is that they may be less responsive to market forces, which, we will show in the next section, are strongly correlated with robot adoption.

In addition, 35 percent of the firms in our sample have a chief executive officer (CEO) who is a member of the Communist Party—another indicator of political connectedness. This factor is positively correlated with robot adoption, as shown in cell 2 of Table 2. The correlation coefficient is not precisely estimated, but the magnitude is sizable: the CEO's party membership is associated with a 2.4 percentage point higher probability of robot adoption. This result becomes stronger after controlling for the CEO's gender, age, and education (not shown here). These results suggest that government policies and political factors should be considered when examining the robot adoption behaviors of Chinese manufacturing firms.

Market Factors: Labor Cost and Others

We now examine this question: To what extent is the adopotion of robots by Chinese firms correlated with market factors, such as the ability to decrease labor costs, improve product quality, and expand production?

Rising labor cost is often cited as a main motivation for robot adoption in China (for example, Bland 2016). We test this conjecture by using three measures of firm labor costs: the total wage bill (in logs) of a firm, whether a firm has a labor union, and the worker turnover rate. Because we control for firm size, the wage cost variable can be viewed as reflecting the average wage cost of a firm.

The wage bill is positively correlated with robot adoption, as shown in cell 3 of Table 2. Specifically, a one-standard-deviation increase in log wage (per worker) is associated with an increase in the probability of robot adoption by 3 percentage points (relative to the mean of 9 percent). While 60 percent of the firms have labor unions, we find no positive correlation between labor union presence and robot adoption (see cell 4 of Table 2), a result consistent with the understanding that labor unions in China lack independence and do not play a critical role in wage bargaining. The lack of strong and independent unions in China may also partly contribute to workers' tolerance of robot adoption.

We also examine how labor turnover affects robot use. Our data shows that voluntary turnover is much more common (with a mean of 0.31 of the annual workforce) than involuntary turnover (with a mean of 0.13). The data on "voluntary" and "involuntary" turnover is based on responses from those who answered the firm-level survey (that is, the managers and their team members). In our worker survey answered by the employees, we also find that voluntary turnover is more common than involuntary turnover. For instance, when being asked why they left their previous jobs, 61 percent of the workers answered that they left voluntarily because they got a better job or wanted to search for a better job, and another 21 percent left voluntarily for other reasons (like returning to their hometown or family matters). Regarding involuntary turnover, 14 percent answered that they left because their firms went out of production or got restructured, while 2 percent cited downsizing payrolls.

Voluntary worker turnover is positively correlated with robot adoption, while involuntary turnover is not. As shown in cell 5 of Table 2, an increase in log *voluntary turnover* by one standard deviation is associated with an increase of 2.7 percentage points in the probability of robot adoption. In contrast (but not shown on the table), there is no significant correlation between robot adoption and *involuntary turnover* (with a coefficient of -0.003 and standard error of 0.011). Because involuntary turnover is more likely to be the consequence of robot adoption while voluntary turnover is more likely to be the cause of robot adoption, these patterns suggest that few workers have been displaced as a result of robot adoption.

We also examine other market factors, such as quality control and production growth, on the likelihood of robot adoption. Firms may adopt robots to meet high quality standards. To examine the role of quality control, we consider whether a firm has a quality control strategy in place, the defect rate of products, and whether a firm is involved in exporting (assuming exported products are of higher quality). We also examine whether a firm is expanding production, as it is likely that high-growth firms may find it difficult to recruit a sufficient number of workers. For this reason, they may be more likely to employ robots than other firms. We use the growth of sales revenue over the past year as a measure of production growth.

As shown in cells 6 and 7 of Table 2, we find no evidence suggesting quality control is correlated with firms' robot adoption behavior, while in cells 8 and 9 we find weak evidence for the correlation between production growth and robot adoption. Of course, our variables do not perfectly measure either quality control or production growth, and we are looking at cross-section data for a single year rather than time-series data. We plan to revisit these hypotheses after we collect more data in future years.

Job Tasks

We next examine the extent to which job tasks are associated with firm adoption of robots, as certain tasks might be more suitable for industrial robots to complete than others. We assess tasks at a given firm using information collected at the worker level. Following the approach of Autor, Levy, and Murnane (2003), we asked detailed questions in the survey related to the task characteristics of each sampled worker. In our analysis, we employ a principal component analysis method to measure the degree to which the job/post of a worker requires a manual, routine, or abstract task. For each worker's job, we assign a value of 1 if it requires that type of task, and 0 otherwise. We then calculate the firm-level aggregated task measures by taking the average of worker-level task measures.[4]

Linking robot adoptions to these firm-level task measures, we find that robots are more prevalent at firms where employees are commonly doing manual tasks, but not those that require routine or abstract tasks. As can be seen in cell 10 of Table 2, the correlation between robot adoption and our manual task measure is positive and significant. In terms of the magnitude, an increase in the manual task measure by one standard deviation is associated with a 1.5 percentage point increase in the probability of robot adoption. In contrast, the correlations between routine/abstract tasks and robot adoption are small in magnitude and not significantly different from zero, as shown in cells 11 and 12. Although some may assume that robots were likely to replace routine tasks a priori, our data do not support this conjecture. One possible reason is that robots have taken on the manual, dirty and health-hazardous tasks, but at least so far have not been able to replace more delicate routine tasks in a cost-effective manner. In addition, because it is difficult for manual workers to express themselves in Chinese society, their voice on robot adoption is unlikely to be heard.

Results are similar when we include all 12 of these factors in the regressions, as shown in Table 3. In column 1 of Table 3, we report a regression with the same dependent variable (a zero-or-one variable indicating the adoption of robots), including all the firm characteristics together as independent variables. The patterns are similar to those we find in Table 2. In the second column of Table 3, we use the (log of) value of robots as an outcome variable and also obtain qualitatively similar results.

[4] These aggregate task measures are correlated as one might expect. The manual task measure is positively associated with the routine task measure, with a correlation coefficient of 0.33. The manual task measure is also negatively correlated with the abstract task measure, with a correlation coefficient of -0.22. For details of summary statistics see the online Appendix available with this paper at the journal website.

Table 3
Firm Characteristic, Robot Adoption, and General Machinery Usage

	Dependent Variable			
	(1) *Robot Use* *(0/1)*	*(2)* *ln (Robot Value)*	*(3)* *ln (Machine Value)*	*(4)* *(3) – (2)*
ln (Workers) std	0.039	0.341	1.308	–0.967
	(0.020)	(0.139)	(0.105)	(0.174)
ln (Asset/Worker) std	0.022	0.167	0.607	–0.440
	(0.009)	(0.062)	(0.063)	(0.077)
SOE	–0.023	–0.323	0.034	–0.358
	(0.027)	(0.168)	(0.167)	(0.234)
CEO: Party member	0.048	0.291	0.046	0.245
	(0.020)	(0.123)	(0.104)	(0.153)
ln (Wage cost) std	0.023	0.071	0.332	–0.261
	(0.015)	(0.111)	(0.091)	(0.142)
Union	–0.023	–0.140	0.204	–0.344
	(0.019)	(0.112)	(0.106)	(0.150)
ln (Workers leaving voluntarily) std	0.027	0.085	–0.007	0.092
	(0.013)	(0.078)	(0.064)	(0.095)
Quality control	0.003	–0.009	0.062	–0.071
	(0.012)	(0.073)	(0.106)	(0.123)
Defect rate std	0.000	0.002	–0.041	0.044
	(0.010)	(0.056)	(0.031)	(0.056)
Exporter	0.033	0.209	–0.153	0.362
	(0.020)	(0.132)	(0.108)	(0.160)
ln (Sales change)	0.011	0.061	–0.040	0.102
	(0.006)	(0.038)	(0.043)	(0.063)
Manual task std	0.013	0.093	–0.013	0.105
	(0.008)	(0.051)	(0.051)	(0.070)
Routine task std	0.000	–0.035	0.072	–0.106
	(0.008)	(0.050)	(0.052)	(0.070)
Abstract task std	0.002	0.023	0.154	–0.131
	(0.009)	(0.068)	(0.051)	(0.083)
Province, industry fixed effects	Y	Y	Y	Y
Observations	911	911	911	911
R_2	0.194	0.157	0.650	0.309

Notes: This table compares robot usage with general machinery usage. We find that the factors driving machinery usage are not the same as those driving robot adoptions. Besides log workers and log capital-labor ratio, the other variables are the same as those in Table 2. Column 4 tests the significance of these differences. Robust standard errors are reported in the parentheses.

Robots versus Other Machines

Next, do the factors correlated with the robot adoption behavior of firms differ from those that are correlated with the general use of machinery? To compare the correlations of different factors with the uses of machinery versus those with robot adoptions, we estimate a regression with the (log of) the value of machinery as the outcome variable and report it in column 3 of Table 3. We also report in column 4 the significance of the difference between robot adoption and machinery use.

There are indeed differences in the correlations between robot and machinery usage. First, although robot adoption is positively associated with firm size and the capital-labor ratio, the correlations are much smaller than those for general machinery. Second, while spending on robots is significantly correlated with the Communist Party membership status of the firm's chief executive officer, a firm's spending on general machinery is not. Third, the role of labor costs is mixed. On the one hand, wage costs have a larger impact on the use of general machinery than on the use of robots; on the other hand, worker turnover appears more important for the use of robots. Finally, replacing manual tasks is more important in explaining robot usage than general machinery usage. As reported in column 4, these differences are not always significant but are large in magnitude, suggesting that it is valuable for future research to study how different dimensions of labor costs and job task characteristics affect the use of robots as opposed to general machinery.

Conclusions

In this paper, we have sought to describe some key patterns in the rise of robots in China. At the aggregate level, the rise of robots has accompanied a decline in the growth of the working-age population and an increase in wages, suggesting that the rising cost of labor is one underlying driver of robot usage in China. Because China is a global leader in the production and consumption of automotive and electronics, the two leading industries in robot adoption, China probably will play an even more important role in the robot market in the future. The Chinese government's industrial policies are also likely to affect both robot adoption and production.

Using the China Employer-Employee Survey (CEES) data, we further provide firm-level evidence of the rise of robots in Chinese manufacturing firms. We believe that the evidence we have found on the roles of government and the market in driving the adoption of robot technology is particularly important. These analyses are some initial steps towards understanding the causes and consequences of the increasing use of robots in Chinese manufacturing. Such consequences include effects on firm productivity, complementarity/substitution between humans and robots, and other labor market outcomes.

At this stage, the threat of job replacement is not a high-priority concern in the mind of China's government or its citizens. Government policies are motivated by the challenges of labor costs and labor shortage, as well as the imperative to lead a new wave of Industrial Revolution. For employers, the labor force challenges are indeed important considerations for robot adoption, as shown by our analysis. For employees, the high voluntary turnover rates and the lack of strong and independent unions may partly contribute to their tolerance of robot adoption. It is conceivable, however, that the short-run consequences are different from those in the long run. We hope to make further progress on these questions by continuing to follow China's manufacturing firms.

■ *We thank the editors Gordon Hanson and Timothy Taylor, as well as Henrik Christensen, Roger Gordon, Barry Naughton, Rui Ruan, and Xin Wang for their comments and suggestions. We also thank our two collaborators on the CEES survey, Yang Du and Albert Park, as well as the hundreds of enumerators, collaborating firm managers, and workers involved in the survey.*

References

Acemoglu, Daron, and Pascual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets." NBER Working Paper 23285.

Acemoglu, Daron, and Pascual Restrepo. 2018. "Demographics and Automation." NBER Working Paper 24421.

Allen, Robert C. 2009. *The British Industrial Revolution in Global Perspective.* Cambridge University Press.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118(4): 1279–1333.

Bland, Ben. 2016. "China's Robot Revolution." *Financial Times,* June 6.

Bloom, Nicholas, Hong Cheng, Mark Duggan, Hongbin Li, and Franklin Qian. 2018. "Do CEOs Know Better? Evidence from China." NBER Working Paper 24760.

Cheng, Hong, Hongbin Li, and Tang Li. Forthcoming. "The Performance of State-Owned Enterprises: New Evidence from the China Employer-Employee Survey." *Economic Development and Cultural Change.*

Giffi, Craig A., Michelle Drew Rodriguez, Bharath Gangula, Aleda V. Roth, and Tim Hanley. 2016. "2016 Global Manufacturing Competitiveness Index." Deloitte Touche Tohmatsu Limited (DTTL) Global Consumer & Industrial Products Industry Group and the Council on Competitiveness.

Graetz, Georg, and Guy Michaels. 2018. "Robots at Work." *Review of Economics and Statistics* 100(5): 753–68.

He, Linpin. 2016. "The Migration of Industries from the Pearl Delta Region to Inland Areas: Can the Migrant Workers Stay Employed?" *People's Daily,* June 12 [in Chinese], http://finance.people.com.cn/n1/2016/0612/c1004-28425817.html.

International Federation of Robotics. 2017. "Robots: China Breaks Historic Records in Automation." IFR Press Release, August 16. https://ifr.org/ifr-press-releases/news/robots-china-breaks-historic-records-in-automation.

Kilian, Pamela Hartmann, Daniel Schiller, and Frauke Kraas. 2012. "Workplace Quality and Labour Turnover in the Electronics Industry of the Pearl River Delta, China." *Zeitschrift für Wirtschaftsgeographie* 56(1–2): 58–79.

Li Hongbin, Lei Li, Binzhen Wu, and Yanyan Xiong. 2012. "The End of Cheap Chinese Labor." *Journal of Economic Perspectives* 26(4): 57–74.

Li Hongbin, Prashant Loyalka, Scott Rozelle, and Binzhen Wu. 2017. "Human Capital and China's Future Growth." *Journal of Economic Perspectives* 31(1): 25–48.

Lin, Olivia. 2018. "Chinese Robotic Firms Thrive, Thanks to Subsidies." *Asia Times,* May 8.

Ministry of Industry and Information Technology (MIIT). 2013. "Guidance on the Promotion and Development of the Robot Industry" [in Chinese]. http://www.miit.gov.cn/n1146285/n1146352/n3054355/n3057585/n3057590/c3617837/content.html.

Ministry of Industry and Information Technology (MIIT), National Development and Reform Commission (NDRC), and Minister of Finance. 2016. *Robotics Industry Development Plan (2016–2020)* [in Chinese]. http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757018/c4746362/content.html.

Seamans, Robert, and Manav Raj. 2018. "AI, Labor, Productivity and the Need for Firm-Level Data." NBER Working Paper 24239.

State Council, The. 2015. "Made in China 2015." http://english.gov.cn/2016special/madeinchina2025/.

State Council, The. 2016. "Outline of the National Strategy of Innovation-Driven Development". [In Chinese]. http://www.xinhuanet.com/politics/2016-05/19/c_1118898033.htm.

Yang, Zi. 2017. "Who Will Satisfy China's Thirst for Industrial Robots?" *Diplomat,* May 19.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 89–114

Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?

Valerie A. Ramey

When the financial crisis hit ten years ago and monetary policy interest rates fell to their near-zero percent lower bound, policymakers around the world turned to fiscal stimulus packages in order to prevent their economies from freefalling into another Great Depression. But then, as declining GDP and tax revenues led to deteriorating government budget deficits and worries about rising sovereign debt, numerous countries abandoned their fiscal stimulus packages and instead adopted fiscal consolidation measures. While attempting to forecast the impacts of these various fiscal programs, policymakers and academics were surprised to discover not only a lack of consensus about the size of the effects of fiscal policy, but also a dearth of research on the topic since the 1960s. A small army of researchers across many countries turned their attention to this important but long-neglected topic.

This paper takes a snapshot of the state of knowledge about the effects of fiscal policy ten years after the global financial crisis, during which time important progress has been made on theory, empirical methods, and data. The theoretical innovations include the analysis of the effects of sticky prices, hand-to-mouth consumers, lower bounds on policy interest rates, currency unions, the type of financing, and anticipations on the reactions of macroeconomic variables to fiscal policy. Contributions in empirical methods include new ways to identify exogenous variation in policy, standardization of methods for computing fiscal multipliers (defined as the ratio

■ *Valerie A. Ramey is Professor of Economics, University of California, San Diego, La Jolla, California, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Her email address is vramey@ucsd.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.89 doi=10.1257/jep.33.2.89

of the change in output to the change in spending or taxes that caused it), and the incorporation of state dependence. On the data front, researchers now have newly constructed historical and cross-sectional datasets, and are also exploiting the rich new data created by the variety of policymakers' fiscal responses to the crisis. These advancements offer the potential to estimate the effects of government spending with more precision and with a better understanding of how the effects depend on the particular context.

In 2011, I surveyed the pre-crisis and early crisis literature in the *Journal of Economic Literature*. In that paper, which focused only on temporary, deficit-financed increases in government purchases, I concluded based on the evidence available from US data at that time that the multiplier was probably between 0.8 to 1.5, but that the data did not reject a range from 0.5 to 2. The current paper refines those estimates and broadens the inquiry to consider the effects of tax and transfer policy, as well as the effects of fiscal consolidations, in developed countries. However, attention is still limited to the short- or medium-run effects, because the methods for estimating long-run effects are quite different.

My summary of the current state of knowledge about the effects of fiscal policies can be divided into three categories: government purchases multipliers, tax rate change multipliers, and fiscal multipliers in the wake of the financial crisis.

For multipliers on general government purchases, the evidence from developed countries suggests that they are positive but less than or equal to unity, meaning that government purchases raise GDP but do not stimulate additional private activity and may actually crowd it out. The bulk of the estimates across the leading methods of estimation and samples lie in a surprisingly narrow range of 0.6 to 1. However, this range widens once one distinguishes country characteristics, such as the exchange rate regime, and the type of government spending, such as infrastructure spending. The evidence for higher spending multipliers during recessions or times of high unemployment is fragile, and the most robust results suggest multipliers of one or below during these periods. The evidence for higher government spending multipliers during periods in which monetary policy is very accommodative, such as zero lower bound periods, is somewhat stronger. Recent time series estimates for the United States and Japan suggest that multipliers could be 1.5 or higher during those times. Estimated and calibrated New Keynesian models for the United States and Europe also imply higher multipliers under certain conditions.

For tax rate change multipliers, the estimates implied by the leading methods do not agree. Narrative methods (which use historical documents to find exogenous changes) for tax rate changes typically yield multiplier estimates that are surprisingly large and surprisingly uniform across a number of countries. The bulk of the empirical estimates vary between –2 and –3. In contrast, most calibrated and estimated New Keynesian dynamic stochastic general equilibrium (DSGE) models imply smaller multipliers, typically below unity for both labor and capital tax multipliers. Time series evidence, theory, and estimated New Keynesian DSGE models all point to tax multipliers being greater in magnitude during expansions than in recessions—that is, these measures suggest that tax multipliers may be *procyclical*.

Fiscal multipliers might be different in the wake of a financial crisis. However, the evidence for larger national multipliers on the 2009 Obama stimulus package is at best weak. Quantitative New Keynesian models do not find larger fiscal multipliers. Multipliers estimated on cross-state data appear larger at first, but shrink once they are adjusted to be nationally representative. The latest studies on multipliers during the fiscal consolidations in Europe suggest that they were not higher than usual, either.

This paper begins by reviewing how theory highlights the dependence of the size of the fiscal multipliers on numerous features of the policy and the economy. The next section summarizes strengths and weaknesses of the leading empirical approaches to identifying exogenous shifts in fiscal policy. The paper then overviews the innovations of the last ten years in estimating fiscal multipliers. One interesting finding is that the wide range of multipliers reported earlier narrows significantly once methods for calculating multipliers are standardized. The following section reviews the leading estimates of spending and tax multipliers, including those based on aggregate time series, estimated theoretical models, and subnational units and households. It also discusses the complexities of drawing aggregate inferences from parameters estimated on household data. The penultimate section asks what we know about whether multipliers were higher in the wake of the financial crisis. The final section offers some brief conclusions.

What Does Theory Predict about Fiscal Multipliers?

If we simply want to know how much GDP changes if we increase government spending by $1 or reduce tax rates by 1 percentage point, why do we need theory? Theory tells us that there is not just one government spending or tax multiplier. Rather, the effect of fiscal changes on output and other variables potentially depends on: 1) the persistence of the change; 2) the type of spending or taxes that changed; 3) how the policy was financed; 4) whether it was anticipated; 5) how the policy was distributed across potentially heterogeneous agents; 6) how monetary policy reacted; 7) the state of the economy when the policy took effect; and 8) other features that characterize the economy such as level of development, exchange rate regime, and openness. Because policymakers cannot conduct randomized control trials, virtually all multiplier estimates are based on time series, narrative, or natural experiment identification using samples determined by historical happenstance. To understand whether a particular estimate of fiscal effects is suitable for use in predicting the effects of a proposed policy, one must understand how the current circumstances differ from those present in the sample used to generate that estimate.

Most researchers and policymakers had their first exposure to the theoretical effects of fiscal policy in the Keynesian cross model of undergraduate textbooks, which assumes that GDP is demand-determined. This model further assumes that the government spending multiplier is the inverse of one minus the marginal propensity to consume: thus, a marginal propensity to consume of 0.5 yields a multiplier of 2. Because taxes enter the multiplier only through their effect on disposable income in this model, the tax multipliers are smaller than the spending multipliers. Expansion of the model to consider the marginal propensity to import, tax rates, and monetary policy reduces those simple multipliers.

Neoclassical models with variable labor supply and capital stock also predict positive spending multipliers and negative (distortionary) tax multipliers, but the mechanism is completely different from the one at the heart of the traditional Keynesian model. In these models, an increase in government spending has a negative wealth effect, because the government is extracting resources from the private sector. This negative wealth effect raises GDP because it causes households to work more. Distortionary tax rate changes can have potentially large effects in these models, but contrary to the simple Keynesian model, they work through "supply side" channels (for example, Baxter and King 1993).

New Keynesian dynamic stochastic general equilibrium (DSGE) models meld the insights from the traditional Keynesian and neoclassical approaches in a rigorous way (for example, Woodford 2011). The standard representative-agent sticky-price New Keynesian model with no financial frictions tends to produce multipliers below one for government spending. Models that add sticky wages and workers who are "off their labor supply curves" generate larger multipliers. In the last decade, representative agent models have been expanded to include heterogeneous agents and financial market frictions. In these models, either "rule-of-thumb" behavior or wealth held in illiquid assets leads agents to have much higher marginal propensities to consume than predicted by the permanent income hypothesis. These features can lead to spending multipliers above one when spending is deficit financed (for example, Galí, López-Salido, and Vallés 2007; Auclert, Rognlie, and Straub 2018). Alternatively, the models have explored the effects of fiscal policy when monetary policy deviates from the standard Taylor rule (higher interest rates when inflation is high and lower interest rates when unemployment is high) because interest rates are constrained by the zero lower bound. Both of these extensions result in higher multipliers, often above unity.

Clearly, when one is trying to estimate the effects of a specific fiscal policy, one must be aware of which macroeconomic model is being used, along with other factors like persistence of a path of government spending, how it is financed, and many other characteristics such as the exchange rate regime.

A Summary of Leading Empirical Approaches

Numerous empirical approaches have been used to estimate the effects of fiscal policies. I group these approaches into three broad categories: 1) aggregate country-level time series or panel estimates; 2) estimated or calibrated New Keynesian dynamic stochastic general equilibrium (DSGE) models; and 3) subnational geographic cross-section or panel estimates.

The first two categories—time series evidence at the national level and estimated/calibrated DSGE models—share the advantage that the estimates produced are directly informative about the national-level multipliers that are the focus of most policymakers. The time series approach has the advantage of not being tied to a particular structural model. On the other hand, the New Keynesian DSGE model approach can be used to perform counterfactuals because it seeks to estimate structural parameters.

However, these two approaches share some of the same weaknesses. Identification of macroeconomic parameters is always difficult, and the estimation of the aggregate effects of fiscal policy is no exception. The time series approach requires exogenous variation in policy. The leading approaches to identifying this exogenous variation are structural vector autoregressions and natural experiment methods, combined with narrative methods that use historical documents to create new data series of exogenous changes. Too often, though, the variations that turn out to be exogenous yield instruments that are not very *relevant*—that is, they have low correlation with the fiscal variable they are trying to explain—and the variations that are relevant are not always exogenous or are anticipated in advance.

Although many papers using estimated dynamic stochastic general equilibrium models never mention the word *identification*, identification is as crucial to this approach as it is to any other approach seeking to estimate a causal relationship. The New Keynesian DSGE approach identifies the effects of fiscal policy by using strong assumptions about the theoretical model structure and the time series processes driving the unobserved shocks. But such estimated quantitative models are not immune to weak identification (for discussion, see Canova and Sala 2009).

The third approach of estimating across subnational units, such as states or provinces, is more similar to applied microeconomics approaches. These approaches typically seek identification using a natural experiment approach or Bartik-style instrumental variables (which are based on interacting the distribution of industry shares across locations with national industry growth rates).[1] These analyses at lower levels of aggregation tend to have much stronger identification, in the sense that the necessary identifying assumptions are typically more plausible and the instruments are relevant. Moreover, these approaches can be used on a variety of datasets. However, this approach does not lead directly to macroeconomic estimates. Why? Any cross-sectional estimating equation includes a constant term, which means that the macroeconomic effects have been netted out and the parameters estimated are only *relative* effects. Such parameters answer the question: If State A is awarded $1 more in defense prime contracts than the average state, by how much does its employment change *relative* to the average state? In order to infer the implied national-level effects from such microeconomic estimates, researchers must then return to macroeconomic New Keynesian DSGE models, which, as discussed above, incorporate their own additional identifying assumptions. There is no "applied micro free lunch" for macroeconomists. Identification of macroeconomic effects must always depend on macroeconomic identification assumptions.

[1] For a description and critical analysis of Bartik instruments, see Goldsmith-Pinkham, Sorkin, and Swift (2018).

To summarize, there are several approaches to estimating the effects of fiscal policy. Each has its strengths and weaknesses. Moreover, some of the estimates are more appropriate for forecasting the effects of specific policies under certain conditions than others. For these reasons, it is useful to consider estimates across a range of different approaches.

Research Innovations and Lessons Learned during the Last Ten Years

Before the financial crisis, only a few isolated researchers studied the macroeconomic effects of fiscal policy and only a few conferences brought these researchers together. As a result, different researchers chose different methods and there was no agreement on a set of best practices. The situation has changed dramatically since the financial crisis, with many conferences devoted to the study of fiscal policies and much more interaction among researchers studying fiscal policy. As a result, the diffusion of knowledge among researchers has been much faster, and the literature has progressed at a very fast pace. In this section, I will highlight some of the new innovations and the lessons learned from this literature.

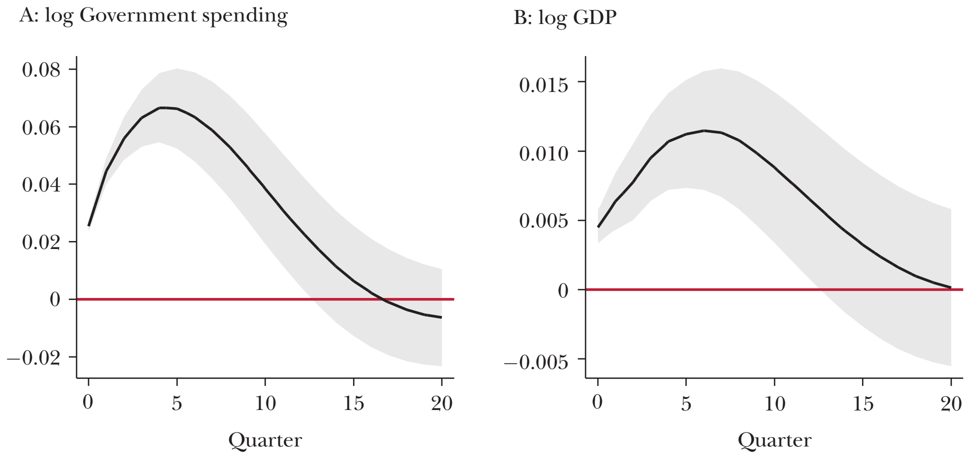
Calculating Multipliers in a Dynamic Environment

One often sees references to the “wide range” of multiplier estimates. The literature has come to realize that differences in reported multiplier estimates are often due not so much to differences in identification methods or samples, but to the methods used to construct multipliers from the raw estimates. In fact, what some researchers call “multipliers” have little to do with the multipliers of interest to policymakers. This section begins with some insights gained over the last decade regarding the computation of multipliers. I begin with spending multipliers and then address a further complication involved with tax multipliers.

Fiscal policy has dynamic effects on output and government budgets. A typical fiscal plan will set into motion a path of spending or taxes over time, and then GDP will respond dynamically to that path. The multiplier must take into account both the multi-year effects of the fiscal plan on the government budget, in order to count the costs fully, as well as the multi-year effects on GDP, in order to count the benefits fully.

Computation of fiscal multipliers was not a focus of research in the decades before the financial crisis. Indeed, in Ramey and Shapiro (1998), when discussing the effects of government spending two decades ago, we did not even mention the word “multiplier.” When describing the patterns of the responses of GDP to spending and tax shocks, Blanchard and Perotti (2002) used the word “multiplier,” but the quantities they calculated were not true dynamic multipliers; instead, Blanchard and Perotti calculated multipliers as the ratio of the output response at a particular horizon, or at its *peak*, to the *impact* effect of the shock on government spending. Many subsequent papers adopted their method, despite the fact that it did not take into account the multi-year path of spending or taxes. Mountford and Uhlig (2009) moved the literature forward by introducing the policy-relevant multipliers, calculated as the present

Figure 1
Estimated Impulse Response Functions for a Shock to Government Purchases



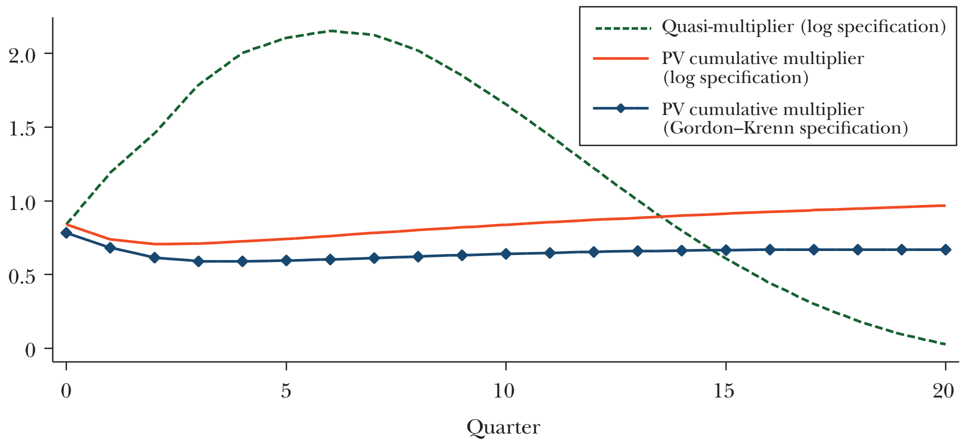
Source: Author.
Note: Estimated impulse responses based on structural vector autoregression (SVAR) estimates using quarterly data from 1939Q1– 2015Q4. The shaded area shows the 95-percent confidence bands. See the text and online appendix available with this paper at the journal website for more detail.

discounted value of the output response over time divided by the present discounted value of the government spending response over time to the shock. In most applications, different interest rates used for this present discounted value—including the use of a zero discount rate—give nearly identical multipliers because the timing of the government spending and output responses is very similar. These multipliers are often known as *present value* or *cumulative multipliers*.

How much do multiplier estimates differ across these various methods of calculating multipliers? It depends importantly on how much government spending rises after the initial impact. Here is one illustration of a situation in which it makes a big difference. I estimate a structural vector autoregression (SVAR) model of the Blanchard and Perotti (2002) type over the period 1939Q1–2015Q4 using the Ramey and Zubairy (2018) dataset. The model contains five endogenous macroeconomic variables: government spending, GDP, and federal tax receipts (with all three deflated by the GDP deflator, divided by population, and in logs), along with the three-month Treasury bill interest rate and inflation (measured as the log change in the GDP deflator). Four lags are included in order to model the dynamics. The exogenous shock to government spending is identified using Blanchard and Perotti's (2002) method, which assumes that any part of government spending not forecasted by lags of any of the variables included in the model is an exogenous shock to government spending.

Figure 1 shows the estimated impulse responses of the log of the government spending variable and the log of the GDP variable (notice that the vertical scales are not the same). The shaded area shows the 95-percent confidence bands. As the

Figure 2
Alternative Definitions of Multipliers: Multipliers by Horizon



Source: Author.
Note: The dotted and solid lines show multipliers calculated based on the log impulse responses shown in Figure 1. The dashed line shows the multiplier given by Blanchard-Perotti's (2002) method, which I call a *quasi-multiplier*. The solid line shows the the Mountford and Uhlig (2009) *present value (PV) cumulative multiplier*. The line with diamonds shows the PV cumulative multiplier using the impulse responses estimated using the Gordon–Krenn specification. See text and online appendix available with this paper at the journal website for more details.

graph illustrates, a positive shock to government spending leads both government spending and GDP to jump up on impact, but then to continue to rise, peaking after about a year. Because the variables are in log form, the impulse responses show elasticities, not the dollar changes required by multipliers, so multipliers cannot be read directly from the graphs. The standard practice until recently had been to use an *ad hoc* "conversion factor." That is, researchers who specified models using logarithms converted the elasticity estimates to multipliers by multiplying the elasticity estimates by the average of the ratio of GDP to total government spending, over the sample. In this illustration, the conversion factor, average Y/G over the sample, is 4.78. I will critique the use of these conversion factors shortly.

Figure 2 shows the multipliers calculated three different ways. The highest multiplier is given by Blanchard-Perotti's (2002) method for calculating a multiplier, which I will call a *quasi-multiplier*.[2] It is calculated as the ratio of the impulse response of output at horizon h to the initial jump in government spending at horizon 0 (multiplied by the average). Their method, shown by the dashed line, essentially traces out a renormalized version of the impulse response of output. In

[2] Note that the Blanchard–Perotti *identification* method is distinct from the Blanchard–Perotti method for calculating multipliers; their method for calculating multipliers could be applied to estimates using any identification method.

this case, it yields multipliers that peak at 2.2 at quarter 6. The Mountford and Uhlig (2009) *present value cumulative multiplier*, shown by the solid line, uses the ratio of the present value of the integral of impulse response of output to the present value of the integral of the impulse response of government spending up to each horizon *h* (again multiplied by the average *Y/G* factor). This multiplier varies between 0.7 and 1, depending on the horizon. The discounting for this multiplier uses the average three-month Treasury bill rate over the sample, 3.6 percent on an annual basis, but because of the timing of the shift, the simple cumulative version is almost identical.

Now let us return to the issues raised by the practice of converting elasticities with the *ad hoc* conversion factor, the average of *Y/G* over the sample. In Owyang, Ramey, and Zubairy (2013), we discovered biases that could arise from this practice. In our historical sample, *Y/G* varied significantly, from 2 to 24, with a mean of 8. Sims and Wolff (2018a, b) also discovered that this practice tends to bias multipliers differentially, making them seem much higher during recessions. The intuition is straightforward: because GDP is cyclical but government spending is not, the movement of *Y/G* is procyclical. However, the practice of using a sample average to convert elasticities to multipliers makes the multipliers appear more countercyclical than they really are. In Owyang, Ramey, and Zubairy, we avoided this problem by using the transformations employed by Hall (2010) and Barro and Redlick (2011): both the change in government spending and the change in GDP are divided by *lagged* GDP. Another transformation that overcomes the problem is Gordon and Krenn's (2010) approach, which divides both government spending and GDP by a measure of potential GDP.

To illustrate the effect of moving from a specification in logarithms that requires the *ad hoc* conversion factor to one that does not, I re-estimate the structural vector autogression (SVAR) model, replacing the logarithms of government spending, GDP, and taxes with the ratios of each of those variables to the Ramey and Zubairy (2018) polynomial trend estimate of potential GDP. The general shape of the estimated impulse responses (not shown) is very similar to those from the log specification, which were shown in Figure 1. The solid line with diamonds in Figure 2 shows the cumulative multiplier estimates based on the impulse responses from this alternative (Gordon–Krenn) specification. These multipliers, which do not rely on a conversion factor, are lower and range from 0.8 on impact down to 0.6.[3]

Thus, deceptively small changes in the method of calculation can make a very big difference in the resulting multipliers. For this application, using Blanchard and Perotti's (2002) quasi-multiplier for government spending on estimated elasticities requiring an *ad hoc* conversion factor produces a multiplier as high as 2.2. That multiplier falls below 0.8 when the fully dynamic Mountford and Uhlig (2009) cumulative multiplier is used on estimates based on data using the Gordon and

[3] This bias also affects the multipliers I reported in Ramey (2011a). The cumulative multipliers based on the elasticity estimates and conversion factor were 1.2. However, in Ramey (2013), I found evidence that private spending fell, which is inconsistent with a multiplier above 1.

Krenn (2010) transformation. Clearly, such differences could have important consequences for the decisions of policymakers.

In addition, even the cumulative multipliers do not fully reflect the consequences for the government budget. If an increase in government spending raises GDP, then we would expect a rise in tax revenues. Thus, even without an exogenous increase in tax rates, we would expect the government budget deficit to rise less than the total amount of government spending. This insight raises a complication when applying these same principles to the computation of tax multipliers. While there is strong feedback from GDP to tax revenue, there is little feedback from GDP to government spending. As a result, the negative effect of a tax cut on tax revenue is tempered by the feedback from the expansionary effect on output. Indeed, Mertens and Ravn (2013) were not able to compute a multiplier for corporate tax cuts because their large positive impact on GDP resulted in no net effect on tax revenues. Because of the presence of these "top of the Laffer curve" effects in some applications, most papers report multipliers using the tax changes measured as the legislative forecasts of the expected cumulative effect on tax revenues, *not* accounting for dynamic feedback from any potential induced GDP changes.

The Importance of Fiscal Foresight

An important innovation in the fiscal literature in the last decade is the recognition that many changes in government spending and taxes are announced in advance. In Ramey (2011a), I showed the importance of anticipations for estimating the effects of government spending shocks, particularly those involving military spending. For example, the responses of key variables such as consumption could change signs if researchers ignored the fact that many changes in government spending are anticipated by at least several quarters. A number of papers also show that "shocks" identified in standard ways are predicted by professional forecasts of government spending. On the tax front, House and Shapiro (2006) and Mertens and Ravn (2012) demonstrated the importance of distinguishing between changes in taxes implemented soon after legislation and changes in taxes implemented with a lag after legislation or phased in slowly. Both papers showed that while unanticipated tax cuts have expansionary effects on output, phased-in tax cuts depress output during the phase-in period because firms and consumers delay their activity until tax rates are lower. Leeper, Walker, and Yang (2013) derived the econometric biases that arise when there is this type of fiscal foresight. As a result of this work, most of the literature tries to address anticipation whenever feasible, either by constructing measures of news (from narratives or bond spreads) or by including professional forecasts of government spending to mitigate the problem.

Improvements in Fiscal Shock Identification

Any analysis that seeks to measure a causal effect must confront identification issues. An example of the problem that arises here is that if governments increase spending in response to a recession, then the simple correlation between government spending and GDP will confound the positive causal effect of government spending on GDP with the negative causal effect of GDP on government spending. In the past, the standard macro approach used to tease out the exogenous rise in government spending was a structural vector autoregression (SVAR). In most applications, this approach is based on the assumption that the exogenous part of government spending was the part of government spending not forecasted by lagged values of spending, GDP, and taxes. Alternatively, to identify exogenous movements in taxes, Blanchard and Perotti (2002) used external estimates of the elasticity of tax revenue to income, which allowed identification of the component of taxes that was not induced by movements in GDP. Several papers have highlighted potential problems with these widely used methods. First, as discussed above, failing to account for fiscal foresight could lead to biased estimates. Second, the tax multiplier estimates were very sensitive to the value of the external tax elasticity estimate used (for example, Mertens and Ravn 2014; Caldara and Kamps 2017). These concerns led to the development of other identification methods using natural experiments and narrative methods. As a result, the standard SVAR identification approach is no longer the first resort in the literature on fiscal multipliers.

In fact, long before structural vector autoregression methods were used, Hall (1980) and Barro (1981) used natural experiment methods to assess the effects of exogenous increases in government spending. Arguing that changes in US defense spending are typically driven by wars rather than the current state of the economy, they used war-induced government spending to estimate causal effects of government spending in US historical data. Ramey and Shapiro (1998) and numerous other follow-up papers built on treating wars as a natural experiment. This method works well for US data, but it does not export well to other countries. Most countries either do not have the substantial fluctuations in defense spending experienced by the United States or they have large variations that are accompanied by war-related destruction of the capital stock, which leads to confounding effects.

Other examples of recent fiscal research that use natural experiment methods abound. For example, Acconcia, Corsetti, and Simonelli (2014) used the central government response to Mafia infiltration as an exogenous change in government spending in Italian provinces. Many of the analyses of the Obama stimulus allocation of funds across states used natural experiment methods. Two analyses of marginal propensities to spend out of the temporary rebates of 2001 and 2008 exploited the randomized timing of the mailing of checks to households (Johnson, Parker, and Souleles 2006; Parker, Souleles, Johnson, and McClelland 2013). The application of these methods has shed significant light on the effects of fiscal policy, particularly at the local and household level.

Romer and Romer (2010) pioneered the use of narrative methods to identify tax changes that are exogenous to the state of the economy. For the post-World War II US economy, they read legislative records to identify whether tax changes were due either to inherited deficits or to beliefs about their ability to promote long-term growth. Their method is easily exported to other countries, and it has now become the standard method for assessing the effects of tax changes across a wide range of countries (for example, Guajardo, Leigh, and Pescatori 2014). Mertens and Ravn (2012) improved their measure by splitting their series into anticipated and unanticipated tax changes, so that the effects of fiscal foresight could be addressed. Alesina, Favero, and Giavazzi (2019) have added to the narrative analysis of fiscal consolidations by creating narrative series of fiscal plans. As they emphasize, most fiscal consolidations involve multi-year plans and those effects should be studied as a whole rather than as independent year-by-year isolated changes.

An additional innovation in the identification of fiscal shocks has been the recognition of the importance of instrument "relevance"—that is, whether the proposed instrument is actually correlated with the variable it is supposed to instrument. While early alarms about weak instruments were raised for macro studies by Nelson and Startz (1990) and for microeconomic studies by Bound, Jaeger, and Baker (1995), most macroeconomists began to pay attention to the issue only in the last five to ten years. The structural vector autoregression methodology hid the fact that the estimation of multipliers was actually an instrumental variables estimation. In Ramey (2016) and Ramey and Zubairy (2018), we showed that cumulative multipliers could be estimated in a one-step instrumental variables method based on local projections: cumulative GDP up to horizon h is regressed on cumulative government spending up to horizon h, using an SVAR shock or a narrative variable as an instrument. However, that recognition highlighted a widespread problem: many of the exogenous measures of fiscal policy are not very relevant instruments, at all or in some subsamples. For example, the military news variable I first introduced in Ramey (2011a) is a weak instrument for the post-1954 period, as are the alternative measures of defense news of Fisher and Peters (2010) and Ben Zeev and Pappa (2017). In contrast, the Blanchard and Perotti (2002) shock is a strong instrument by its nature, particularly at short horizons, since it is simply the one-step ahead forecast error of government spending.

In sum, research on the effects of fiscal policy has made significant strides in methodology. The literature now exploits many new datasets. It has imported some innovations from the applied microeconomics literature, and has extended them in important ways that account for anticipations and dynamics. Moreover, those estimates are now converted to multipliers defined in a way that is relevant for policymakers.

A Summary of Estimates of Spending and Tax Multipliers

This section summarizes the actual estimates of fiscal multipliers obtained from the leading methods. I begin with estimates based on aggregate data. I first review the estimated multipliers on government purchases, initially averages and then by state-dependence. Next, I move on to the effects of tax changes and transfer payments. I then discuss estimates of the effects of the American Recovery and Reinvestment Act of 2009 and the fiscal consolidations in Europe.

Government Spending Multipliers Based on Aggregate Data

Table 1 shows a sampling of estimates of government spending multipliers, grouped by method. Virtually all estimates shown are based on present value or undiscounted cumulative multipliers; in some cases, I updated the original estimates to apply best practices. As shown in Figure 2, the cumulative multipliers usually do not vary greatly across horizons up to five years, so there is little difference between average or peak multipliers. The estimates in Panel A show that the estimated multipliers are not very different across the various methods for identifying government spending shocks in time series. Panel B displays estimates based on New Keynesian DSGE models. The multiplier estimates from these models are similar to those from Panel A. On balance, the table shows that for a variety of samples, identification methods, and countries, most of the estimates are around one or below. A few estimates are noticeably above one, such as the Ben Zeev and Pappa (2017) estimate, but they tend to be less precise and are not statistically different from one. Not shown in the table are numerous multiplier estimates based on key features of a country. For example, Iltzetzki, Mendoza, and Végh (2010) estimate how multipliers change across various important features, such as whether an economy has fixed or flexible exchange rates. They find multipliers that vary between 0.1 on impact to 1.4 in the long run (with a 90-percent confidence interval from around 0.75 to 2.1) for fixed exchange rates and from 0.1 to –0.7 for flexible exchange rates. Thus, the range of estimated multipliers may become much wider when one begins to distinguish by key country characteristics.

The results shown in Table 1 are for total government spending or government consumption. Earlier work by Aschauer (1989), Pereira and Flores de Frutos (1999), and others found high returns to public investment. There is surprisingly little recent aggregate evidence on multipliers for public investment. As one example, Iltzetzki, Mendoza, and Végh (2010) found multipliers for public investment that ranged between 0.4 in the short-run to 1.6 in the long-run in their panel of countries.

Even if government spending multipliers are probably one or below on average, might they be higher during bad economic times? In estimating fiscal multipliers, some key states studied by recent papers are recessions or periods of excess slack (typically measured by unemployment rates), constraints on the monetary policy accommodation (such as the zero lower bound), and the ratio of public debt to GDP.

First consider multipliers during recessions or periods of slack. Auerbach and Gorodnichenko (2012), who conducted the pioneering study on this question, used a nonlinear time series model in which the parameters changed across expansions and recessions. They reported a multiplier of 2.2 in recessions and –0.3 in expansions (based on some simplifying assumptions about the state of the economy not changing after the shock). Various other studies have found high multipliers during recessions (for example, Auerbach and Gorodnichenko 2013; Fazzari, Morley, and Panovski 2015; Caggiano, Castelnuovo, Colombo, and Nodari 2015). However, subsequent research has found many of the state-dependent results to be very

Table 1
Estimates of Government Spending Multipliers Using Aggregate Data, No State Dependence
(almost all are cumulative multipliers, typically over horizons between 0 to 20 quarters)

Method/Sample	*Multipliers*	*Comments*
A: Time series analysis		
Updated implementation of Blanchard and Perotti (2002) identified SVAR		The tax response is positive for the 1939Q1–2015Q4 period, but is essentially 0 for the later periods.
1939Q1–2015Q4	0.6 to 0.8	
1947Q1–2015Q4	0.6 to 0.7	
Military news shocks, local projections Ramey and Zubairy (2018) military news		Tax response is positive for 1939Q1–2015Q4 period.
1889Q1–2015Q4	0.6 to 0.8	SE from 0.04 to 0.06
1939Q1–2015Q4	0.7 to 0.8	SE from 0.05 to 0.1
1947Q1–2015Q4	0.5 to 0.7	SE from 0.15 to 0.2
Ben Zeev and Pappa (2017) news, 1947Q1–2007Q4[a]	1.1 to 2	SE from 0.6 to 1
Hall (2019), Barro and Redlick (2011)—based on regressions using annual defense spending.	0.6 to 0.7	The Barro–Redlick analysis nets out effects of changes in tax rates.
Mountford and Uhlig (2009), SVAR with sign restrictions	0.65	Deficit-financed increase in government spending.
Iltzetzki, Mendoza, and Végh (2013), Blanchard–Perotti identification in SVAR, quarterly data, 1960–2007, 44 countries high-income countries	0.3 to 0.7	
Corsetti, Meier, and Müller (2012)	0.7	Based on unconditional model results reported in their Figure 1.
Leigh et al. (2010), Guajardo, Leigh, and Pescatori (2014), 17 OECD countries, 1978–2009, narrative method identification of spending-based fiscal consolidations	0.3	
Alesina, Favero, and Giavazzi (forthcoming). Narrative analysis of austerity plans, 16 OECD economies from 1978–2014.	0.3	
B: Estimated New Keynesian DSGE models		
Cogan et al. (2010), estimated Smets–Wouters DSGE model on US data	0.6 to 0.7	Based on my visual inspection of figures 2, 3, and 4.
Coenen et al. (2012), large-scale macro models used by central banks and IMF, United States and Europe	0.7 to 1	Based on the two-year cumulative multipliers shown in the upper left graph in figure 6.
Zubairy (2014), estimated medium-scale DSGE model on US data	0.7 to 1.05	Deficit financed, model features deep habits.
Leeper, Traum, and Walker (2017), estimated DSGE model on US data	0.7 to 1.36	Active monetary policy, table 7
Sims and Wolff (2018a)	1.07	The multiplier above 1 is due to estimated complementarity of government spending with private consumption.

Note: SVAR is structural vector autoregression. DSGE stands for "dynamic stochastic general equliibrium."
[a] These estimates are based on the analysis in Ramey (2016) using Ben Zeev and Pappa's estimated news series.

fragile to small changes in specification or to improvements in the methods for computing the multipliers from the basic estimates (Alloza 2017; Owyang, Ramey, and Zubairy 2013; Ramey and Zubairy 2018, and associated online appendix). The more robust methods generally fail to produce multipliers above one during recessions or times of slack.

Perhaps these empirical results should not be surprising, given some other results of theory and quantitative models. The only theoretical models that predict countercyclical markups are ones that include significant frictions. For example, Michaillat (2014) presents a stylized model with labor market frictions and finds that the aggregate employment effect of government hiring is countercyclical. However, the multipliers are always below one. Canzoneri, Collard, Dellas, and Diba (2016) present a model with financial frictions that does generate sizeable, though fleeting, multipliers during recessions. They find significantly higher-impact multipliers during recessions, near 2, but the cumulative multipliers fall below 1 after only a few quarters. Standard new Keynesian models do not predict higher multipliers during recessions. Indeed, Sims and Wolff (2018a) employ a medium-scale New Keynesian DSGE model with high-order terms in the approximations and find that this otherwise standard model implies mildly *procyclical* multipliers.

The situation is different with respect to periods when interest rates are near the zero lower bound or when monetary policy accommodates government spending increases (such as during World War II in the United States). Numerous New Keynesian DSGE models show that multipliers can be higher than one when monetary policy is constrained by the zero lower bound on interest rates. At the zero lower bound, an increase in government spending provides extra stimulus by increasing expected inflation, which lowers the real interest rate (Farhi and Werning 2016). Calibrated models such as the ones analyzed by Christiano, Eichenbaum, and Rebelo (2011) and Coenen et al. (2012) can produce multipliers that range between 2 and 3 when the period of monetary accommodation is sufficiently long. Some recent empirical work has found some evidence of higher multipliers, ranging from 1.5 to 2.5 at the zero lower bound for Japan (Miyamoto, Nguyen, and Sergeyev 2018) and around 1.5 for historical samples in the United States (Ramey and Zubairy 2018).

Finally, there is evidence that government spending multipliers may be negatively related to the public debt-to-GDP ratio. For example, Iltzetzki, Mendoza, and Végh (2013) find that countries with a government debt-to-GDP ratio above 60 percent have an impact multiplier of 0 and a long-run multiplier of –3 (estimated less precisely but still statistically below 0).

In summary, most estimates of government spending multipliers for general categories of government spending for averages over samples are in the range of 0.6 to 0.8, or perhaps up to 1. The evidence for multipliers above one during recessions or times of slack is typically not robust. However, some initial explorations suggest that government spending multipliers could be higher at times when monetary policy accommodates fiscal policy, such as during periods at the zero lower bound of interest rates or wartime.

Tax and Transfer Multipliers Based on Aggregate Data

I now turn to the leading estimates of tax and transfer multipliers at the aggregate level. Tax multipliers are generally negative since an increase in taxes lowers GDP. Table 2 shows the estimates from time series and New Keynesian DSGE estimates for tax rate changes. In contrast to government spending multipliers, which vary only a small amount across horizons, many estimates of tax multipliers start out low on impact but then build. Thus, I report the cumulative multipliers for the horizon where they peak. I should also note that most of the multipliers are calculated without allowing feedback from induced output changes to revenue but several (noted in the table) allow for dynamic feedback.

Most of the time series estimates based on narrative methods of identification are quite high (in absolute value), generally between –2 and –3. These narrative-based estimates are striking not only for their magnitudes, but also for their uniformity across countries and even across various methods of estimation. These estimates are much higher (in absolute value) than the tax multipliers reported by Blanchard and Perotti (2002). As discussed above, those estimates were based both on their assumed elasticity of tax revenue to output and on their unusual way of computing multipliers. Barro and Redlick (2011) estimate multipliers around –1.1. It may be that their use of various approximations and constraints on dynamics account for their smaller estimate. On the other hand, Mountford and Uhlig's (2009) estimates using sign restrictions are –5.

In contrast, the New Keynesian DSGE model estimates are much lower. Panel B of Table 2 shows that most New Keynesian model estimates yield multipliers that are below 1 in absolute value. Thus, there is a conflict between the narrative-based time series estimates and the New Keynesian estimates.

There is a small literature on whether tax multipliers differ by the state of the economy. So far, this literature offers fairly uniform answers. Eskandari (2015) and Demirel (2016) find, using the Romer and Romer (2010) narrative tax shocks, that tax multipliers are *greater* during times of low unemployment than times of high unemployment. Alesina, Azzalini, Favero, Giavazzi, and Miano (2018) also find higher multipliers in expansions using their narrative of fiscal plans across OECD countries. These results are consistent with the one New Keynesian analysis of this issue using the dynamic stochastic general equilibrium approach. Sims and Wolff (2018b) obtain estimates of tax multipliers that are procyclical: for example, their capital tax multiplier is 1 in recessions and almost 2 in expansions.

There has been very little work on the aggregate effects of transfers. Romer and Romer (2016) used changes in Social Security benefit increases to study the effects on macroeconomic variables. They found that permanent increases in benefits led to a roughly equal rise in consumption in the short-run, but the effect dissipated quickly. Temporary increases in benefits had no significant effect on aggregate consumption. Coenen et al. (2012) studied general transfers and directed transfers across the various New Keynesian DSGE models used at policy institutions. They found that general transfers had multipliers between 0.2 and 0.6, with the higher ones occurring with monetary accommodation. In contrast, targeted transfers (to

Table 2
Estimates of Tax Change Multipliers Using Aggregate Data, No State Dependence
(† denotes multipliers computed using the cumulative actual response of tax revenues or deficits in the denominator)

Method/Sample	*Largest cumulative multiplier within first 5 years*	*Comments*
A: Time Series Methods		
Mountford and Uhlig (2009), SVAR with sign restrictions, US data	–5†	
Romer and Romer (2010), narrative series of tax changes unrelated to current economy, US data, 1950 to 2007, dynamic single equation model or VAR	–2.5 to –3	The output effects take time to build.
Barro and Redlick (2011), historical annual US data, tax rate shocks.	–1.1	
Mertens and Ravn (2013, 2014), refinement of Romer and Romer series used in a proxy SVAR	–2.5 to –3†	The peak output effects occur in the first 18 months.
Cloyne (2013), narrative, UK	–2.5	
Hayo and Uhl (2013), narrative, Germany	–2.4	
Guajardo, Leigh, and Pescatori (2014), 17 OECD countries, 1978–2009, narrative taxed-based consolidations	–3	
Riera-Crichton, Vegh, and Vuletin (2016), narrative analysis of fiscal consolidations in 15 industrialized countries from 1980 to 2009, with focus on VAT rate changes	–3.5	
Alesina, Azzalini, Favero, Giavazzi, and Miano (2018), narrative analysis of austerity plans, 16 OECD economies from 1978 to 2014, taxed-based consolidations		
Based on static primary surplus	–1 to –1.6	
Based on actual response of primary surplus	–2.3 to –3.7†	
B: Estimated New Keynesian DSGE models		
Coenen et al. (2012), large-scale macro models used by central banks and IMF, United States, and Europe. Two-year cuts in tax, no monetary accommodation		
Consumption tax	–0.2 to –0.4	
Labor tax	–0.2 to –0.4	
Corporate income tax	0 to –0.15	
Zubairy (2014)		
Labor tax	–0.7 to –1	
Capital tax	–0.2	
Sims-Wolff (2018b), medium scale New Keynesian DSGE model that allows for higher-order terms.		Steady-state multipliers
Consumption tax	–0.6	
Labor tax	–1	
Capital tax	–1.5	

Note: SVAR is structural vector autoregression. VAR is vector autoregression. VAT is value added tax. DSGE stands for dynamic stochastic general equilibrium.

households that were financially constrained) yielded multipliers as high as 2 in some models when there was monetary accommodation.

In sum, most time series estimates of tax rate change multipliers indicate that they are very large, at least –2 to –3. This contrasts with the results from estimated New Keynesian dynamic stochastic general equilibrium, where the multipliers (in absolute value) are typically below 1 and never higher than 1.5. There is not much aggregate time series evidence for sizeable multipliers for temporary transfers, though calibrated New Keynesian models suggest they can be high if they are targeted and if monetary policy is accommodative.

Multiplier Estimates Based on Subnational Data

One of the important innovations in the fiscal multiplier literature, as mentioned earlier in this paper, has been the application of applied microeconomics-type identification methods to the estimation of parameters of use for macroeconomics. These include studies of panels or cross-sections of US states or provinces in other countries, as well as household-level estimates of marginal propensities to spend out of temporary transfers.

Chodorow-Reich (forthcoming) summarizes the panel and cross-section multipliers from individual studies, so I refer the reader to his tables. Many of the subnational multipliers for government purchases, temporary tax rebates, and transfers lie between 1.5 to 2. Thus, they tend to be higher than the aggregate-level estimates of multipliers.

As noted earlier, subnational multipliers are not the same as aggregate multipliers. The relationship between subnational multipliers and aggregate multipliers depends on many features, including how the spending is financed, whether there are spillovers across regions, whether there is a currency union, and whether the aggregate economy is at the zero lower bound. For discussion of some of the theoretical considerations when drawing implications from subnational multiplier estimates to aggregate estimates, see Nakamura and Steinsson (2014), Farhi and Werning (2016), and Chodorow-Reich (forthcoming). In some instances, the subnational multipliers are expected to be higher than the aggregate multipliers, whereas in other instances they are expected to be lower. There is no general rule. Dupor and Guerrero (2017) conduct an empirical investigation in which they directly compare estimates based on a state-level panel to those obtained when the state data are aggregated to the national level. They obtain similar multiplier estimates across the two datasets, though quite low, between 0 and 0.5.

Multipliers in the Wake of the Financial Crisis

A number of researchers and commentators have argued that the effects of the stimulus from the American Recovery and Reinvestment Act of 2009 and the subsequent fiscal consolidations in European countries were much larger than indicated by multipliers during average times. A common theme is that the high unemployment rates and lower bound on interest rates combined to raise the multipliers. But as shown in the previous sections, there is no robust evidence of higher multipliers during recessions or times of slack, for either spending or taxes. In fact, all studies of state dependence for tax multipliers find *higher* multipliers during expansions. However, there is evidence from historical periods in the United States and from Japan, as well as from New Keynesian models, that multipliers can be higher than one during periods of monetary accommodation such as the zero lower bound on interest rates. Thus, it is possible that multipliers could have been higher after the financial crisis.

Consider first the fiscal consolidations in Europe, aimed at reducing government deficits and debt. Blanchard and Leigh (2013, 2014) presented evidence that countries that implemented bigger fiscal consolidations grew more slowly than forecasted by the IMF and other organizations. They concluded that the models used by forecasters assumed values of multipliers that were too small. Górnicka, Kamps, Koster, and Leiner-Killinger (2018) gathered data on the forecasters' assumed values of multipliers and found that they were very low, around 0.25. They then calculated that the "true" multipliers were higher, though they never exceeded one.

The conclusions of Górnicka et al. (2018) are consistent with some other analyses of the size of multipliers in the European fiscal consolidations. For example, Alesina, Favero, and Giavazzi (2019) use their narrative dataset of fiscal consolidation plans across OECD countries to study whether fiscal multipliers were greater in the immediate post-financial crisis years. They find no evidence that multipliers were greater. At this point, the evidence does not suggest that multipliers were larger than normal for the fiscal consolidations in Europe.

The American Recovery and Reinvestment Act (ARRA) of 2009 was the leading stimulus program in the US economy. This program was a mix of spending and transfers to states and individuals. As Table 3 shows, none of the New Keynesian DSGE models find multipliers above 1 for this program, with the exception of one experiment by Coenen et al. (2012) that included two years of monetary accommodation. While interest rates were indeed at the zero lower bound during those years, Swanson and Williams (2014) present evidence that yields on one- and two-year Treasury bills were unconstrained from 2008 to 2010, "suggesting that monetary policy and fiscal policy were about as effective as usual during this period."

In contrast, the cross-state estimates of the effects of the American Recovery and Reinvestment Act are typically much higher. Chodorow-Reich (forthcoming) presents an extremely valuable standardization and synthesis of the leading estimates of the effects of the stimulus act on job creation across US states. This literature emphasizes employment effects, mainly because the employment data have less measurement error than gross state product. These estimates are based on strong applied microeconomic methods. His cross-state natural experiment estimates indicate multipliers from 1.7 to 2 for gross state product and $50,000 per job-year created. Building on Farhi and Werning's (2016) theoretical analysis, Chodorow-Reich (forthcoming) argues that these subnational multipliers are *lower bounds* on the national multipliers during a liquidity trap. Thus, he argues that the multiplier from the American Recovery and Reinvestment Act was at least 2.

Table 3
Multipliers for the American Recovery and Reinvestment Act (ARRA)

Method/Sample	*Peak cumulative multipliers within first 5 years*	*Comments*
Cogan et al. (2010)	0.6 to 0.7	
Coenen et al. (2012), large-scale macro models used by central banks and IMF, US, and Europe		From figure 7. These are the peak instantaneous multipliers.
No monetary accommodation	0.3 to 0.5	
1 year monetary accommodation	0.4 to 0.6	
2 years monetary accommodation	0.5 to 1.8	
Drautzburg and Uhlig (2015), medium-scale New Keynesian DSGE model, with ZLB, credit constraints	0.5	Multipliers become negative in the long run because of the necessary increase in taxation.
Chodorow-Reich (forthcoming), based on cross-state estimates and theoretical arguments about the relationship between subnational and national multipliers at the ZLB.		
Gross State Product multiplier	1.7 to 2	
Cost per job year	2 job-years per $100K	

Note: ZLB is zero lower bound.

But there is reason to suspect that the state-level estimates of the effects of the American Recovery and Reinvestment Act presented by Chodorow-Reich (forthcoming) are probably overestimates for the national-level multipliers. His cross-state estimates answer one question: "How much extra employment was induced in the average state by each *$1 of ARRA spending* by the federal government?" But the question relevant for the aggregate effects is a different one: "How much extra aggregate employment was generated by each *$1 of government spending induced by ARRA spending* by the federal government?" Chodorow-Reich uses per capita values of spending and employment in each state, and his cross-state estimates give equal weight to North Dakota and California, which is fine for answering the first cross-state question. But if there is heterogeneity in the treatment effects, the estimates will not give estimates that are nationally representative.[4] The data need to be weighted by population or in some other way to obtain nationally representative results. A second issue is that Chodorow-Reich's measure of spending is federal ARRA spending, which again is appropriate for measuring the first cross-state question. However, ARRA spending stimulated state and local spending more than dollar for dollar (Leduc and Wilson

[4] Most of the literature using cross-sectional estimates has used per capita estimates and has not weighted the estimates. However, Dupor and Mehkari (2016) started weighting the estimates and discovered that weighted estimates of the American Recovery and Reinvestment Act of 2009 are much lower than unweighted estimates.

Table 4
Conversion of Chodorow-Reich Estimates to Nationally Representative Estimates

	Cumulative employment multiplier estimates (number of job-years created per $100K of ARRA spending)
Chodorow-Reich (forthcoming) headline estimates (his table 1, column 4)	2.01 (0.59)
Weighted estimates (using December 2008 population of state)	1.15 (0.72)
Weighted estimates based on total spending, including induced spending by states	0.89 (0.45)

Note: ARRA is American Recovery and Reinvestment Act. Estimates presented in the last two rows are the author's estimates, based on Chodorow-Reich's programs and data in his forthcoming paper. See the text and online appendix for more detail and programs.

2017). Thus, multipliers that use only the ARRA transfers to the states will overestimate the multiplier per dollar spent across all levels of government.

Table 4 shows the effects of adjusting the employment response estimates to make them more suitable for answering the question about aggregate effects of federal government spending. The first row shows Chodorow-Reich's (forthcoming) preferred estimates, which use all three of the leading instruments for estimating cross-state effects of the American Recovery and Reinvestment Act: Medicaid formulas, Department of Transportation formulas, and a combination of multiple agency formulas. The estimates are for job-years created for each $100,000 of ARRA spending. Thus, the estimate of 2.01 implies that each $100,000 of ARRA spending creates two job-years of employment. The second row of Table 4 shows the results of my re-estimating Chodorow-Reich's model (using his replication files) but weighting by initial state population (in December 2008) to make the estimates representative of national data. The point estimate falls to 1.15 and the standard error is higher at 0.72. The third row of Table 4 shows the estimates when spending across the levels of government are substituted for the ARRA spending. Here, I use the Chodorow-Reich combination of instruments, and I weight by initial state population. The jobs multiplier estimate is now 0.89 with a standard error of 0.45. Chodorow-Reich's method for converting jobs multipliers to output multipliers is nearly one-for-one, so the 0.89 estimate also implies an output multiplier around 0.9. Thus, once the cross-state estimates are made nationally representative and include all spending, they look very much like the aggregate estimates and lie below unity.

Two important caveats about these adjusted estimates are in order. First, reweighting by population gives very large influence to just a few of the 50 states. Second, the great instrument relevance in Chodorow-Reich's analysis disappears once I add state and local spending to American Recovery and Reinvestment Act spending. In other words, the instruments that are so good at explaining ARRA spending are not very good at explaining total government spending in the state.

Thus, it appears that the natural experiments exploited by the ARRA literature are rich enough to answer questions about the effects of ARRA spending on a cross-state basis, but not to answer questions about the aggregate effects of government spending induced by the ARRA.

In sum, a number of commentators and researchers have argued that multipliers may have been higher than usual after the financial crisis. I interpret most of the evidence at this point as suggesting that they were not higher than usual.

Conclusion

The fiscal literature has made tremendous progress in the ten years since the start of the global financial crisis. The range of estimates for average fiscal multipliers has been reduced considerably, particularly for government purchases. On average, government purchases multipliers are likely to be between 0.6 and 1. Narrative-based time series estimates point to tax rate change multipliers between –2 and –3, though these are significantly greater in magnitude than those predicted by New Keynesian DSGE models. However, there is still ongoing debate about specific contexts, such as the size of fiscal multipliers during "bad" times and the effects of other characteristics, such as exchange rate regimes.

Across industrialized countries, most of the temporary stimulus packages enacted from 2007 to 2009 in response to the global financial crisis took the form of transfer payments or lump-sum tax rebates (Oh and Reis 2012). Policymakers were "flying blind" in that they had little research to guide them at that time. Had they known then some of the results now emerging from the literature, they might have fashioned the stimulus packages differently, perhaps relying more on tax rate cuts and less on expenditures.

I believe the literature would benefit from progress in three main areas. First, the literature needs to catch up to the current policy discussions by focusing more on the short-run and long-run effects of infrastructure investment. The few studies at the aggregate and subnational levels suggest that these multipliers can be very large in some contexts (for example, Leduc and Wilson 2013). Second, researchers need to be careful about their implementation decisions. Seemingly small changes, such as how multipliers are actually calculated, can make a big difference. Finally, researchers should continue to innovate along the lines they have pursued in the last ten years, exploiting new datasets, extending theoretical models, and improving estimation techniques. As part of this innovation, researchers should continue to analyze the link between micro estimates and aggregate effects.

■ *I am grateful for helpful comments from Alberto Alesina, Gabriel Chodorow-Reich, Martin Eichenbaum, Carlo Favero, Mark Gertler, Francesco Giavazzi, Gordon Hanson, Daniel Leff Yaffe, Karel Mertens, Maury Obstfeld, Timothy Taylor, Linda Tesar, Sarah Zubairy, and participants at the July 2018 NBER Conference "Global Financial Crisis @10."*

References

Acconcia, Antonio, Giancarlo Corsetti, and Saverio Simonelli. 2014. "Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-Experiment." *American Economic Review* 104(7): 2185–2209.

Alesina, Alberto, Gualtiero Azzalini, Carlo Favero, Francesco Giavazzi, and Armando Miano. 2018. "Is it the 'How' or the 'When' that Matters in Fiscal Adjustments?" *IMF Economic Review* 66(1): 144–88.

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019. *Austerity: When It Works and When It Doesn't.* Princeton University Press.

Alloza, Mario. 2017. "Is Fiscal Policy More Effective in Uncertain Times or During Recessions?" Banco de España Working Paper 1730.

Aschauer, David Alan. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23(2): 177–200.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2018. "The Intertemporal Keynesian Cross." NBER Working Paper 25020.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4(2): 1–27.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2013. "Fiscal Multipliers in Recession and Expansion." Chap. 2 in *Fiscal Policy After the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi. University of Chicago Press.

Barro, Robert J. 1981. "Output Effects of Government Purchases." *Journal of Political Economy* 89(6): 1086–1121.

Barro, Robert J., and Charles J. Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics* 126(1): 51–102.

Baxter, Marianne, and Robert G. King. 1993. "Fiscal Policy in General Equilibrium." *American Economic Review* 83(3): 315–34.

Ben Zeev, Nadav, and Evi Pappa. 2017. "Chronicle of a War Foretold: The Macroeconomic Effects of Anticipated Defence Spending Shocks." *Economic Journal* 127(603): 1568–97.

Blanchard, Olivier J., and Daniel Leigh. 2013. "Growth Forecast Errors and Fiscal Multipliers." *American Economic Review* 103(3): 117–20.

Blanchard, Olivier J., and Daniel Leigh. 2014. "Learning about Fiscal Multipliers from Growth Forecast Errors." *IMF Economic Review* 62(2): 179–212.

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117(4): 1329–68.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90(430): 443–50.

Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari. 2015. "Estimating Fiscal Multipliers: News from a Nonlinear World." *Economics Journal* 125(584): 746–76.

Caldara, Dario, and Christophe Kamps. 2017. "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers." *Review of Economic Studies* 84(3): 1015–40.

Canova, Fabio, and Luca Sala. 2009. "Back to Square One: Identification Issues in DSGE Models." *Journal of Monetary Economics* 56(4): 431–49.

Canzoneri, Matthew, Fabrice Collard, Harris Dellas, and Behzad Diba. 2016. "Fiscal Multipliers in Recessions." *Economic Journal* 126(590): 75–108.

Chodorow-Reich, Gabriel. Forthcoming. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy.*

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo. 2011. "When is the Government Spending Multiplier Large?" *Journal of Political Economy* 119(1): 78–121.

Cloyne, James. 2013. "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom." *American Economic Review* 103(4): 1507–28.

Coenen, Günter, Christopher J. Erceg, Charles Freedman, Davide Furceri, Michael Kumhof, René Lalonde, Douglas Laxton, Jesper Lindé, Annabelle Mourougane, Dirk Muir, Susanna Mursula, Carlos de Resende, John Roberts, Werner Roeger, Stephen Snudden, Mathias Trabandt, and Jan in't Veld. 2012. "Effects of Fiscal Stimulus in Structural Models." *American Economic Journal: Macroeconomics* 4(1): 22–68.

Cogan, John F., Tobias Cwik, John B. Taylor, and Volker Wieland. 2010. "New Keynesian versus Old Keynesian Government Spending Multipliers." *Journal of Economic Dynamics and Control* 34(3): 281–95.

Corsetti, Giancarlo, André Meier, and Gernot J. Müller. 2012. "What Determines Government Spending Multipliers?" *Economic Policy* 27(72): 521–65.

Demirel, U. Devrim. 2016. "The Short-Term Effects of Tax Changes—Evidence for State Dependence." Congressional Budget Office Working Paper 2016-03.

Drautzburg, Thorsten, and Harald Uhlig. 2015. "Fiscal Stimulus and Distortionary Taxation." *Review of Economic Dynamics* 18(4): 894–920.

Dupor, Bill, and Rodrigo Guerrero. 2017. "Local and Aggregate Fiscal Policy Multipliers." *Journal of Monetary Economics* 92: 16–30.

Dupor, Bill, and M. Saif Mehkari. 2016. "The 2009 Recovery Act: Stimulus at the Extensive and Intensive Labor Margins." *European Economic Review* 85: 208–28.

Eskandari, Ruhollah. 2015. "State-Dependent Macroeconomic Effects of Tax Changes: Estimates Based on Narrative Records in the US." In *Three Essays on Tax Policy, Macroeconomics, and Corporate Policies*. University of Milan dissertation.

Farhi, Emmanuel, and Iván Werning. 2016. "Fiscal Multipliers: Liquidity Traps and Currency Unions." Chap. 31 in *Handbook of Macroeconomics*, Vol. 2, edited by John B. Taylor and Harald Uhlig, 2417–92. Amsterdam: Elsevier.

Fazzari, Steven M., James Morley, and Irina Panovska. 2015. "State-Dependent Effects of Fiscal Policy." *Studies in Nonlinear Dynamics and Econometrics* 19(3): 285–315.

Fisher, Jonas D. M., and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *Economic Journal* 120(544): 414–36.

Gali, Jordi, J. David López-Salido, and Javier Vallés. 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association* 5(1): 227–270.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2018. "Bartik Instruments: What, When, Why, and How." NBER Working Paper 24408, revised July 2018.

Gordon, Robert J., and Robert Krenn. 2010. "The End of the Great Depression 1939–41: Policy Contributions and Fiscal Multipliers." NBER Working Paper 16380.

Górnicka, Lucyna, Christophe Kamps, Gerrit Koster, and Nadine Leiner-Killinger. 2018. "Learning about Fiscal Multipliers during the European Sovereign Debt Crisis: Evidence from a Quasi-Natural Experiment." European Central Bank Working Paper 2154, May.

Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. "Expansionary Austerity? International Evidence." *Journal of the European Economic Association* 12(4): 949–68.

Hall, Robert E. 1980 "Labor Supply and Aggregate Fluctuations." *Carnegie-Rochester Conference Series on Public Policy* 12: 7–33.

Hall, Robert E. 2009. "By How Much Does GDP Rise If the Government Buys More Output?" *Brookings Papers on Economic Activity* 40(2): 183–231.

Hayo, Bernd, and Matthias Uhl. 2013. "The Macroeconomic Effects of Legislated Tax Changes in Germany." *Oxford Economic Papers* 66(2): 397–418.

House, Christopher L., Christian Proebsting, and Linda L. Tesar. 2017. "Austerity in the Aftermath of the Great Recession." NBER Working Paper 23147.

House, Christopher L., and Matthew D. Shapiro. 2006. "Phased in Tax Cuts and Economic Activity." *American Economic Review* 96(5): 1835–49.

Ilzetzki, Ethan, Enrique G. Mendoza, and Carlos A. Végh. 2013. "How Big (Small?) Are Fiscal Multipliers?" *Journal of Monetary Economics* 60(2): 239–254.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review* 96(5): 1589–1610.

Leduc, Sylvain, and Daniel Wilson. 2013. "Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment." *NBER Macroeconomics Annual* 27(1): 89–142.

Leduc, Sylvain, and Daniel Wilson. 2017. "Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act." *American Economic Journal: Economic Policy* 9(2): 253–92.

Leeper, Eric M., Nora Traum, and Todd B. Walker. 2017. "Clearing up the Fiscal Multiplier Morass." *American Economic Review* 107(8): 2409–54.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. 2013. "Fiscal Foresight and Information Flows." *Econometrica* 81(3): 1115–45.

Leigh, Daniel, Pete Devries, Charles Freedman, Jaime Guajardo, Douglas Laxton, and Andrea Pescatori. 2010. "Will it Hurt? Macroeconomic Effects of Fiscal Consolidation." Chap. 3 of *World Economic Outlook: Recovery, Risk, and Rebalancing*, October 2010. International Monetary Fund.

Mertens, Karel, and Morten O. Ravn. 2012. "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks." *American Economic Journal: Economic Policy* 4(2): 145–81.

Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103(4): 1212–47.

Mertens, Karel, and Morten O. Ravn. 2014. "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers." *Journal of Monetary Economics* 68

(December Supplement): S1–S19.

Michaillat, Pascal. 2014. "A Theory of Countercyclical Government Multiplier." *American Economic Journal: Macroeconomics* 6(1): 190–217.

Miyamoto, Wataru, Thuy Lan Nguyen, and Dmitriy Sergeyev. 2018. "Government Spending Multipliers Under the Zero Lower Bound." *American Economic Journal: Macroeconomics* 10(3): 247–77.

Mountford, Andrew, and Harald Uhlig. 2009. "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics* 24(6): 960–92.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions." *American Economic Review* 104(3): 753–92.

Nelson, Charles R., and Richard Startz. 1990. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica: Journal of the Econometric Society* 58(4): 967–76.

Oh, Hyunseung, and Ricardo Reis. 2012. "Targeted Transfers and the Fiscal Response to the Great Recession." *Journal of Monetary Economics* 59(Supplement): S50–S64.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy. 2013. "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data." *American Economic Review* 103(3): 129–34.

Parker, Jonathan A., Nicolas S. Souleles, David S. Johnson, and Robert McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103(6): 2530–53.

Pereira, Alfredo M., and Rafael Flores de Frutos. 1999. "Public Capital Accumulation and Private Sector Performance." *Journal of Urban Economics* 46(2): 300–322.

Ramey, Valerie A. 2011a. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126(1): 1–50.

Ramey, Valerie A. 2011b. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49(3): 673–85.

Ramey, Valerie A. 2013. "Government Spending and Private Activity." Chap. 1 in *Fiscal Policy after the Financial Crisis,* edited by Alberto Alesian and Francesco Giavazzi. University of Chicago Press.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." Chap 2 in *Handbook of Macroeconomics*, edited by John B. Taylor and Harald Uhlig. Vol. 2, 71–162. Amsterdam: Elsevier.

Ramey, Valerie A. Forthcoming. "Fiscal Policy: Tax and Spending Multipliers in the United States." In *Rethinking Macroeconomic Policy IV.* MIT Press.

Ramey, Valerie A., and Matthew D. Shapiro. 1998. "Costly Capital Reallocation and the Effects of Government Spending." *Carnegie Rochester Conference on Public Policy* 48(1): 145–194.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy* 126(2): 850–901.

Riera-Crichton, Daniel, Carlos A. Vegh, and Guillermo Vuletin. 2016. "Tax Multipliers: Pitfalls in Measurement and Identification." *Journal of Monetary Economics* 79: 30–48.

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100(3): 763–801.

Romer, Christina D., and David H. Romer. 2016. "Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Increases, 1952–1991." *American Economic Journal: Macroeconomics* 8(4): 1–42.

Sims, Eric, and Jonathan Wolff. 2018a. "The Output and Welfare Effects of Government Spending Shocks over the Business Cycle." *International Economic Review* 59(3): 1403–35.

Sims, Eric and Jonathan Wolff. 2018b. "The State-Dependent Effects of Tax Shocks." *European Economic Review* 107: 57–85.

Swanson, Eric T., and John C. Williams. 2014. "Measuring the Effect of the Zero Lower Bound on Medium-and Longer-Term Interest Rates." *American Economic Review* 104(10): 3154–85.

Woodford, Michael. 2011. "The Simple Analytics of the Government Expenditure Multiplier." *American Economic Journal: Macroeconomics* 3(1): 1–35.

Zubairy, Sarah. 2014. "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model." *International Economic Review* 55(1): 169–195.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 115–140

Rising Government Debt: Causes and Solutions for a Decades-Old Trend

Pierre Yared

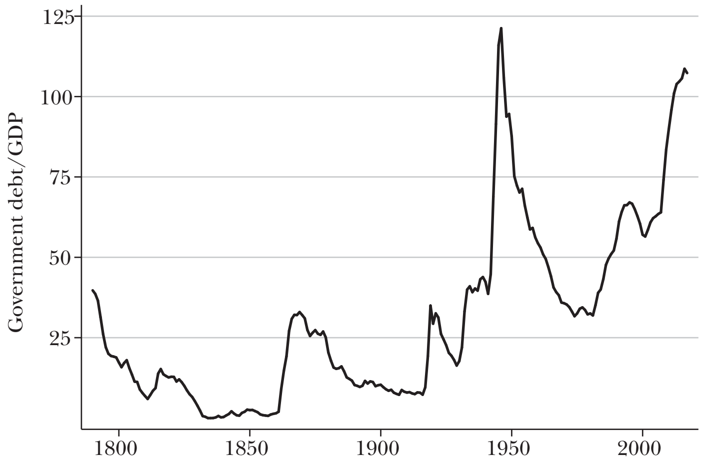
Since US government debt as a fraction of GDP reached a trough in the mid-1970s, it has been on a generally upward trajectory. As shown in Figure 1, it is now approaching levels not reached since World War II, and is projected to continue to increase significantly over the coming decade (Congressional Budget Office 2018, table 4.1). This is largely the result of a secular expansion of government spending—in particular, mandatory spending programs such as Social Security, Medicare, and Medicaid (as discussed in Blahous 2013)—with tax revenue not rising as rapidly. Between 1968 and 2017, spending on these three categories as a share GDP increased by an average annual rate of 2.4 percent, while tax revenue as a share of GDP grew by an average annual rate of 0.16 percent (based on Congressional Budget Office 2018, tables E.1 and E.5).

The United States is not alone. Advanced economies as a group have experienced a long-term increase in government debt to GDP, with France and Germany singled out as examples in Figure 2. The increase in government debt in most of these countries is also the result of tax revenue not keeping pace with the expansion of government spending. For example, between 1972 and 2016, central government tax revenue as a share of GDP increased in France, in Germany, and, more broadly, in the Organisation for Economic Cooperation and Development countries (based on World Bank data).

■ *Pierre Yared is a Professor of Business at Columbia Business School and a Director of the Richard Paul Richman Center for Business, Law, and Public Policy at Columbia University, New York, New York. His email address is pyared@columbia.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.115 doi=10.1257/jep.33.2.115

Figure 1
Gross Debt of the US Government
(percent)



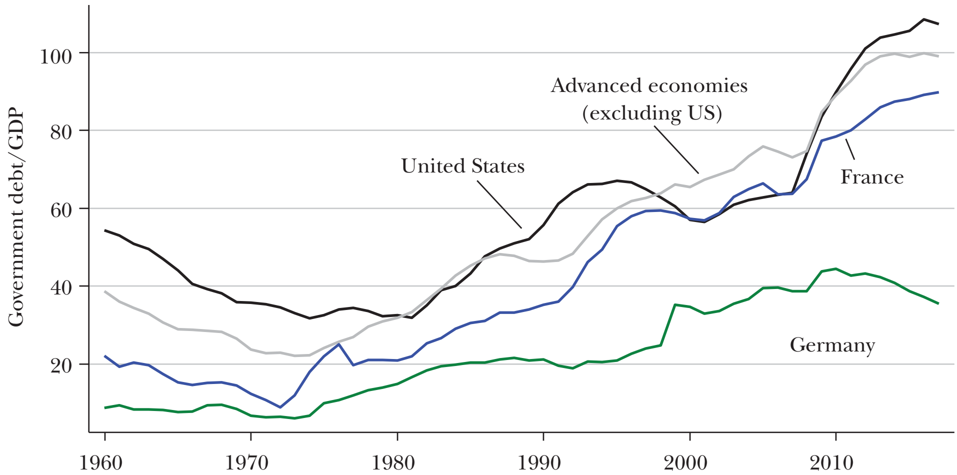
Source: Government debt to GDP is gross central government debt as a percentage of GDP from Reinhart and Rogoff (2011) for 1790–2010, updated for 2011–2017, with the growth rate in debt to GDP from the International Monetary Fund.
Note: I focus on gross central government debt as this measure is available for the broadest cross-section of advanced economies. But for the United States, all empirical observations in the paper are robust to replacing this gross measure with federal debt held by the public.

Large debt buildups can eventually lead to diminished economic activity, either by crowding out private capital investment or by forcing an increase in distortive taxes and a decrease in public investment to facilitate repayment.[1] Moreover, a government carrying such a high debt load may be constrained in responding to future catastrophes, such as financial crises, natural disasters, or wars (see Obstfeld 2013; Battaglini and Coate 2016; Romer and Romer 2018). In extreme cases, the result is default through explicit debt repudiation or inflation. There are many historical cases of default in advanced economies (Reinhart, Reinhart, and Rogoff 2015). The costs of default include increased stress on financial institutions, lower international financing for domestic firms, and decreased export market access (for discussion, see Borensztein and Panizza 2008; Hébert and Schreger 2017 and the references cited therein).

Has the rise in government debt over the past four decades served a socially beneficial purpose that would compensate for the risks of the added debt burden? In the first part of this article, I review normative macroeconomic theories in which government debt serves three possible functions: it can facilitate tax-smoothing, provide a safe asset, or sustain dynamic efficiency. I argue that, while the increased debt in certain periods may have been an optimal response to specific

[1] For an analysis of the empirical relationship between economic growth and public debt, see Reinhart, Reinhart, and Rogoff (2012) in this journal and Eberhardt and Presbitero (2015).

Figure 2
Government Debt in Advanced Economies
(percent)



Source: Government debt to GDP is gross central government debt as a percentage of GDP from Reinhart and Rogoff (2011) for 1960–2010, updated for 2011–2017 with the growth rate in debt to GDP from International Monetary Fund. GDP is from Feenstra, Inklaar, and Timmer (2015) for 1960–2014, and the 2014 GDP weight is assigned to 2015–2017.
Note: The sample of advanced economies is a balanced panel which includes Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States. The line for advanced economies (excluding the US) represents the GDP-weighted average for each observation year.

macroeconomic shocks, the broad-based long-run *trend* in debt accumulation seems inconsistent with these theories of optimal government debt policy.

I then review political economy theories of government debt. I argue that an increasingly older population, rising political polarization, and rising electoral uncertainty can explain the long-run trend in government debt across advanced economies. A resonating theme of these political economy theories is the time-inconsistency of government policy. Current governments want to be fiscally *irresponsible*, while simultaneously hoping that future governments will be fiscally *responsible*.

Thus, governments across the world have adopted fiscal rules—such as mandated deficit, spending, or revenue limits—to curtail future increases in government debt. In 2015, 92 countries had fiscal rules in place, a dramatic increase from 1990, when only seven countries had them (Lledó, Yoon, Fang, Mbaye, and Kim 2017). Fiscal rules must balance commitment not to overspend with flexibility to react to shocks. In the final part of this article, I describe some recent research on the optimal design of fiscal rules, elucidating the commitment-versus-flexibility tradeoff in theory and in practice. This discussion touches on how rules should be conditioned on public information, how they should be enforced, how they should be applied at a supranational level, whether they should feature escape clauses, and whether they should be based on fiscal policy tools or targets.

Optimal Government Debt Policy

Behind any theory of optimal government debt lurks the Ricardian equivalence proposition (Barro 1974). This proposition states that the level of government debt is irrelevant and has no effect on real economic activity because government borrowing can be undone by private actors. For example if the government cuts taxes and borrows today, the private sector anticipates a tax increase in the future by the government that needs to repay the debt. As a consequence, the private sector uses the tax cut today to save through government bonds to finance a higher future tax burden, and the government's decision to borrow more has no effect on consumption, labor, and capital investment decisions.

Ricardian equivalence requires three strong conditions that do not hold in practice. First, it assumes that raising tax revenue entails no deadweight loss, which is why the timing of revenue-raising does not directly distort consumption, labor, or capital investment decisions. Second, households and firms are assumed to be financially unconstrained and can thus borrow and lend freely at the same terms as the government. Finally, households and firms care about the level of taxes infinitely far into the future. I now turn to theories of optimal government debt that relax each of these three conditions and consider whether any of them can justify the overall pattern of rising government debt.

Tax-Smoothing: Unanticipated and Anticipated Fiscal Needs

The tax-smoothing argument is the most widely used theory of optimal government debt. If lump-sum taxes are ruled out so that raising tax revenue distorts economic decisions, whereas selling government bonds does not, then government debt allows the government to smooth the deadweight loss from raising tax revenue across time (for early examples, see Barro 1979 and Lucas and Stokey 1983; for recent examples, see Bhandari, Evans, Golosov, and Sargent 2017 and the references cited therein). However, the logic of this argument plays out differently if the fiscal needs are unanticipated versus anticipated. Let's discuss both cases.

The tax-smoothing argument suggests that a government facing unanticipated, temporary spending needs should respond optimally by increasing government debt. The logic is that financing these needs through immediate revenue-raising would be more costly for the economy in the short-term, and so it is better to issue debt to spread these costs into the future, when fiscal needs are lower.

It's easy enough to think of several unanticipated temporary fiscal needs that have caused government debt to increase across advanced economies. The global financial crisis, which started in 2007, put downward pressure on government revenues and upward pressure on the potential benefits of fiscal stimulus. In the United States, gross central government debt as a fraction of GDP increased from 64 percent in 2007 to 90 percent in 2010. During the same time frame, government debt to GDP in the euro area also increased, not only in countries heavily affected by the crisis such as Greece, Ireland, Italy, Portugal, and Spain, but also in countries less affected such as Germany and France (based on data in Reinhart and Rogoff

2011). Prior to the global financial crisis, the unanticipated wars in Afghanistan (2001–present) and Iraq (2003–2011) contributed to rising US government debt. US military spending as a fraction of GDP increased from 2.9 percent in 2000 to 3.8 percent in 2007 (based on World Bank data).

But while unanticipated temporary fiscal needs resulting from the global financial crisis and war can explain some of the increase in US debt in certain periods, they cannot explain either the long-term trend in government debt since the mid-1970s across advanced economies or the projected rise in the future.

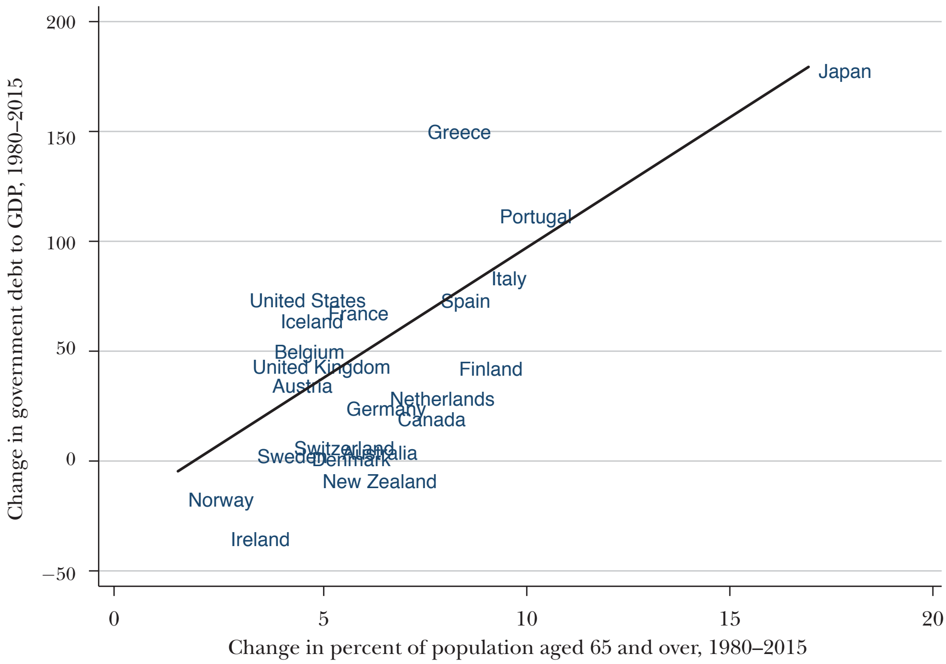
Can a combination of tax-smoothing theory and anticipated fiscal needs explain the long-term trend in public debt? The difficulty here is that, according to tax-smoothing theory, a government facing rising future fiscal pressures should pay down a larger portion of the debt in the present so as to alleviate forecasted fiscal strain.

Across advanced economies, the reduction in fertility rates and the extension of life spans have increased the elderly population, which in turn is leading to a long-term rise in fiscal pressures. In the sample of advanced economies used for Figure 2, the share of the population aged 65 and older has doubled, rising from 9.3 percent in 1960 to 18.5 percent in 2015 (based on data from the United Nations). Government spending on programs to assist the elderly have risen accordingly. Between 1980 and 2015, cash benefits to the elderly as a fraction of GDP across the OECD increased from 4.6 percent to 6.6 percent (as reported in the OECD "Social Expenditure Database"). In the United States, Social Security spending as a fraction of GDP increased from 2.6 percent in 1968 to 4.9 percent in 2017, while Medicare spending as a fraction of GDP during that time increased from 0.6 percent to 3.7 percent (Congressional Budget Office 2018, table E.5). This increase in mandatory spending was anticipated by historical US government forecasts which, on average, predicted larger increases than were realized (Congressional Budget Office 2017, table 2).

Rising government spending driven by promised payments to the elderly is likely to continue. These future commitments dwarf on-balance-sheet government debt. Hamilton (2014, table 5) estimates that in 2012, future Social Security and Medicare obligations were 4.8 times the size of on-balance-sheet debt. The European Central Bank (Lojsch, Rodríguez-Vives, and Slavík 2011, tables 4 and 11) estimates that in 2007, future pension entitlements in the euro area were five times the size of on-balance-sheet debt.

In the face of these well-anticipated demographic changes, tax-smoothing theory would have prescribed a general decumulation—not accumulation—of government debt during the past several decades. Moreover, tax-smoothing theory would have predicted lower debt accumulation in countries anticipating greater strain due to an aging population. Nevertheless, the cross-sectional data illustrated in Figure 3 shows the opposite: countries experiencing a greater increase in population aging, such as Japan, have accumulated more debt as a percentage of GDP than those experiencing a lower demographic strain, such as Canada. In sum, the long-term secular trend in government debt accumulation in the United States and across advanced economies cannot reflect an optimal policy response to either unanticipated or anticipated fiscal needs.

Figure 3
Change in Government Debt and Change in Elderly Population
(percentage point change)



Source: See Figure 2 for the government debt to GDP data source. Percent of population aged 65 and over is from the United Nations.
Note: The sample represents advanced economies, and is the same as for Figure 2. The regression represented by the fitted line yields a coefficient of 11.85 (standard error = 2.38), $N = 22$, and $R^2 = 0.55$.

Safe Asset Provision

A second theory of optimal government debt considers the role of public debt when the private sector cannot borrow or lend freely at the same terms as the government (for early examples, see Woodford 1990; Aiyagari and McGrattan 1998; Holmström and Tirole 1998; for recent examples, see Azzimonti and Yared forthcoming and the references cited therein). This theory builds on the fact that governments can borrow more cheaply than the private sector. As a result, when the government issues bonds, it slackens financial constraints on borrowers who now receive additional resources from the government (through tax cuts or government loans). In addition, the safe asset provision theory suggests that if financial constraints become tighter, an optimal policy response increases public debt to counteract the shrinking supply of safe assets for creditors, while simultaneously providing more liquidity to increasingly constrained borrowers. The safe asset role of optimal debt arises in various contexts: in the aftermath of the global financial crisis, during financial deregulation, under changing income risk, and during rising cross-border capital flows. I now examine whether these considerations in the context of the safe asset provision theory can justify the observed long-term trend in government debt in advanced economies.[2]

For example, the safe asset provision theory can certainly offer a justification for the increase in public debt in response to the global financial crisis. However, in the previous decades from 1980 and 2007, financial conditions did not tighten, but in general loosened through a global process of financial deregulation (for discussion, see Philippon and Reshef 2012). This deregulation came hand in hand with an increase in private sector leverage. The US economy, for example, saw household debt as a percent of income rise from 62 percent in 1980 to 123 percent in 2007 (Ahn, Batty, and Meisenzahl 2018). The safe asset provision theory suggests that such a relaxation of financial constraints should have been met with a decrease, as opposed to an increase, in public debt.

The safe asset provision theory also suggests that public debt should increase in response to rising income risk, because households and businesses facing greater income risk develop a stronger precautionary motive to save, driving down interest rates. The optimal policy response increases the supply of public debt to satisfy the increased demand for safe assets, as Azzimonti, de Francisco, and Quadrini (2014) illustrate in a quantitative model.

But evidence from US administrative data suggests that household income risk actually *declined* in the decades after 1980 (for example, Sabelhaus and Song 2010; Guvenen, Ozkhan, and Song 2014), while business-level analyses of trends in risk have found mixed results (for example, Comin and Philippon 2005; Davis, Haltiwanger, Jarmin, and Miranda 2006 and the references cited therein). From this perspective, the safe asset provision theory thus offers little support for an increase in public debt.

The safe asset provision theory also applies in the context of the dramatic expansion of cross-border flows in the last four decades, a response to the reduction of international barriers in trade and finance. This trend accelerated in the aftermath of the Asian financial crisis of 1997 and the introduction of China into the World Trade Organization in 2001, when many nations began substantially increasing their US dollar reserves. The ensuing large capital inflows into advanced economies—a phenomenon known as the "global saving glut"—led to a deterioration of net foreign asset position for some advanced economies and to a decline in global interest rates (Bernanke 2005). For example, between 1995 and 2015, US net foreign assets decreased from –5 percent of GDP to –42 percent of GDP (based on data from Lane and Milesi-Ferretti 2018).

[2] For this discussion, I am implicitly considering the implications for an economy with heterogeneous households consisting of borrowers and lenders. An alternative approach considers hand-to-mouth homogeneous households in an open economy. Because the government's objective in this case is to smooth private consumption over time through taxes and transfers matched with fluctuating government borrowing from abroad, the analysis of this environment is isomorphic to a tax-smoothing framework. For further discussion on the isomorphism between tax-smoothing and consumption-smoothing frameworks, see Barro (1979) and Aiyagari, Marcet, Sargent, and Seppälä (2002).

From the perspective of safe asset provision theory, the optimal policy response to greater globalization and capital inflows is ambiguous. These phenomena should reduce the cost of public borrowing for two reasons: 1) an increase in asset demand by foreigners reduces interest rates and the cost of issuing public debt; and 2) globalization expands the market for safe assets, thereby reducing the marginal interest rate response to additional public debt issuance. But on the other side, additional borrowing by the domestic private sector (in response to lower interest rates) means that domestic borrowers suffer more from marginal interest rate increases induced by higher public debt.[3]

Beyond this theoretical ambiguity, there are other reasons that the long-term trend in public debt across advanced economies does not appear to be an optimal policy response to globalization. First, government debt in advanced economies had been on an upward trajectory well before the onset of the global saving glut in the late 1990s, as shown earlier. Second, prior to the late 1990s, the degree of cross-border public debt holdings had been relatively stable, suggesting that the globalization of public debt markets was limited up until that point. For example, in the case of the United States, the fraction of government debt that was held by foreigners remained around 15 to 20 percent between 1980 and 1995 and increased significantly thereafter, reaching 46 percent in 2009 (Aizenman and Marion 2011, figure 7). Finally, the safe asset provision theory would predict that, all else fixed, smaller countries respond to globalization by increasing public debt proportionately more than larger countries, because globalization decreases the interest rate response to debt issuance by more for small countries. However, the relationship between country size and debt issuance for advanced economies during this period is actually positive: in the advanced economy sample from Figure 2, the change in debt to GDP from 1980 to 2017 has a correlation of 0.41 with (the log of) 1980 GDP. Two large economies with especially large increases in their public debt-to-GDP ratios over this time are the United States and Japan.

Dynamic Efficiency

A final theory, less explored in the research literature, considers the role of public debt when the private sector does not internalize the effect of fiscal policy infinitely far into the future (for example, Diamond 1965; Blanchard 1985). In such an environment, older households do not face the future tax cost of issuing government debt today, because any taxes will be repaid by future generations. As a consequence, an increase in government debt tilts the lifetime consumption profile towards older generations, while also increasing interest rates and crowding out capital investment. Under some conditions, the possibility of a bubble in

[3] The three channels highlighted here, together with an ambiguous optimal policy response, emerge if one extends the two-period model of Azzimonti and Yared (2017) by introducing foreign asset demand (details available upon request). Azzimonti, de Francisco, and Quadrini (2014) also illustrate the second channel in a model with symmetric countries individually choosing policy. Another approach to this question additionally considers the risk of default and inflation by the government (for example, see Farhi and Maggiori 2018).

government debt arises, whereby one generation is willing to hold government debt purely because future generations are also expected to do so.

From this perspective, if an economy is dynamically inefficient and has overaccumulated capital, increasing government debt can be optimal. Inefficient capital overaccumulation can emerge in equilibrium when agents have finite horizons, in which case a bubble in government debt can improve welfare (for discussion, see Tirole 1985). However, there is no evidence of capital overaccumulation in the United States or advanced economies in the post-World War II period (Abel, Mankiw, Summers, and Zeckhauser 1989).[4]

Political Economy Forces behind Rising Government Debt

The absence of a clear normative reason for the trend in government debt across advanced economies suggests that political forces are behind this pattern. In this section, I review political economy theories of government debt, with a focus on rational theories driven by political self-interest. I argue that, over the past four decades, changes in specific political factors can explain the long-run trajectory of government debt.[5]

In theoretical terms, the political factors that I describe imply that a government behaves similarly to an agent with present-biased and dynamically inconsistent preferences, which economists often analyze using a hyperbolic discounting model (for example, Laibson 1997). In the context of fiscal policy, quasi-hyperbolic preferences imply that the government at a given date t weighs periods $\{t, t+1, t+2, \ldots\}$ according to discount factors $\{1, \beta\delta, \beta\delta^2, \ldots\}$, for some time preference factor $\delta \in (0,1)$ and present bias $\beta \in (0,1)$. This creates a familiar problem of dynamic inconsistency. Consider the weight the government assigns to date $t+2$ relative to date $t+1$. From the perspective of date t, this weight is $(\beta\delta^2)/(\beta\delta) = \delta$, but from the perspective of date $t+1$, this weight is $\beta\delta < \delta$. Thus, a government subject to present bias will always want to apply the discount factor δ to future time periods, in line with what is socially optimal. However, when those time periods actually arrive and become the present, the present bias β becomes relevant. The government becomes like a person who always wants to start exercising or eating healthier tomorrow, but never wants to start today.

In fiscal policy, any political factor that amplifies the present bias results in larger deficits (from higher spending or lower taxes) and changes the long-term trend in government debt. In addition, a government with this kind of present bias

[4] Geerolf (2018) reaches the same result when applying the methodology of Abel, Mankiw, Summers, and Zeckhauser (1989) to more recent US data. Using a different methodology and data, however, this work finds less-strong evidence in favor of dynamic efficiency.

[5] In contrast to rational theories, "fiscal illusion" theory emphasizes voters' behavioral biases and their potential inability to understand the long-term costs of deficits (for example, Buchanan and Wagner 1977). This theory does not lead voters to demand commitment devices, such as the fiscal rules discussed in the next section. Moreover, it is not clear whether the time-series and cross-country patterns in behavioral biases—to the extent these could be measured—would explain the empirical evidence on public debt.

will recognize that it would like to be more patient in the future, but will probably be unable to do so, and thus will be interested in implementing fiscal rules as a commitment device (as discussed in the next main section of this paper). Examples of fiscal policy applications that make use of quasi-hyperbolic preferences include Aguiar and Amador (2011) and Halac and Yared (2014, 2018a, 2019).

In the next subsections, I describe several political factors that provide a microfoundation for the present bias and the dynamic inconsistency of government preferences. I document how these factors have evolved and offer an explanation for the long-run trend in government debt. I focus here on long-run considerations and ignore variation in present bias over the political business cycle. For a starting point in that literature, see Ales, Maziero, and Yared (2014, and the references cited therein).[6]

Aging and Heterogeneous Discounting

Households differ in how much they weigh the present relative to the future. These differences can be the result of demographics, with older households caring less about the future than younger households. This is consistent with survey evidence on intergenerational differences in policy preferences, with younger households placing a larger value on fiscal responsibility than older households (Parker 2012; Wolter, Hansen, Campbell, and Ansolabehere 2013). In a political environment in which policy is chosen sequentially without commitment, as is common in a representative democracy that has not imposed long-term fiscal rules on itself, this heterogeneity implies a present bias together with dynamically inconsistent preferences for the government.

Conceptually, heterogeneity in discount rates means that impatient households wield disproportionate influence in policymaking in the present period. If commitment were possible, impatient households would agree in advance to allow the patient households to have more political influence in the future, because those households value the future more. However, nothing can stop impatient households from also deciding to influence policy when later time periods become the present. Jackson and Yariv (2014, 2015) formalize this idea and show that with any heterogeneity in preferences, every nondictatorial aggregation method that respects unanimity must be time-inconsistent; moreover, any such method that is time-separable must lead to a present bias.

This theory suggests that the greater fraction of old impatient households relative to young patient households, the more shortsighted is the government, the larger

[6] Alesina and Passalacqua (2016) offer a survey of the literature on the political economy of public debt. Even in the absence of the long-run forces that I describe, government debt can deviate from the normative benchmark if a government is benevolent but lacks commitment to the path of interest rates or to repaying debt (for a starting point in that literature, see Chari and Kehoe 1993; Debortoli, Nunes, and Yared 2017 and the references cited therein). However, whether this form of lack of commitment on its own leads to debt that is higher or lower than is optimal is ambiguous and depends on various economic considerations. For this reason, I focus on how lack of commitment combined with additional political factors leads to excessive debt.

are government deficits, and the faster is government debt accumulation. Arguments along these lines emerge in the models of Cukierman and Meltzer (1989) and Tabellini (1991).[7] This theory explains the long-term trend in government debt in advanced economies as a result of an aging population. In addition, this theory is consistent with the cross-country trends displayed earlier in Figure 3, where government debt has grown faster in countries experiencing a larger increase in the elderly population.

Tragedy of the Commons

Shortsighted policymaking can also result from a version of the tragedy of the commons in which political parties acting independently engage in excessive targeted government spending since they do not internalize the shared financing costs of government debt. Weingast, Shepsle, and Johnsen (1981) take this approach in a static fiscal framework, while Velasco (2000) offers an example of this approach in a dynamic framework.

As an illustration, consider N symmetric parties that can make targeted deficit-financed spending appropriations to their constituencies in the present, simultaneously and without coordinating. Then each party fails to internalize the total cost of additional debt because the burden of this debt is shared equally across parties in the future; from the party's perspective, the cost of one additional unit of debt due to targeted spending is $1/N$ of the total cost. The result is excessive spending and government debt accumulation, which would be alleviated if parties jointly committed in advance to limiting borrowing. This lack of coordination leads the government to be present-biased and time-inconsistent in its fiscal policy. Moreover, this present bias is amplified when there is greater disagreement in spending priorities across political parties (Hertzberg 2016).[8]

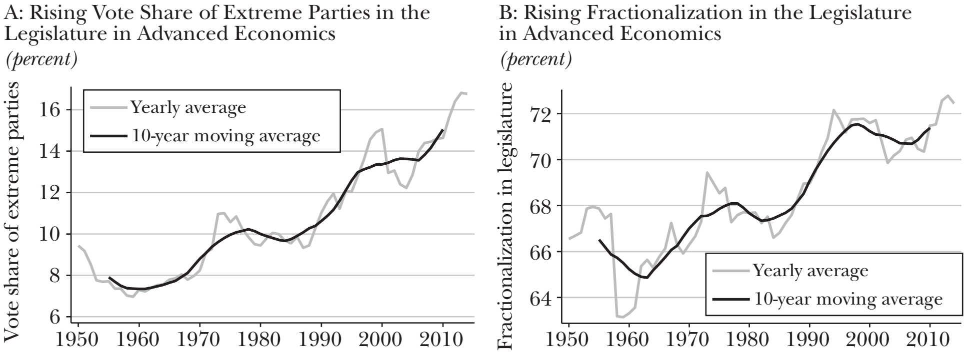
Even in the absence of domestic political disagreement, a related coordination problem can emerge across countries, particularly if these countries are highly integrated financially, as in the euro area. Individual countries may fail to internalize the impact of their borrowing decisions on the shared interest rates, inflation rates, or probability of financial contagion. Azzimonti, de Francisco, and Quadrini (2014) discuss excessive borrowing in the context of shared interest rates, as do we in Halac and Yared (2018a), while Beetsma and Uhlig (1999), Chari and Kehoe (2007), and Aguiar, Amador, Farhi, and Gopinath (2015) do so in the context of shared inflation rates. Either way, the result is inefficiently high public debt accumulation across countries. These mechanisms also apply to subnational governments that can issue their own debt (Dovis and Kirpalani 2017).

The tragedy of the commons predicts that countries with a large number of constituencies or deep disagreements in fiscal priorities across constituencies will

[7] Song, Storesletten, and Zilibotti (2012) show that this present bias can be mitigated if current generations care more about future generations than future generations care about current generations.

[8] Hertzberg (2016) captures disagreement as the relative weight placed on targeted-transfers versus mutually beneficial public goods. This work establishes an equivalence result which, under certain assumptions on preferences, links the intertemporal behavior of multiple time-consistent agents suffering from the tragedy of the commons with that of a single time-inconsistent agent with quasi-hyperbolic preferences.

Figure 4
Polarization and Fractionalization in the Legislature in Advanced Economies



Source: Measures on the y-axis come from Funke, Schularick, and Trebesch (2016).
Note: The sample represents a balanced panel of advanced economies used for Figure 2 with available vote share and fractionalization data. The sample excludes Greece, Iceland, New Zealand, Portugal, and Spain, for which data is not available for all years. For Figure 4A, the measure is the percent of the popular vote for extreme parties on the far right or the far left for the most recent election in the lower legislature. For Figure 4B, the fractionalization measure represents the probability that two members of the lower legislature are from different political parties.

incur larger government deficits, resulting in faster government debt accumulation. This prediction is consistent with empirical work that has found that larger deficits are associated with countries with more ministers, with greater ideological polarization in the executive, and with a proportional (as opposed to majoritarian) election system (for evidence, see Woo 2003; Persson and Tabellini 2004; Crivelli, Gupta, Mulas-Granados, and Correa-Caro 2016).

Through the lens of this theory, the long-term trend in government debt in advanced economies is a result of the increase in political polarization and fragmentation across these economies. Evidence on this rise in polarization in the United States comes from many sources. The Partisan Conflict Index of Azzimonti (2018), which is based on the number of newspaper articles reporting political disagreement in a given month, shows an increase in partisan conflict in the United States since the late 1960s. This trend is consistent with evidence from other advanced economies, which have witnessed a declining influence of centrist political parties. Figure 4A shows that across advanced economies, the share of the legislative vote going to parties of the extreme left or extreme right has been on the rise since the 1960s. Figure 4B calculates the probability that any two members of the legislature are from different political parties, and by this measure finds a pattern of increasing political fractionalization in legislatures since the 1960s in advanced economies.

Political Turnover

A large literature focuses on political turnover as an explanation for rising government debt; early examples include Persson and Svensson (1989) and Alesina and Tabellini (1990), while more recent examples include Battaglini and Coate

(2008) and Yared (2010). In this case, present bias results from the interaction of two factors: 1) the *temporary* concentration of political authority in one political party, which derives additional benefits from spending while in power by boosting its popularity, concentrating government resources on preferred initiatives, or increasing wasteful rents; and 2) the inability of parties to make binding (intertemporal) commitments to one another.

Conceptually, the realization (or threat) of political turnover causes the current government to be impatient, since the party holding power recognizes that it may not have the opportunity to benefit from spending in the future.[9] This present bias is more severe if the temporary benefits from spending and rent-seeking while in office are large, if there are more parties competing for power, if only a subset of parties can make decisions at any time (as in Battaglini and Coate 2008), or if there is more political risk. In addition to overweighing the present relative to the future, government preferences are dynamically inconsistent; that is, the party presently in power would prefer that future governments be fiscally responsible, but future governments cannot commit to future policy. In this sense, the combination of lack of commitment together with political risk causes the government to be present-biased and time-inconsistent. Aguiar and Amador (2011) offer a formal analysis along these lines.

This theory predicts that countries with more rent-seeking, political fragmentation, or political risk will incur larger government deficits, resulting in faster government debt accumulation. These predictions are in line with empirical cross-country studies on the determinants of government deficits (for example, Drazen 2000; Alt and Lassen 2016).

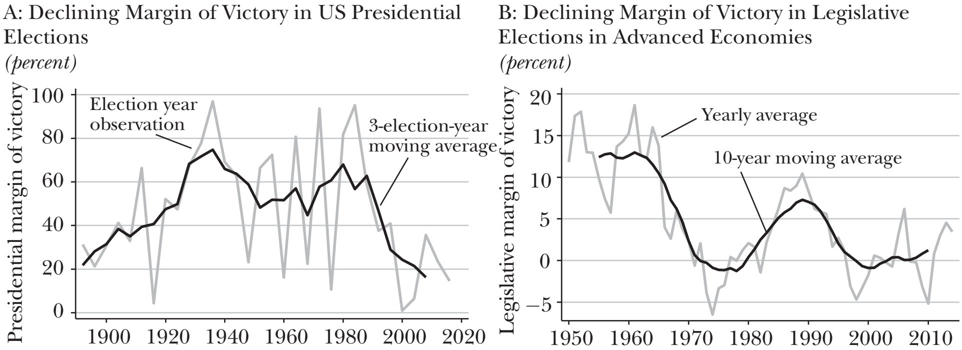
This theory can explain the long-term trend in government debt in advanced economies as a result of rising political uncertainty for parties in power. The margin of victory in US presidential elections has been in decline since the mid-1980s, as shown in Figure 5A, suggesting that elections have become closer and less predictable. Similarly, analysis of US congressional elections has documented a declining incumbency advantage since the mid-1980s (for example, Jacobson 2015). This trend is consistent with the evidence from advanced economies in Figure 5B, which displays a decline in the average popular vote margin in legislative elections for the governing party or coalition.

Assessment

Political economy theories of government debt can qualitatively explain the long-term trend in government debt accumulation across advanced economies. Nevertheless, these theories leave several unanswered questions for future research.

[9] Persson and Svensson (1989) and Müller, Storesletten, and Zilibotti (2016) argue that the present bias may be more severe if the current party in power leans to the right and puts higher relative weight on tax cuts versus government spending increases.

Figure 5
Declining Margin of Victory in Elections



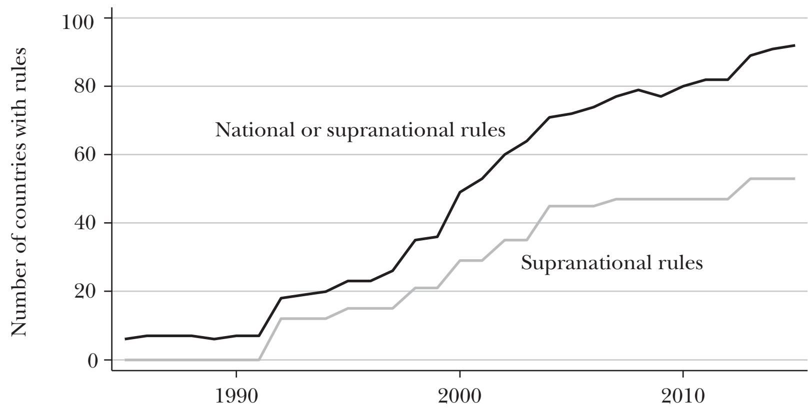
Source: Electoral margin of victory for presidential elections is from US Electoral College. Margin of victory for the most recent election in the lower legislature is from Funke, Schularick, and Trebesch (2016).
Note: For the US presidental margin of victory (Figure 5A), the measure is the difference in electoral votes received by the winner and the runner-up as a percentage of the sum of votes received by the winner and the runner-up. For the legislative margin of victory in advanced economies (Figure 5B), the margin is the difference between popular votes received by the legislators in the governing party or governing coalition and the votes received by those in the opposition party or coalition, as a percentage of the sum of votes received by the two groups. The sample for Figure 5B represents a balanced panel of the advanced economies from Figure 2 with available margin of victory data. The sample excludes Finland, Greece, Iceland, New Zealand, Portugal, and Spain, for which data is not available for all years.

First, it is unclear whether certain political economy models can quantitatively match the time-series and cross-sectional patterns in advanced economy government debt. Second, political economy theories do not explain why polarization and electoral uncertainty have increased in advanced economies, nor how this development may have been caused by certain economic trends or policies. For example, McCarty, Poole, and Rosenthal (2006) argue that polarization and income inequality reinforce each other, and Baker, Bloom, Canes-Wrone, Davis, and Rodden (2014) provide evidence that higher government spending, taxes, and polarization have contributed to increased policy-related economic uncertainty in the United States. Finally, current political economy theories do not directly address the change in the composition of government spending, which has become increasingly concentrated in old-age government assistance programs. A plausible explanation is that increasingly competitive political parties both change the composition of government spending and increase government debt in their efforts to appeal to an aging constituency.

Fiscal Rules to Constrain Rising Debt

Every political explanation for rising debt discussed in the previous section is based on time-inconsistency in government preferences. Current governments want

Figure 6
Number of Countries with Fiscal Rules



Source: Data is from the International Monetary Fund.
Note: A country is classified as having a fiscal rule if it is subject to an expenditure rule, a revenue rule, a budget balance rule, or a debt rule.

to be fiscally *irresponsible*, while simultaneously hoping that future governments be fiscally *responsible*. Thus, governments across the world have sought to adopt fiscal rules—such as mandated deficit, spending, or revenue limits—to restrict future fiscal policy and curtail the increase in government debt. Figure 6 illustrates the growing number of countries that have imposed fiscal rules. For a complete description of the fiscal rule adopted in each country see Lledó et al. (2017).

Fiscal rules have been adopted at the subnational, national, and supranational levels. In some countries, such rules have been an effective force. For example, in Switzerland the ratio of government debt to GDP rose from 13 percent in 1990 to 29 percent in 2003, but after a fiscal rule was adopted that year, the Swiss debt-to-GDP ratio declined back to 20 percent of GDP by 2016 (based on World Bank data; see Pfeil and Feld 2016 for a discussion). In contrast, the United States is currently subject to national-level spending caps passed in the Budget Control Act of 2011, which were subsequently increased by Congress in 2013, 2015, and 2018. These caps do not apply to most mandatory spending items underlying the growth in debt (Capretta 2014). For a broader discussion of the effectiveness of national and supranational rules at reducing debt, see Wyplosz (2012) and Eyraud, Debrun, Hodge, Lledó, and Pattillo (2018). For analysis of subnational rules, see Primo (2007, and the references therein) for the US experience, and Grembi, Nannicini, and Troiano (2016) for the case of Italy.

In this section, I describe research on the optimal design of fiscal rules. I begin with the fundamental tradeoff of fiscal rules between commitment and flexibility. My discussion then touches on how fiscal rules should be conditioned on public information, how they should be enforced, how they should be applied at a supranational level, whether they should feature escape clauses, and whether they should be based on fiscal policy tools or targets.

Commitment versus Flexibility

Fiscal rules entail a fundamental tradeoff. On the one hand, rules provide commitment to counteract the present bias in policymaking; on the other hand, there is a cost of reduced flexibility because fiscal rules cannot spell out policy prescriptions for every possible shock or contingency, and so some discretion may be optimal. This commitment-versus-flexibility tradeoff is familiar in macroeconomics. For example, it also arises in discussions of monetary policy rules (for an early example, see Rogoff 1985; for recent examples, see Kocherlakota 2016 and the references cited therein). It also arises in principal-agent theory in the study of delegation (for an early example, see Holmström 1977; for recent examples, see Amador and Bagwell 2013 and the references cited therein).

There are two approaches to the theoretical analysis of this tradeoff. One approach restricts the structure of a fiscal rule to a form used in practice—such as a deficit limit—and evaluates the stringency of an optimal rule (for examples, see Azzimonti, Battaglini, and Coate 2016; Halac and Yared 2018a and the references cited therein). The other approach does not restrict the structure of a fiscal rule and uses mechanism design to characterize simultaneously the structure and the stringency of an optimal rule (for example, Amador, Werning, and Angeletos 2006; Halac and Yared 2014, 2016, 2018b, 2019). This second approach distinguishes between fiscally relevant information on which a fiscal rule can explicitly depend—such as the level of public debt or GDP—and relevant information on which a fiscal rule cannot explicitly depend—such as the depth of a financial crisis or the wartime needs of the military. This latter type of information can be thought of as the government's *private information.* Such information may be observable but not contractible, or it may be literally private information. The latter case arises, for example, if the exact cost of public goods is only known to the government, or if the government has superior information about the aggregate preferences of heterogeneous citizens (as in Sleet 2004; Piguillem and Schneider 2016). In any case, an optimal fiscal rule then is represented as a policy prescription that maximizes social welfare subject to the government's private information and degree of present bias.

The advantage of the first approach is that it can be used to assess real world rules and evaluate the costs and benefits of partial reform in a framework that incorporates a rich set of macroeconomic and political forces. The advantage of the second approach is that it can be used to evaluate the costs and benefits of global—as opposed to partial—reform. This second approach also elucidates how other considerations, on top of private information and present bias, may contribute to the determination of an optimal rule. These two approaches complement each other and provide useful lessons for the optimal design of fiscal rules.[10]

In the next subsections, I discuss what the tradeoff between commitment and flexibility implies for various features of fiscal rules in theory and in practice.

[10] The difference between these two approaches is analogous to that between the Ramsey and Mirrlees approaches to optimal taxation, a distinction discussed by Mankiw, Weinzerl, and Yagan (2009) in this journal.

Conditioning on Information

An optimal fiscal rule relies on fiscally relevant information that is observable and contractible, like the level of debt and GDP. Because not all fiscally relevant information can be easily observed and verified, this rule may allow some discretion. Amador, Werning, and Angeletos (2006) show that, under certain assumptions on the distribution of private information and the government's preferences, the optimal rule takes a simple threshold form, such as the deficit, spending, or revenue limits observed in practice. (Without these assumptions, an optimal rule is more complex and can involve multiple policy thresholds.) In general, the optimal threshold is tighter the smaller is the volatility of the government's private information and the more severe is the government's present bias, as in both cases the value of commitment is increased relative to the value of flexibility.

Setting optimal fiscal thresholds is challenging. First, there are practical questions regarding implementation. Recent research has been devoted to examining which macroeconomic measures should be used to set a threshold, how to weigh the relative importance of these measures, and how to set the numerical targets so as to afford sufficient flexibility while simultaneously preventing excessive debt growth. For example, Azzimonti, Battaglini, and Coate (2016) analyze the short- and long-term costs and benefits of adopting a balanced budget amendment in the United States. Alfaro and Kanczuk (2016) compare the performance of a debt-independent deficit limit to a pure debt limit for Brazil. Eyraud, Baum, Hodge, Jarmuzek, Kim, Mbaye, and Türe (2018) offer a general discussion of the challenges in calibrating fiscal rules.

Second, there are questions regarding the dynamic determination of optimal fiscal rules when some fiscally relevant information is not contractible. In Halac and Yared (2014), we show that if the government's private information is persistent over time, an optimal fiscal rule should condition on the extent to which past policies agreed with fiscal targets, even if this measure is irrelevant for optimal policy determination. This implies fiscal thresholds that change in response to past policy decisions, unlike the thresholds that would be optimal when the government's private information is independent over time. How to incorporate such considerations into real-world fiscal rules in a practical way is an interesting area for future research.

Enforcement

According to the International Monetary Fund, governments comply with their fiscal rules only about half of the time (Eyraud, Debrun, Hodge, Lledó, and Pattillo 2018). Violation of fiscal rules can trigger either a formal or informal enforcement mechanism. For example, in the European Union, an Excessive Deficit Procedure—a sequence of costly fiscal adjustments and potential sanctions—is set in motion when a rule is breached (as described in Lledó et al. 2017, p. 81). In Chile, penalties for fiscal rule violation have been informal. In Halac and Yared (2017), we describe an episode in 2009 in which breach of the fiscal rule by the Chilean administration was informally punished by the next administration, which continued to ignore the rule. This example highlights a potential self-enforcement mechanism: a current government may follow the fiscal rule because it does not wish to set a precedent of rule abandonment to be followed by future governments.

How should fiscal rules be structured under limited enforcement? In Halac and Yared (2019), we explore a commitment-versus-flexibility framework where punishments for rule violation are limited and socially costly. We show that under some conditions, the optimal rule is a maximally enforced threshold—namely a deficit, spending, or revenue limit that triggers the largest feasible penalty whenever violated. Whereas graduated punishments would be less socially costly, they would also induce less fiscal discipline. Furthermore, we show that fiscal thresholds that are never violated by the government may be suboptimal. This is the case if extreme shocks to the economy are sufficiently rare and a lax fiscal rule achieves little discipline. Tightening the rule so that it is violated under extreme shocks is then beneficial; as the expected cost of punishment following violation is small relative to the gain in fiscal discipline in normal times.

There are several issues to take into account when considering punishments for breaking fiscal rules. First, whether or not rules have been broken might be unclear. There are numerous examples of how governments can use creative accounting to circumvent rules. Frankel and Schreger (2013) describe how euro-area governments use overoptimistic growth forecasts to comply with fiscal rules. Many US states compensate government employees with future pension payments, which increases off-balance-sheet entitlement liabilities not subject to fiscal rules (Bouton, Lizzeri, and Persico 2016). In 2016, President Dilma Rousseff of Brazil was impeached for illegally using state-run banks to pay government expenses and bypass the fiscal responsibility law (Leahy 2016). Given this transparency problem, many countries have established independent fiscal councils to assess and monitor compliance with fiscal rules (Debrun et al. 2013).

A second issue to consider is the credibility of punishments. As an example, the Excessive Deficit Procedure against France and Germany in 2003 was stalled by disagreement between the European Commission and the European Council; consequently, French and German deficits persisted without penalty (as discussed in Gros, Mayer, and Ubide 2004). In Halac and Yared (2019), we argue that in the absence of institutionalized penalties, the temporary abandonment of rules combined with overspending—as in the Chilean case previously described—can serve as its own deterrent for breaking a fiscal rule. Unlike sanctions that are harmful to all parties, a punishment in the form of future rule abandonment and overspending may be credible, as it benefits the recipients of this overspending.

A third issue is the response of the private sector to the violation of rules, which can also serve as a form of punishment. For example, Eyraud, Debrun, Hodge, Lledó, and Pattillo (2018) find that the violation of fiscal rules is associated with a significant increase in interest rate spreads for sovereign borrowing. Such an increase in financing costs immediately penalizes a government for breaching a rule. This idea can be formalized in a model of government debt and default which features multiple equilibria resulting from self-fulfilling market expectations (as in Calvo 1988).

Coordinated Rules

More than half of the countries with fiscal rules are subject to rules that apply at a supranational level: examples include the European Union's Stability and Growth Pact, the West African Economic and Monetary Union, the Central African Economic and Monetary Community, and the Eastern Caribbean Currency Union. Among European countries under EU fiscal rules, more than a dozen also have additional rules at the national level. For example, Germany is constrained by its own constitutionally mandated "debt brake," which imposes a tighter limit on the government's structural deficit than the EU Stability and Growth Pact (Truger and Will 2013).

The main argument for imposing rules at a supranational level relates to the tragedy of the commons argument presented earlier. Individual countries in an integrated economic region do not internalize the impact of their borrowing decisions on the shared interest rates, inflation rates, or probability of financial contagion. Supranational fiscal rules can limit this externality.

However, supranational fiscal rules come with numerous challenges. First, the imposition of uniform thresholds for multiple countries under a supranational rule may be inappropriate if countries are likely to differ in the level or volatility of their fiscal needs or in the severity of their government's present bias. Hatchondo, Martinez, and Roch (2017) argue that conditioning thresholds on market signals, like the interest spread on sovereign borrowing, allows supranational rules to be tailored more effectively to individual countries.

Second, the design of rules at a supranational level must account for the disciplining effect of interest rates (Halac and Yared 2018a). Excessively tight supranational rules not only reduce flexibility, but they promote fiscal irresponsibility by reducing regional interest rates and governments' cost of funding. For example, Fernández-Villaverde, Garicano, and Santos (2013) argue in this journal that the drop in interest rates that followed European integration led to the abandonment of reforms and institutional deterioration in the peripheral European countries. In addition, countries that complement supranational rules with more stringent rules at the national level—as in the case of Germany in the European Union—exert an externality by driving down regional interest rates and reducing fiscal discipline in other countries. In Halac and Yared (2018a), we show that when this interest rate externality and the resultant international imbalances are large enough, supranational rules must be made more stringent to reduce imbalances.[11]

Finally, whether supranational rules are easier or harder to enforce than national rules is an open theoretical and empirical question. On one hand, the international economic system provides more tools for sanctioning, and the supranational sanctioning authority may be less subject to domestic political pressures. On the other hand, the enforcement of supranational rules faces a collective action

[11] The same logic can also justify rules that limit current account surpluses (and indirectly budget surpluses), such as the Macroeconomic Imbalance Procedure in the European Union (European Commission 2016, table 3.2).

problem, and disagreement over whether enforcement should occur may be an impediment, as in the case described previously concerning the enforcement of the European Union's Excessive Deficit Procedure in 2003.

Escape Clauses

Many governments' fiscal rules feature an escape clause that allows violating the rule under exceptional circumstances (Lledó et al. 2017). Triggering an escape clause typically involves a review process, which culminates in a final decision by an independent fiscal council, a legislature, or citizens via a referendum. In Switzerland, for example, the government can deviate from a fiscal rule with a legislative supermajority in the cases of natural disaster, severe recession, or changes in accounting method.

The cost of triggering an escape clause deters governments from using them too frequently. Moreover, because these costs largely involve a facilitation of information gathering to promote efficient fiscal policy, escape clauses are useful even in the presence of perfect rule enforcement.

In Halac and Yared (2016), we study fiscal rules that make use of escape clauses in a commitment-versus-flexibility framework; Coate and Milton (2017) also study this. These papers find that introducing escape clause provisions is generally optimal if (privately observed) fiscal shocks are sufficiently volatile, the government's present bias is sufficiently severe, and the resource cost of triggering an escape clause is sufficiently low. In such a situation, a rule with an escape clause dominates a pure threshold rule by allowing for more flexibility in response to extreme economic conditions.

In practice, the use of escape clause provisions can be challenging. The interpretation of events in which escape clauses can be triggered is subjective, and the political deliberation surrounding an appropriate fiscal response can be uncertain and induce delay. As an example, Primo (2007) discusses the problems in implementing escape clauses in the fiscal rules of US states. Whether these costs can outweigh the benefits of using escape clauses is an open empirical question.

Instrument-Based and Target-Based Rules

How should fiscal rules be applied? Should the government face constraints directly on instruments of policy, such as spending, or should the fiscal rule concern targets of policy, such as deficits? Which instruments and targets ought to be addressed?

In practice, fiscal rules can constrain different instruments of policy, such as specific categories of government spending or tax rates. Different instruments may call for different thresholds, as the associated commitment-versus-flexibility tradeoff may not be the same (as Galperti 2019 explains in the context of personal budgeting). For instance, due to volatile geopolitical conditions, military spending needs may be less forecastable than other spending needs, and may thus demand more flexibility. Capital spending is another category where allowing increased flexibility may be optimal, as the benefits of capital spending accrue well into the future and are thus subject to a less-severe present bias. Thus, many countries have "golden rules," which limit spending net of a government's capital expenditure. Poterba (1995) and Bassetto and Sargent (2006) address the benefits of a "golden rule" in the context of US states.

Overall, the evidence suggests that rules that distinguish across categories are indeed associated with better fiscal and macroeconomic outcomes (for discussion, see Eyraud, Lledó, Dudine, and Peralta 2018). Moreover, it can be optimal to set multiple layers of rules, for example specifying a fiscal threshold for individual categories of taxes and spending as well as on the total level of taxes and spending in the form of a (forecasted) deficit rule. Multiple-layer rules are particularly beneficial when there are complementarities across different fiscal instruments.

Similar principles apply to the analysis of target-based rules, which identify targets for outcomes of policy, such as the deficit-to-GDP ratio. A target-based rule specifies an economic goal, while giving the government greater instrument discretion to respond to changing macroeconomic conditions. However, given the risk of macroeconomic surprises, a government may be penalized for rule breach despite its best efforts. Therefore, an optimal target threshold should be tight enough that it induces the government to rein in its present bias, but not so tight that it is excessively prone to violations due to macroeconomic surprises.

In Halac and Yared (2018b), we develop a theoretical framework to compare these different classes of rules, using an extended delegation setting that incorporates a noisy observable outcome. We show that target-based rules dominate instrument-based rules if the government is sufficiently well informed, so that instrument discretion is beneficial and punishment due to macroeconomic surprises is relatively unlikely. We also show how a simple hybrid rule—which allows for an instrument threshold that is relaxed whenever a target threshold is satisfied—would do better than either of these two classes. Bohn and Inman (1996) analyze fiscal rules of US states and find that target-based rules, in the form of end-of-the-year fiscal requirements, perform better than instrument-based rules, in the form of beginning-of-the-year fiscal requirements.

Concluding Remarks

Over the past four decades, government debt as a fraction of GDP has been on an upward trajectory in advanced economies, and the US government debt to GDP is approaching levels not reached since World War II. This paper has argued that political economy theories can explain this long-run trend as resulting from an aging population, rising political polarization, and rising electoral uncertainty.

Many countries have adopted fiscal rules to rein in growing debts. Most of these rules were recently introduced, and time will tell whether they lead to sustainable government finances and to a reversal of this decades-old trend. Their success depends, in part, on whether they appropriately balance the tradeoff between commitment and flexibility underpinning these rules, and whether they address other challenges that I have highlighted, such as enforceability.

This discussion suggests several interesting questions for future research. First, while I have focused on fiscal rules as a solution to growing debts, the introduction of fiscal rules should be combined with additional reforms to budgetary procedures. How specific procedural rules, such as voting or amendment rules, complement or thwart the effect of fiscal rules is an important issue to consider; for example, Capretta (2014) suggests reforms to the US budget process that would allow Congress to change entitlement policy more easily. Second, a government's deficit bias is not constant, because it evolves over time in response to factors such as changing polarization and electoral uncertainty. Understanding how these underlying political forces are impacted by fiscal policy and by the introduction of fiscal rules is important for governments contemplating rule adoption. Finally, the introduction and implementation of fiscal rules requires a level of political consensus and stability, which often occurs when the need for a fiscal rule is less salient. How to take advantage of the occasions to adopt and improve fiscal rules when they arise, rather than letting them pass and missing the opportunity, is critical for limiting the growth of government debt.

■ *I would like to thank Marina Azzimonti, V. V. Chari, Steve Coate, François Geerolf, Mark Gertler, Fatih Guvenen, Marina Halac, Gordon Hanson, Andrew Hertzberg, Leo Martinez, Chris Moser, Suresh Naidu, Ziad Obermeyer, Serdar Ozkan, Facundo Piguillem, Ken Rogoff, Jesse Schreger, Timothy Taylor, Christoph Trebesch, and Alan Viard for comments. Trisha Sinha and George Vojta provided excellent research assistance.*

References

Abel, Andrew B., N. Gregory Mankiw, Lawrence H. Summers, and Richard J. Zeckhauser. 1989. "Assessing Dynamic Efficiency: Theory and Evidence." *Review of Economic Studies* 56(1): 1–20.

Aguiar, Mark, and Manuel Amador. 2011. "Growth in the Shadow of Expropriation." *Quarterly Journal of Economics* 126(2): 651–97.

Aguiar, Mark, Manuel Amador, Emmanuel Farhi, and Gita Gopinath. 2015. "Coordination and Crisis in Monetary Unions." *Quarterly Journal of Economics* 130(4): 1727–79.

Ahn, Michael, Michael Batty, and Ralf R. Meisenzahl. 2018. "Household Debt-to-Income Ratios in the Enhanced Financial Accounts." FEDS Notes, Board of Governors of the Federal Reserve System, January 11, 2018. https://doi.org/10.17016/2380-7172.2138.

Aiyagari, Rao S., Albert Marcet, Thomas J. Sargent, and Juha Seppälä. 2002. "Optimal Taxation without State-Contingent Debt." *Journal of Political Economy* 110(6): 1220–54.

Aiyagari, Rao S., and Ellen R. McGrattan. 1998. "The Optimum Quantity of Debt." *Journal of Monetary Economics* 42(3): 447–69.

Aizenman, Joshua, and Nancy Marion. 2011. "Using Inflation to Erode the US Public Debt." *Journal of Macroeconomics* 33(4): 524–41.

Ales, Laurence, Pricila Maziero, and Pierre Yared. 2014. "A Theory of Political and Economic Cycles." *Journal of Economic Theory* 153: 224–51.

Alesina, Alberto, and Andrea Passalacqua. 2016. "The Political Economy of Government

Debt." Chap. 33 in *Handbook of Macroeconomics*, vol. 2, edited by John B. Taylor and Harald Uhlig. Amsterdam: Elsevier.

Alesina, Alberto, and Guido Tabellini. 1990. "A Positive Theory of Fiscal Deficits and Government Debt." *Review of Economic Studies* 57(3): 403–414.

Alfaro, Laura, and Fabio Kanczuk. 2016. "Fiscal Rules and Sovereign Default." NBER Working Paper 23370.

Alt, James E., and David Dreyer Lassen. 2006. "Transparency, Political Polarization, and Political Budget Cycles in OECD Countries." *American Journal of Political Science* 50(3): 530–50.

Amador, Manuel, and Kyle Bagwell. 2013. "The Theory of Optimal Delegation with an Application to Tariff Caps." *Econometrica* 81(4): 1541–99.

Amador, Manuel, Iván Werning, and George-Marios Angeletos. 2006. "Commitment vs. Flexibility." *Econometrica* 74(2): 365–96.

Azzimonti, Marina. 2018. "Partisan Conflict and Private Investment." *Journal of Monetary Economics* 93: 114–31.

Azzimonti, Marina, Marco Battaglini, and Stephen Coate. 2016. "The Costs and Benefits of Balanced Budget Rules: Lessons from a Political Economy Model of Fiscal Policy." *Journal of Public Economics* 136: 45–61.

Azzimonti, Marina, Eva de Francisco, and Vincenzo Quadrini. 2014. "Financial Globalization, Inequality, and the Rising Public Debt." *American Economic Review* 104(8): 2267–2302.

Azzimonti, Marina, and Pierre Yared. 2017. "A Note on Optimal Fiscal Policy in an Economy with Private Borrowing Limits." *Economics Letters* 151: 62–65.

Azzimonti, Marina, and Pierre Yared. Forthcoming. "The Optimal Public and Private Provision of Safe Assets." *Journal of Monetary Economics.*

Baker, Scott R., Nicholas Bloom, Brandice Canes-Wrone, Steven J. Davis, and Jonathan Rodden. 2014. "Why Has US Policy Uncertainty Risen Since 1960?" *American Economic Review* 104(5): 56–60.

Barro, Robert J. 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy*, 82(6): 1095–1117.

Barro, Robert J. 1979. "On the Determination of the Public Debt." *Journal of Political Economy* 87(5): 940–971.

Bassetto, Marco, and Thomas J. Sargent. 2006. "Politics and Efficiency of Separating Capital and Ordinary Government Budgets." *Quarterly Journal of Economics* 121(4): 1167–1210.

Battaglini, Marco, and Stephen Coate. 2008. "A Dynamic Theory of Public Spending, Taxation, and Debt." *American Economic Review* 98(1): 201–236.

Battaglini, Marco, and Stephen Coate. 2016. "A Political Economy Theory of Fiscal Policy and Unemployment." *Journal of the European Economic Association* 14(2): 303–337.

Beetsma, Roel, and Harald Uhlig. 1999. "An Analysis of the Stability and Growth Pact." *Economic Journal* 109(458): 546–71.

Bernanke, Ben S. 2005. "The Global Saving Glut and the U.S. Current Account Deficit." Remarks at the Sandridge Lecture, Virginia Association of Economists, Richmond, Virginia, March 10, 2015. https://www.federalreserve.gov/boarddocs/speeches/2005/200503102/.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas J. Sargent. 2017. "Fiscal Policy and Debt Management with Incomplete Markets." *Quarterly Journal of Economics* 132(2): 617–63.

Blahous, Charles P. 2013. "Why We Have Federal Deficits: The Policy Decisions that Created Them." Research Paper, Mercatus Center, George Mason University. https://www.mercatus.org/publication/why-we-have-federal-deficits-policy-decisions-created-them.

Blanchard, Olivier J. 1985. "Debt, Deficits, and Finite Horizons." *Journal of Political Economy* 93(2): 223–47.

Bohn, Henning, and Robert P. Inman. 1996. "Balanced-Budget Rules and Public Deficits: Evidence from the U.S. States." *Carnegie-Rochester Conference Series on Public Policy* 45: 13–76.

Borensztein, Eduardo, and Ugo Panizza. 2008. "The Costs of Sovereign Default." *International Monetary Fund Staff Papers* 56(4): 683–741.

Bouton, Laurent, Alessandro Lizzeri, and Nicola Persico. 2016. "The Political Economy of Debt and Entitlements." NBER Working Paper 22570.

Buchanan, James M., and Richard E. Wagner. 1977. *Democracy in Deficit: The Political Legacy of Lord Keynes.* New York: Academic Press.

Calvo, Guillermo A. 1998. "Servicing the Public Debt: Role of Expectations." *American Economic Review* 78(4): 647–61.

Capretta, James C. 2014. "Reforming the Budget Process." *National Affairs* 21(Fall): 64–80.

Chari, V. V., and Patrick J. Kehoe. 1993. "Sustainable Plans and Mutual Default." *Review of Economic Studies* 60(1): 175–95.

Chari, V. V., and Patrick J. Kehoe. 2007. "On the Need for Fiscal Constraints in a Monetary Union." *Journal of Monetary Economics* 54(8): 2399–2408.

Coate, Stephen, and Ross Milton. 2017. "Optimal Fiscal Limits with Overrides." http://rossmilton.com/optfisclim17.pdf.

Comin, Diego, and Thomas Philippon. 2005. "The Rise in Firm-Level Volatility: Causes and Consequences." *NBER Macroeconomics Annual* 20: 167–201.

Congressional Budget Office. 2017. "An Evaluation of CBO's Past Outlay Projections." November. www.cbo.gov/publication/53328.

Congressional Budget Office. 2018. "The Budget and Economic Outlook: 2018 to 2028." April. www.cbo.gov/publication/53651.

Crivelli, Ernesto, Sanjeev Gupta, Carlos Mulas-Granados, and Carolina Correa-Caro. 2016. "Fragmented Politics and Public Debt." IMF Working Paper 16-190.

Cukierman, Alex, and Allan H. Meltzer. 1989. "A Political Theory of Government Debt and Deficits in a Neo-Ricardian Framework." *American Economic Review* 79(4): 713–32.

Davis, Steven J., John Haltiwanger, Ron Jarmin, and Javier Miranda. 2006. "Volatility and Dispersion in Business Growth Rates: Publicly Traded Versus Privately Held Firms." *NBER Macroeconomics Annual* 21: 107–179.

Debortoli, Davide, Ricardo Nunes, and Pierre Yared. 2017. "Optimal Time-Consistent Government Debt Maturity." *Quarterly Journal of Economics* 132(1): 55–102.

Debrun, Xavier, Tidiane Kinda, Teresa Curristine, Luc Eyraud, Jason Harris, and Johann Seiwald. 2013. "The Functions and Impact of Fiscal Councils." IMF Policy Paper, July 16.

Diamond, Peter A. 1965. "National Debt in Neoclassical Growth Model." *American Economic Review* 55(5): 1126–50.

Dovis, Alessandro, and Rishabh Kirpalani. 2017. "Fiscal Rules, Bailouts, and Reputation in Federal Governments." NBER Working Paper 23942.

Drazen, Allan. 2000. "The Political Business Cycle after 25 Years." *NBER Macroeconomics Annual* 15: 75–117.

Eberhardt, Markus, and Andrea F. Presbitero. 2015. "Public Debt and Growth: Heterogeneity and Non-Linearity." *Journal of International Economics* 97: 45–58.

European Commission. 2016. "The Macroeconomic Imbalance Procedure: Rationale, Process, Application: A Compendium." European Commission Institutional Working Paper 039. https://ec.europa.eu/info/sites/info/files/file_import/ip039_en_2.pdf.

Eyraud, Luc, Anja Baum, Andrew Hodge, Mariusz Jarmuzek, Young Kim, Samba Mbaye, and Elif Türe. 2018. "How to Calibrate Fiscal Rules: A Primer." How-To Notes, March. Fiscal Affairs Department, International Monetary Fund.

Eyraud, Luc, Xavier Debrun, Andrew Hodge, Victor Lledó, and Catherine Pattillo. 2018. "Second-Generation Fiscal Rules: Balancing Simplicity, Flexibility, and Enforceability." International Monetary Fund Staff Discussion Note 18-04.

Eyraud, Luc, Victor Duarte Lledó, Paolo Dudine, and Adrian Peralta. 2018. "How to Select Fiscal Rules: A Primer." How-To Notes, March, Fiscal Affairs Department, International Monetary Fund.

Farhi, Emmanuel, and Matteo Maggirori. 2018. "A Model of the International Monetary System." *Quarterly Journal of Economics* 133(1): 295–355.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105(10): 3150–82. The Database. Penn World Table version 9.0, https://www.ggdc.net/pwt/ (accessed June 6, 2018).

Fernández-Villaverde, Jesús, Luis Garicano, and Tano Santos. 2013. "Political Credit Cycles: The Case of the Eurozone." *Journal of Economic Perspectives* 27(3): 145–66.

Frankel, Jeffrey, and Jesse Schreger. 2013. "Over-Optimistic Official Forecasts and Fiscal Rules in the Eurozone." *Review of World Economics* 149(2): 247–72.

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2016. "Going to Extremes: Politics after Financial Crises, 1870–2014." *European Economic Review* 88: 227–60.

Galperti, Simone. 2019. "A Theory of Personal Budgeting." *Theoretical Economics* 14: 173–210.

Geerolf, François. 2018. "Reassessing Dynamic Efficiency." https://fgeerolf.github.io/r-g.pdf.

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano. 2016. "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics* 8(3): 1–30.

Gros, Daniel, Thomas Mayer, and Angel Ubide. 2004. "The Nine Lives of the Stability and Growth Pact: A Special Report of the CEPS Macroeconomic Policy Group." Centre for European Policy Studies Paperpacks. https://www.ceps.eu/publications/nine-lives-stability-pact.

Guvenen, Fatih, Serdar Ozkhan, and Jae Song. 2014. "The Nature of Countercyclical Income Risk." *Journal of Political Economy* 122(3): 621–60.

Halac, Marina, and Pierre Yared. 2014. "Fiscal Rules and Discretion under Persistent Shocks." *Econometrica* 82(5): 1557–1614.

Halac, Marina, and Pierre Yared. 2016. "Commitment vs. Flexibility with Costly Verification." NBER Working Paper 22936.

Halac, Marina, and Pierre Yared. 2017. "Fiscal Rules and Discretion under Self-Enforcement." NBER Working Paper 23919.

Halac, Marina, and Pierre Yared. 2018a. "Fiscal Rules and Discretion in a World Economy." *American Economic Review* 108(8): 2305–34.

Halac, Marina and Pierre Yared. 2018b. "Instrument-Based vs. Target-Based Rules." NBER

Working Paper 24496.

Halac, Marina, and Pierre Yared. 2019. "Fiscal Rules and Discretion under Limited-Enforcement." NBER Working Paper 25463.

Hamilton, James D. 2014. "Off-Balance-Sheet Federal Liabilities." *Cato Papers on Public Policy* 3: 1–40.

Hatchondo, Juan Carlos, Leonardo Martinez, and Francisco Roch. 2017. "Fiscal Rules and the Sovereign Default Premium." https://drive.google.com/file/d/0B28k8Kom2VDBZzhGQnFLeVFCLUE/view.

Hébert, Benjamin, and Jesse Schreger. 2017. "The Costs of Sovereign Default: Evidence from Argentina." *American Economic Review* 107(10): 3119–45.

Hertzberg, Andrew. 2016. "Time-Consistent Individuals, Time-Inconsistent Households." https://www0.gsb.columbia.edu/faculty/ahertzberg/papers/HH_Paper_Sept_2016_v9.pdf.

Holmström, Bengt. 1977. "On Incentives and Control in Organizations." PhD dissertation, Stanford University.

Holmström, Bengt, and Jean Tirole. 1998. "Private and Public Supply of Liquidity." *Journal of Political Economy* 106(1): 1–40

International Monetary Fund. No date. "Central Government Debt." https://www.imf.org/external/datamapper/CG_DEBT_GDP@GDD/CHN/FRA/DEU/ITA/JPN/GBR/USA (accessed January 15, 2019).

International Monetary Fund. No date. "Fiscal Rule Dataset." International Monetary Fund Fiscal Affairs Department. http://www.imf.org/external/datamapper/fiscalrules/map/map.htm (accessed June 6, 2018).

Jackson, Matthew O., and Leeat Yariv. 2014. "Present Bias and Collective Dynamic Choice in the Lab." *American Economic Review* 104(12): 4184–04.

Jackson, Matthew O., and Leeat Yariv. 2015. "Collective Dynamic Choice: The Necessity of Time Inconsistency." *American Economic Journal: Microeconomics* 7(4): 159–178.

Jacobson, Gary C. 2015. "It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections." *Journal of Politics* 77(3): 861–73.

Kocherlakota, Narayana. 2016. "Rules versus Discretion: A Reconsideration." *Brookings Paper on Economic Activity*, Fall.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112(2): 443–478.

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2018. "The External Wealth of Nations Revisited: International Financial Integration in the Aftermath of the Global Financial Crisis." *IMF Economic Review* 66(1): 189–222. https://link.springer.com/article/10.1057%2Fs41308-017-0048-y (accessed January 21, 2019).

Leahy, Joe. 2016. "What Is Brazil's President Dilma Rousseff Accused Of?" *Financial Times*, May 12.

Lledó, Victor, Sungwook Yoon, Xiangming Fang, Smaba Mbaye, and Young Kim. 2017. "Fiscal Rules at a Glance." March. International Monetary Fund Background Paper. https://www.imf.org/external/datamapper/fiscalrules/Fiscal%20Rules%20at%20a%20Glance%20-%20Background%20Paper.pdf.

Lojsch, Dagmar Hartwig, Marta Rodríguez-Vives, and Michal Slavík. 2011. "The Size and Composition of Government Debt in the Euro Area." ECB Occasional Paper 132, European Central Bank.

Lucas Jr., Robert E., and Nancy L. Stokey. 1983. "Optimal Fiscal and Monetary Policy in an Economy without Capital." *Journal of Monetary Economics* 12(1): 55–93.

Mankiw, N. Gregory, Matthew Weinzierl, and Danny Yagan. 2009. "Optimal Taxation in Theory and in Practice." *Journal of Economic Perspectives* 23(4): 147–74.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge: MIT Press.

Müller, Andreas, Kjetil Storeletten, and Fabrizio Zilibotti. 2016. "The Political Color of Fiscal Responsibility." *Journal of the European Economic Association* 14(1): 252–302.

Obstfeld, Maurice. 2013. "On Keeping Your Powder Dry: Fiscal Foundations of Financial and Price Stability." *Monetary and Economic Studies* 31: 25–37.

Organisation for Economic Cooperation and Development. No date. "Social Expenditure Database." http://www.oecd.org/social/expenditure.htm (accessed January 9, 2019).

Persson, Torsten, and Lars E.O. Svensson. 1989. "Why a Stubborn Conservative Would Run a Deficit: Policy with Time-Inconsistent Preferences." *Quarterly Journal of Economics* 104(2): 325–45.

Persson, Torsten, and Guido Tabellini. 2004. "Constitutional Rules and Fiscal Policy Outcomes." *American Economic Review* 94(1): 25–45.

Pew Research Center. 2012. "The Big Generation Gap at the Polls Is Echoed in Attitudes on Budget Tradeoffs: But Public Sees Only Modest Conflict between Young and Old." Washington, DC: Pew Research Center. http://assets.pewresearch.org/wp-content/uploads/sites/3/2012/12/FINAL_policies_report.pdf.

Pfeil, Christian F., and Lars P. Feld. 2016. "Does the Swiss Debt Brake Induce Sound Federal

Finances? A Synthetic Control Analysis." CESIFO Working Paper 6044. https://www.cesifo-group.de/DocDL/cesifo1_wp6044.pdf.

Philippon, Thomas, and Ariell Reshef. 2012. "Wages and Human Capital in the U.S. Finance Industry: 1909–2006." *Quarterly Journal of Economics* 127(4): 1551–1609.

Piguillem, Facundo, and Anderson Schneider. 2016. "Coordination, Efficiency and Policy Discretion." EIEF Working Papers Series 1306, Einaudi Institute of Economics and Finance.

Poterba, James M. 1995. "Capital Budgets, Borrowing Rules, and State Capital Spending." *Journal of Public Economics* 56(2): 165–187.

Primo, David M. 2007. *Rules and Restraints: Government Spending and the Design of Institutions.* University of Chicago Press.

Reinhart, Carmen M., Vincent Reinhart, and Kenneth Rogoff. 2012. "Public Debt Overhangs: Advanced-Economy Episodes since 1800." *Journal of Economic Perspectives* 26(3): 69–86.

Reinhart, Carmen M., Vincent Reinhart, and Kenneth Rogoff. 2015. "Dealing with Debt." *Journal of International Economics* 96(1): S43–S55.

Reinhart, Carmen M. and Kenneth S. Rogoff. 2011. "From Financial Crash to Debt Crisis." *American Economic Review* 101(5): 1676–1706.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* (10094): 1169–89.

Romer, Christina D., and David H. Romer. 2018. "Why Some Times Are Different: Macroeconomic Policy and the Aftermath of Financial Crises." *Economica* 85(337): 1–40.

Sabelhaus, John, and Jae Song. 2010. "The Great Moderation in Micro Labor Earnings." *Journal of Monetary Economics* 57(4): 391–403.

Sleet, Christopher. 2004. "Optimal Taxation with Private Government Information." *Review of Economic Studies* 71(4): 1217–39.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2012. "Rotten Parents and Disciplined Children: A Political-Economic Theory of Public Expenditure and Debt." *Econometrica* 80(6): 2785–2803.

Tabellini, Guido. 1991. "The Politics of Intergenerational Redistribution." *Journal of Political Economy* 99(2): 335–57.

Tirole, Jean. 1985. "Asset Bubbles and Overlapping Generations." *Econometrica* 53(6): 1499–1528.

Truger, Achim, and Henner Will. 2013. "The German 'Debt Brake': A Shining Example for European Fiscal Policy?" *Revue de l'OFCE,* no. 127, pp 153–88.

US Electoral College. No date. "Historical Election Results." National Archives and Records Administration of U.S. Electoral College. https://www.archives.gov/federal-register/electoral-college/votes/index.html (accessed June 6, 2018).

United Nations. 2017. No date. "World Population Prospects 2017." United Nations Population Division. https://population.un.org/wpp/Download/Standard/Population/ (accessed January 9, 2019).

Velasco, Andrés. 2000. "Debts and Deficits with Fragmented Fiscal Policymaking." *Journal of Public Economics* 76(1): 105–25.

Weingast, Barry R., Kenneth A. Shepsle, and Christopher Johnsen. 1981. "The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics." *Journal of Political Economy* 89(4): 642–64.

Wolter, Kirk, Mark Hansen, Andrea Campbell, and Stephen Ansolabehere. 2013. "2012 NORC Presidential Election Study: Americans' Views on Entitlement Reform and Health Care." http://www.norc.org/PDFs/NORC_Inauguration%20Report-FINAL.pdf.

Woo, Jaejoon. 2003. "Economic, Political, and Institutional Determinants of Public Deficits." *Journal of Public Economics* 87(3–4): 387–426.

Woodford, Michael. 1990. "Public Debt as Private Liquidity." *American Economic Review* 80(2): 382–88.

World Bank. No date. "World Development Indicators." https://data.worldbank.org/products/wdi (accessed January 21, 2019).

Wyplosz, Charles. 2012. "Fiscal Rules: Theoretical Issues and Historical Experiences." In *Fiscal Policy After the Financial Crisis,* edited by Alberto Alesina and Francesco Giavazzi, 495–529. University of Chicago Press.

Yared, Pierre. 2010. "Politicians, Taxes and Debt." *Review of Economic Studies* 77(2): 806–840.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 141–162

Effects of Austerity: Expenditure- and Tax-based Approaches

Alberto Alesina, Carlo Favero, and Francesco Giavazzi

Sometimes governments need to reduce their budget deficits aggressively. These policies are labeled "austerity." Almost always austerity is needed because excessive debt has been accumulated, as a result of policy mistakes and political distortions (Alesina and Passalacqua 2016; Yared, in this issue).

The austerity policies embraced by several European countries starting in 2010 have generated an extraordinarily harsh policy debate. One side has argued that austerity is (almost) always a bad idea. From this perspective, even European countries that were experiencing serious difficulties in financial markets—either by being totally cut off from borrowing like Greece, or by paying high risk premia like Portugal, Spain, Ireland, and Italy—should have continued to stimulate their economies with high levels of government spending. Austerity, the argument continues, was self-defeating because the recessions it induced, or extended, only increased government debt as a ratio of GDP. Blanchard and Leigh (2014) argued that this round of austerity was particularly costly: in other words, fiscal multipliers were especially high. The other side argued that postponing austerity would have caused

■ *Alberto Alesina is Professor of Political Economy, Harvard University, Cambridge, Massachusetts. Carlo Favero is Professor of Finance, and Francesco Giavazzi is Professor of Economics, both at Bocconi University, Milan, Italy. Alesina and Giavazzi are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. All three authors are Fellows at the Innocenzo Gasparini Institute for Economic Research (IGIER), Università Bocconi, Milan, Italy, and also Research Fellows at the Centre for Economic Policy Research, London, United Kingdom. Their email addresses are aalesina@harvard.edu, carlo.favero@unibocconi.it, and francesco.giavazzi@unibocconi.it.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.141 doi=10.1257/jep.33.2.141

debt defaults and bank runs, another round of financial collapses and, possibly, the crumbling of the European monetary union with unpredictable and potentially disastrous economic and political consequences.

In this paper, we argue that the focus on austerity as such misses an important distinction: austerity based upon spending cuts is much less costly than that based upon tax increases. In our 2019 book, *Austerity: When It Works and When it Doesn't*, we documented close to 200 austerity plans in 16 high-income OECD economies from the late 1970s until 2014. These plans have been reconstructed consulting original documents concerning about 3,500 individual fiscal measures.[1] Our analysis of these episodes finds a large and statistically significant difference between the effects on output of expenditure-based and tax-based austerity plans. On average, an expenditure-based austerity plan the size of 1 percent of GDP implies a loss of about 1/4 of a percentage point of GDP and lasts less than two years. In contrast, tax-based austerity plans of the same size on average generate losses of more than two percentage points of GDP and the effect lasts 3–4 years. Of course, these averages conceal a broader range of outcomes. We even find a few cases of "expansionary austerity"—namely cases in which the output costs associated with an expenditure-based austerity plan have instead turned out to be output gains. Examples include Ireland, Denmark, Belgium, and Sweden in the 1980s and Canada in the 1990s. There has been vitriolic criticism of the possibility that expansionary austerity could ever exist. This dispute has sometimes distracted from what we see as the most policy-relevant result: the enormous difference, on average, between expenditure- and tax-based austerity plans. Our conclusions are very consistent with the findings of the literature on tax versus spending multipliers as reviewed by Ramey in this symposium.

We begin with a brief overview of some reasons why one might plausibly expect the effects of spending cuts on output, and eventually on the debt/GDP ratio, to differ from those of tax increases. We then turn to three key methodological issues that arise in measuring the effects of austerity empirically: 1) endogeneity, which in this case involves separating the effects on output of fiscal tightening from those of changes in output on the fiscal balance; 2) multiyear horizons, namely embracing the fact that austerity plans are almost always multiyear events involving a mixture of announcements of future changes in policy and immediate changes; and 3) the choice of the empirical model needed to design the experiment to measure the macroeconomic effects of austerity. In each case, we describe some common approaches in the earlier literature and how our own recent work draws upon them.

After having walked the reader through these three aspects of the austerity debate, we move to a more detailed discussion of our own findings. Together with our overall finding that expenditure-based austerity has on average smaller effects on output than tax-based austerity, we look for the channels that might be responsible for this result. We find that a main difference between expenditure- and tax-based austerity plans is the reaction of private capital investment. We also find

[1] Our data are available in a user-friendly form at http://www.igier.unibocconi.it/fiscalplans.

that the smaller negative effect of expenditure-based austerity plans looks much the same both before and after the Great Recession. In this discussion, we also review how the policy and the academic debates about austerity have evolved over time. In a concluding section, we offer some additional thoughts about how our framework of expenditure- and tax-based austerity plans relates to issues of redistribution, the electoral consequences of austerity, the case of Greece, and whether nations of Europe should have been slower after the Great Recession to seek out austerity.

Why Might Expenditure- and Tax-based Austerity Have Different Effects?

When analyzing austerity measures, it has been common for both policymakers and researchers to consider only the overall change in the fiscal balance, while paying much less attention to how that change is achieved. There are a number of reasons to suspect that the effects of expenditure-based and tax-based austerity may not be the same. The basic workhorse IS–LM model informs much of the public debate amongst politicians and the vast majority of the public. This model implies that spending cuts are (much) more recessionary than tax increases because spending multipliers in the model are (much) higher, in absolute value, than tax multipliers.

This assumption on the size of multipliers has been called into question in recent research, as Ramey (in this issue) points out. Why? Several arguments concerning both the demand side and the supply side are in order. First, with expenditure-based austerity, forward-looking households will react to the lower path of spending by realizing that future taxes will not rise as much as previously expected, or may even fall. Thus, the permanent expected income of consumers increases—supporting more private consumption. This, however, is not true for hand-to-mouth consumers who cut spending one-to-one when their disposable income falls and do not react to changes in permanent income. In the case of expenditure-based austerity, investors will also perceive their future tax burden reduced, or at least not increased as much as in the case of tax-based austerity. These effects will be stronger the more credible and long-lasting the expenditure cuts are perceived to be. Tax-based austerity, which does not tackle automatic increases in spending programs (like entitlements), will generate expectations of additional taxes in the future, thus having the opposite effects on consumers and investors.

Second, debt consolidation policies often occur in a state of crisis or close to it when investors (and consumers) are worried and uncertain about the future. Imagine an economy that finds itself on an unsustainable path with an exploding public debt (for instance, as described in Alesina and Drazen 1991; Blanchard 1990). The longer it waits before launching a fiscal stabilization, the bigger the future austerity package will need to be. When the stabilization eventually occurs, it removes the uncertainty about further delays, which would have increased its costs even more. Croce, Kung, Nguyen, and Schmid (2012) show that increases in government expenditure generate tax risks for firms: the extent of this uncertainty depends on the government's ability to pin down long-run tax dynamics. The removal of uncertainty is another force that boosts entrepreneurs' confidence and supports investment spending.

Third, demand-side effects may differ at different points of the cycle. For instance, the relative shares of the two types of consumers (forward-looking and hand-to-mouth) may vary over the business cycle, with hand-to-mouth consumers likely to be more numerous during recessions.

Fourth, spending cuts and tax increases have different supply-side effects. Tax distortions may affect the supply in a variety of ways. In the case of labor taxes, the elasticity of prime age males is low, but it is higher for the second earner in a family. Faced with higher labor taxes, youngsters may delay their entry into the labor market, weighing on their family income, and the elderly may retire sooner, putting additional burden on already stressed social security systems.

Fifth, for both these demand-side and supply-side effects, the consequences of expenditure- and tax-based plans vary with the persistence of the fiscal adjustment. Expenditure-based plans are less recessionary the longer-lived is the reduction in government spending, because the longer lasting the spending cuts, the larger the expected reduction in taxes for consumers and investors. On the other hand, the distortions associated with tax-based plans are larger the longer lasting is the increase in the tax burden.[2]

Austerity policies are rarely implemented in isolation: accompanying policies matter. Clearly monetary policy has a role: by lowering interest rates and buying government bonds, the central bank can help. This help is harder to come by at the zero lower bound, like in the latest round of austerity in Europe. The behavior of the exchange rate matters as well, especially for small open economies. Austerity policies that are more successful in reducing interest rates—for instance, by reducing risk premia on domestic bonds—may lead to a devaluation, which may help net exports. Austerity programs have often been accompanied by structural reforms, like labor or goods market liberalizations, which may affect the growth rate. The question is whether systematic differences in accompanying policies can explain the different output effects of expenditure-based versus tax-based austerity. We will show that the answer to this question is negative. This of course does not mean that accompanying policies are irrelevant, but simply that they do not explain the differences between the two types of austerity policies.

Measuring the Effects of Austerity: Three Issues

Because different theoretical models imply different multipliers—not only in size but sometimes even in sign—empirical evidence is critical in selecting among

[2] To the extent that fiscal adjustments are carried out in the form of multiyear plans, and thus perceived to be relatively permanent, a standard neo-Keynesian model implies that spending cuts are less recessionary than tax increases, as shown in Alesina, Barbiero, Favero, Giavazzi, and Paradisi (2017).

different theories. To this end, the empirical specification and the design of the identification strategy should be chosen independently from any specific theoretical mechanism behind differential effects of austerity. In this section, we discuss the design of empirical evidence on the macroeconomic consequences of fiscal adjustments. Such a design is difficult for three reasons: 1) endogeneity, (2) the multiyear nature of fiscal adjustments, and 3) that the problem of choosing the empirical model used to measure the effects of exogenous shifts in fiscal policy on output growth.

Endogeneity

The issue of endogeneity arises from the two-way interaction between fiscal policy and output growth. Suppose you observe a reduction in the government deficit together with an economic boom. It would be preposterous to jump to the conclusion that the policies that reduced deficits also generated growth. The causality is quite likely to run the other way: different factors (other than fiscal policy) increased economic growth, and by doing so led to higher tax revenues (for given tax rates), or reduced spending, say for unemployment compensation or welfare. This question has of course been at the very core of all empirical work on the effects of fiscal policy. Identification assumptions are thus needed to measure the effects of shifts in fiscal policy on output growth, and such assumptions should be as neutral as possible with respect to competing theories.

An early literature addressed this issue by considering episodes of large reductions in the cyclically adjusted budget deficit, arguing that this approach would, supposedly, mute the reverse effects of the business cycle on the government balance. (The cyclically adjusted budget deficit is the deficit a country would have if the economy ran at full potential.) For example, Giavazzi and Pagano (1990) analyzed three cases of fiscal consolidations, which occurred in the 1980s, two in Denmark and one in Ireland. They argued that reductions in the budget deficit signal that taxes may be lower in the future, with positive effects on consumers' permanent income and thus on consumption. Later, Alesina and Ardagna (1998) identified five additional episodes of large fiscal consolidation: Belgium 1984–86, Canada 1986–88, Italy 1989–92, Portugal 1984–86, and Sweden 1983–89. In each case, the cyclically adjusted primary deficit two years after the consolidation was at least 4 percentage points of GDP smaller than before the adjustment. These episodes were accompanied by growth of private consumption and investment in almost every year of the adjustment, sometimes with a year delay or so.

One finding of this early literature was that deficit reductions implemented via spending cuts were much less costly than those based upon tax increases, and that the former were sometimes associated with an expansion of GDP, even on impact. For example, McDermott and Wescott (1996) and Lambertini and Tavares (2005) show that a fiscal consolidation that concentrates on the expenditure side, and especially on transfers and government wages, is more likely to succeed in reducing the public debt ratio than tax-based consolidation. These results were robust to alternative measures of the cyclically adjusted budget deficit (for example, Alesina and

Perotti 1996). Using panel data for 19 OECD countries, Perotti (1999) estimates a model that predicts expansionary adjustments in "bad times" and contractionary adjustments in "good times." In bad times, when public debt is growing rapidly, a tax hike that rules out an even larger tax hike in the future can induce a positive response of consumption. In this vein, Alesina and Ardagna (2010) study a panel of OECD countries from 1970 to 2007. They define a fiscal adjustment as a year in which the primary cyclically adjusted budget balance improves by at least 1.5 percent of GDP. They find that these fiscal adjustments are expansionary only when they are spending-based. Moreover, they find that the fiscal adjustments associated with higher GDP growth are those in which a larger share of the reduction in the primary deficit is accounted for by cuts in current government spending, rather than in investment spending. They also find a positive reaction of private investment spending to government spending cuts.

One problem with this early literature is that cyclically adjusted measures of the deficit likely suffer from measurement error (Perotti 2013) in a way that means they still suffer from endogeneity. Even if the change in the fiscal balance is cyclically adjusted, so that it excludes changes induced by automatic stabilizers, it still includes other legislated changes in taxes and spending that are motivated by the state of the economy.

An alternative way of identifying exogenous fiscal adjustments—and one that has been increasingly adopted in recent years—is the "narrative approach" launched by Romer and Romer (2010). These authors have recovered exogenous shifts in taxes from a painstaking analysis of the motivations that US legislatures have offered for each of their tax decisions. This approach has been labeled "narrative identification." The motivation underlying each tax decision is assessed consulting original sources: budget documents, records of Congressional debates, speeches, and so on. The authors define as exogenous—that is, not related to the business cycle—all episodes of changes (up and down) in US federal taxes from 1947 to 2007 that were motivated by the aim of either improving long-run growth or reducing an inherited deficit. Economists in the research department of the IMF (Devries, Guarjardo, Leigh, and Pescatori 2011) used the narrative methodology to construct a panel of exogenous shifts in fiscal variables for 17 OECD countries over the sample 1978–2009. These data only cover episodes of deficit reduction, but include both changes in taxes and expenditures.

In Alesina, Favero, and Giavazzi (2019), we addressed endogeneity using the narrative approach.[3] We have documented close to 200 exogenous austerity plans—that is plans not adopted by considerations related to the business cycle—in 16 OECD

[3]Jordá and Taylor (2016) use narratively identified shifts in fiscal variables as an instrument for studying the effects of the changes in the cyclically adjusted fiscal balance. They overlook the difference between tax-based and expenditure-based austerity and concentrate on the issue of the recessionary versus expansionary effect of fiscal consolidations. These authors also question the validity of the narrative fiscal instrument used by Guajardo, Leigh, and Pescatori (2014) and propose an alternative econometric strategy. We discuss their criticism and their proposed strategy in the online Appendix available with this paper at the journal website.

economies (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Portugal, Spain, Sweden, the United Kingdom, and the United States) from the late 1970s until 2014. To construct this time series of exogenous shifts in fiscal variables, we took the Devries et al. (2011) dataset as a starting point and extended it in many important dimensions. First, we added the period 2010 to 2014, which is of course critical given the large amount of austerity plans that occurred in those years. Second, going back to the original sources, we complemented their data so as to keep track of the implementation of austerity plans over time—an issue whose importance we address in the next section. Third, we disaggregated these austerity plans depending on their composition. Our main disaggregation is between austerity plans mostly based on expenditure cuts and plans mostly based on tax hikes. But in addition, spending measures were further disaggregated between cuts in transfers and cuts in other government consumption (and investment). The measures on the tax side were broken down into indirect and direct taxes. While doing this, we double-checked the Devries et al. (2011) classifications and introduced some modifications.

Multiyear Austerity

Much of the literature on fiscal policy in general, and austerity in particular, has evaluated the effects of individual shifts in taxes or spending on a year-by-year (or even quarter-by-quarter) basis. Especially when this approach is applied to austerity policies, it overlooks two important facts. One is the multiyear nature of fiscal adjustments. Virtually all austerity programs are multiyear plans announced in advance and sometimes revised along the way. Because expectations matter for consumers' and investors' decisions, these announcements and the multiyear nature of these plans need to be taken into account. The other point is that decisions about how much to cut spending and how much to raise taxes are interconnected and cannot be assumed to be independent of one another. Typically, a legislature first decides by how much the deficit should be reduced (in the case of EU countries, this target has to be agreed upon with the European Commission). Given this target, its allocation between spending cuts and tax hikes is then decided through political bargaining in the legislature.

For example, the round of austerity which took place in Europe around 2010–2014 typically took the form of three-year plans of deficit reduction, announced by various countries in agreement with the EU or the "Troika" (the European Commission, the European Central Bank, and the International Monetary Fund). In some cases, these signposts were a precondition for receiving financing from the EU and the IMF, as happened in Greece, Portugal, and Ireland. In other cases, they were the conditions needed to avoid the "excessive deficit procedure," a status that implies automatic deficit reduction targets. In these agreements, the Troika did not care much about the composition of deficit reduction policies: they just cared about the bottom line in terms of multiyear deficit targets.

We used our narrative data to construct multiyear austerity plans. For each of the austerity plans, the total fiscal adjustment over time was divided into three categories: 1) measures announced and implemented immediately; 2) announcements of measures to be implemented in future periods; and 3) measures which had been legislated in the past but are implemented in the current year. (For simplicity of exposition here, we use only two time periods, one for the present and one for the future, but in our empirical work we consider three-year plans.) Of course, not all austerity plans need involve all three of these components. For each of these three categories, we estimate separately the increase in taxes and the cut in expenditures. Tax increases are measured by the expected revenue effect of each change in the tax code, either due to a change in tax rates or in the tax base, as a percent of GDP the year before the tax change is introduced. Ideally, one would want to distinguish between changes in the tax base and the tax rate because they may have different economic effects (Riera-Crichton, Vegh, and Vuletin 2016), but this was not feasible with our data. Spending cuts are changes in expenditure relative to the level that would have occurred absent the change in policy, as is standard. We then calculate, for each plan, which component dominates, whether spending cuts or tax increases. In the data, very few plans are close to being half and half, and our results are robust to dropping them. Note that considering tax hikes and expenditure cuts as independent would overlook the fact that they are linked by the decided target of deficit reduction. In any event, our results are robust (although more difficult to interpret) if tax hikes and spending cuts were considered as independent variables (for details, see Alesina, Favero, and Giavazzi 2019).

We call the first category of measures (those announced and immediately implemented) unexpected policy changes, where the total adjustment, or primary deficit reduction e, is given by the sum of tax increases τ and spending cuts g (that is, $e^u_{i,t} = \tau^u_{i,t} + g^u_{i,t}$). Of course, we recognize that even a measure announced and implemented immediately could have been anticipated based upon the legislative discussions that preceded its adoption ($e^a_{i,t,t-1} = \tau^a_{i,t,t-1} + g^a_{i,t,t-1}$). For the second category (announcements of policies to be adopted in the future), we make the admittedly restrictive assumption that these announcements are believed by economic agents, even though we of course also take account of changes when they happen. An important improvement in this line of research would be a better characterization of expectations of the public and different degrees of credibility of policy announcements. Finally, we consider policy announced at time t, to be implemented in the following years ($e^a_{i,t,t+1} = \tau^a_{i,t,t+1} + g^a_{i,t,t+1}$).

Consider a specific example: the fiscal consolidation in Belgium in 1992–94. The first column of Table 1 shows in 1992, 1993, and 1994 new immediate overall reductions in the primary deficit equivalent respectively to 1.85, 0.52, and 0.38 percent of GDP in the previous year are announced and carried out. The second column says that no previously announced austerity was carried out in 1992, while in 1993 and 1994 previously announced measures for, respectively, 0.47 and 0.83 of GDP, were carried out. The third column notes that further deficit reductions to be implemented in the following year are announced in 1992 and 1993, equivalent respectively to 0.47 and 0.83 percent of GDP. The next three columns show tax increases. For instance, for 1992, the tax increase carried out immediately, the

Table 1
The Multiyear Plan Introduced in Belgium in 1992 (% of GDP)

	Adjustment (e)			*Tax increase (τ)*			*Expenditure cut (g)*			
Year	e_t^u	$e_{t-1,t}^a$	$e_{t,t+1}^u$	τ_t^u	$\tau_{t-1,t}^a$	$\tau_{t,t+1}^a$	g_t^u	$g_{t-1,t}^a$	$g_{t,t+1}^a$	
1992	1.85	0	0.47	1.03	0	0.05	0.82	0	0.42	*EB*
1993	0.52	0.47	0.83	0.40	0.05	0.55	0.12	0.42	0.28	*TB*
1994	0.38	0.83	0	0	0.55	0	0.38	0.28	0	*EB*

Note: e_t^u is the unanticipated adjustment implemented in year t; $e_{t-1,\ t}^a$ is the anticipated adjustment, announced in the year $t - 1$ and implemented in year t; $e_{t,t+1}^a$ is the adjustment announced in the year t for implementation in the year $t + 1$; and analogously for the tax increases τ and expenditure cuts g. To decide whether something is expenditure-based (EB) or tax-based (TB), we sum over all g and τ measures and pick the category whose sum is larger.

zero tax increase that had been preannounced in an earlier year, and the small tax increase announced for the future. The final three columns show the spending cuts: the immediate spending cuts, the cuts that had been announced in an earlier year, and the spending cuts announced for the future. The adjustments in 1992 and 1994 are counted as expenditure-based (EB) because the sum over all measures on the expenditure side are larger $((.82 + 0 + .42 > 1.03 + 0 + .05)$ and $(.38 + .28 + 0 > 0 + .55 + 0))$, while the adjustment in 1993 is tax-based (TB) because $(.40 + .05 + .55 > .12 + .42 + .28)$.

The Model

To analyze the effects of austerity, one needs an empirical model to generate two paths for macroeconomic variables: in the presence and in the absence of the shift in fiscal variables.[4] The difference between these two paths is the impulse response that describes the dynamic reaction of the economy to the policy correction.

One example is the model used by Romer and Romer (2010) in their study: a truncated moving average representation of output growth in terms of (narratively identified) tax changes only. In practice, they estimate an ordinary least squares regression of output growth on (three-year lags) of exogenous changes in taxes. The implicit assumption is that narratively identified changes in taxes are orthogonal to all other structural shocks in the economy. The truncation at a three-year horizon is not a problem provided the variables excluded are correlated with the included narrative adjustments. Their regression also lumps together unexpected shifts in taxes and announcements, assuming that the responses of economic agents to the two kinds of policy shifts are identical. These assumptions have been relaxed in a

[4] In the online Appendix available at the journal website, we describe several alternative empirical models that can be used to simulate the macroeconomic effects of a plan. Along with moving averages and vector autoregressions, we discuss an alternative empirical strategy, the local projection method proposed by Jordá (2005), which implies computing impulse responses through the estimation of a battery of single equations, each of them capturing the effect of an exogenous shift in fiscal variables at a given horizon.

number of subsequent contributions. For example, Mertens and Ravn (2013) find that unexpected changes in taxes produce short-run effects on aggregate output that are larger than those associated with announcements. Favero and Giavazzi (2012) avoid the truncation problem by including narrative shocks in a vector autoregression, which includes government expenditure, government receipts, output growth, inflation, and the average interest cost of the public debt.

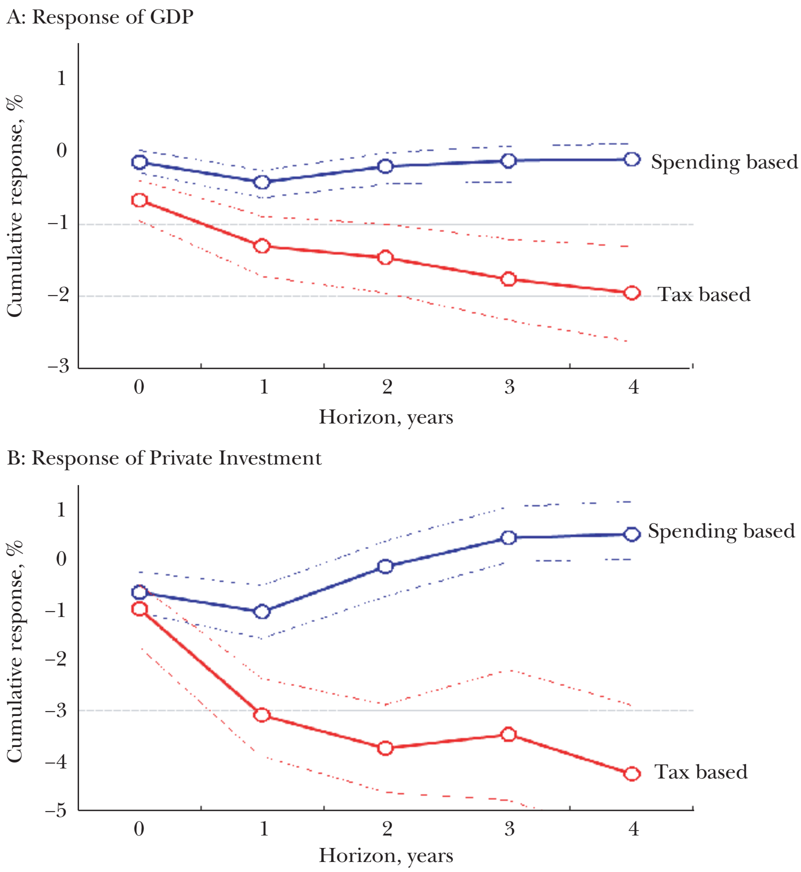
In our view, dynamic models such as vector autoregressions have several advantages. First, the estimated coefficients on the narratively identified shifts in fiscal variables measure the effect on output growth of the component of such shifts that is orthogonal to lagged included variables: thus, the estimated multipliers are not affected by the possible predictability of plans on the basis of past information. Second, by including in the vector autoregression changes in revenues and spending (as a fraction of GDP), one can track the effect of the narratively identified shifts in fiscal variables on total revenues and total spending. This allows a researcher to check the strength of narratively identified instruments: for instance verifying if, following a positive shift in taxes, revenues indeed increase. Finally, dynamic models allow a researcher to reconstruct the response of the debt/GDP ratio to a fiscal adjustment.

In our book, we detail how to insert narratively identified expenditure- and tax-based austerity plans into a fiscal vector autoregressive model (Alesina, Favero, and Giavazzi 2019). The parameters estimated in such a vector autoregression can then be used to generate two alternative paths for the macroeconomic and policy variables, in the presence or absence of the austerity plan. This vector autoregression can be linear or nonlinear. The nonlinearity allows for the dynamic response to a fiscal plan to differ depending on the regime the economy is in when the plan is introduced—for example, during an expansion or a recession, or with an increasing or stable debt/GDP ratio.[5]

Finally, how to measure the fiscal multipliers? In this symposium, Ramey discusses several alternatives used in the literature. The most common approach considers the total output response over time to a given fiscal adjustment (typically 1 percent of GDP), as in Romer and Romer (2010). We prefer the option of looking at the total output response over time divided by total change in fiscal variables over time, an approach suggested by Woodford (2011) and used by Mountford and Uhlig (2009), Uhlig (2010), Fisher and Peters (2010), and Auerbach and Gorodnichenko (2013). This approach has the advantage of taking into account the response of taxes and spending to the fiscal plan, as well as considering the persistence of fiscal shocks.

[5] In a multiyear plan, unexpected measures are typically accompanied by the announcement of future measures. This means that one cannot simulate the effect of an unexpected measure in isolation (unless it is unaccompanied by any other announcement). Doing so would assume that unexpected measures and announcements are uncorrelated—which they are not in our data. This problem can be addressed by exploiting the in-sample correlation between announcements and unexpected measures. More specifically, one can estimate parameters that relate announcements to unanticipated shifts in fiscal variables. Then, when simulating the effects of an unexpected measure, one can accompany it with an "artificial" announcement constructed using the value estimated in the sample.

Figure 1
The Response to Two Different Austerity Plans



Source: Based on the author's simulation of a panel vector autoregression approach for about 200 episodes of austerity across the 16 countries in our sample for the period 1978–2014.
Note: The Figure shows the effect on per capita GDP (panel A) and on private investment (panel B) of an expenditure-based austerity plan (blue line) and a tax-based austerity plan (red line). The continuous lines show the response to a plan which reduces the deficit-over-GDP ratio of 1 percent relative to the path that these variables would have followed in the absence of the fiscal plan. The dotted lines indicate the 90 percent confidence intervals.

Tax-based versus Expenditure-based Austerity: Results

Several graphs summarize our key results (for details, see Alesina, Favero, and Giavazzi 2019). Figure 1 shows the effect on per capita GDP (panel A) and on private investment (panel B) of an expenditure-based austerity plan (blue line) and a tax-based austerity plan (red line). The two continuous lines in the figures show the response of GDP (and private investment) to a plan that reduces the deficit-to-GDP ratio 1 percent relative to the path that these variables would have followed in the absence of the fiscal plan. The figures are based on the simulation of a panel vector autoregression approach for about 200 episodes of austerity across the 16 countries in our sample for the period 1978–2014, as mentioned above. The difference between the effects of expenditure- and tax-based plans is striking. As we can see from the dotted lines indicating the 90 percent confidence intervals, the two types of austerity plans are statistically different from one another. Tax-based austerity plans lead to deep and prolonged recessions, lasting several years. Expenditure-based plans on average exhaust their mild recessionary effect within two years after a plan is introduced. Of course, these findings represent averages of many plans. We provide details on specific episodes and detailed case studies in our book (Alesina, Favero, and Giavazzi 2019).[6]

The detailed data from our narrative approach also allows us to look at some broad categories of spending and taxes. For example, when we distinguish the effect of cuts in expenditure on goods, services, and investment from cuts in transfer payments (where transfers include both monetary transfers, such as social security, and in-kind transfers, such as health expenditures), we find that the results are broadly similar, although cuts in transfers imply even lower costs in terms of GDP growth than cuts in spending on goods and investment. This finding suggests that if one wishes to aggregate transfers with other items of the government budget, they ought to be aggregated with spending and not considered akin to negative taxes. In constructing our expenditure-based austerity plans, we would have liked to separate current government consumption from public investment but there are almost no austerity plans where the main component is a cut in public investment. Across the austerity plans we consider, when aggregating cuts in government consumption and investment, the former component represents around 80 percent of the total correction. The spending-based plans we study thus describe austerity programs mostly based on cuts in current government spending. The effects of cuts in public investment spending is obviously an important question for future research since they may have long-term costs that are not considered here.

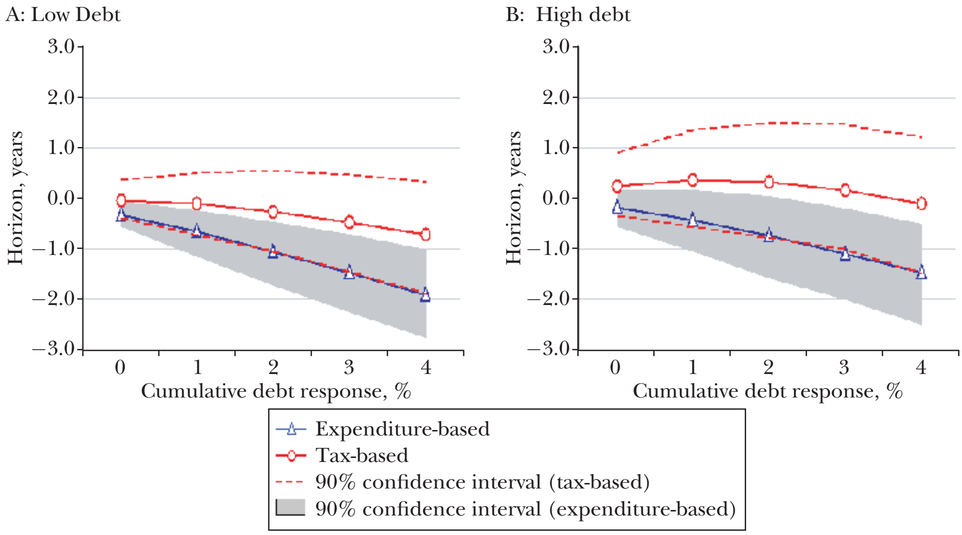
The component of aggregate demand that mostly drives the heterogeneity between the effects of tax- and expenditure-based austerity is private investment. Figure 1B reports the responses of private investments to fiscal plans and illustrates an even stronger heterogeneity than that observed for output growth. Private consumption instead behaves relatively similarly in the two cases of austerity. Net exports also do not behave differently during expenditure- and tax-based episodes. This fact makes it unlikely that movements in the exchange rate are an important factor in explaining the differences in the effects of expenditure- versus tax-based austerity.

Figure 2 shows the effects of tax- and expenditure-based austerity plans on the debt/GDP ratio. The effects vary depending on the initial level of debt and its cost. We consider two situations: a low level of debt (around 60 percent of GDP) and a high level of debt (around 120 percent of GDP) at the time the fiscal adjustment is implemented. In both cases, the cost of debt service is the same and assumed

[6] Estimating the truncated moving average representation of these variables, as done in Romer and Romer (2010), instead of simulating a vector autoregression, gives very similar results.

Figure 2

Public Debt Response to a Fiscal Consolidation Plan with Low and High Initial Debt-to-GDP Ratios (Assuming a Low Cost of Debt in Each Case)



Source: : Authors.
Note: Figure 2 shows the effects of tax- and expenditure-based austerity plans on the debt-to-GDP ratio. The effects vary depending on the initial level of debt and its cost. We consider two situations: a high debt (around 120 percent of GDP), and a low level of debt (around 60 percent of GDP), at the time the fiscal adjustment is implemented. In both cases, the cost of debt service is the same and assumed to be relatively low. Figure 2 is derived from a vector autoregression that includes taxes, government expenditure, net interest expenses on government debt, output growth, and inflation, along with the narratively identified austerity plans.

to be relatively low. Figure 2 is derived from a vector autoregression that includes taxes, government expenditure, net interest expenses on government debt, output growth and inflation, along with the narratively identified austerity plans. In the scenario with high public debt, an expenditure-based austerity plan (blue line) has a stabilizing effect on the debt dynamics, as debt goes down, while a tax-based plan (red line) has a destabilizing effect: that is, public debt tends to *increase* in the first years following a tax-based austerity plan. In the scenario with low public debt, the expenditure-based adjustment remains stabilizing, while the effect of a tax-based plan becomes neutral and not statistically different from zero.

Summing up: the anti-austerity argument—namely, that the latter creates large recessions and is self-defeating because it does not reduce the debt/GDP ratio—applies only to tax-based austerity, not to expenditure-based austerity. This distinction has been vastly overlooked. To our knowledge, only the government of Ireland in presenting its austerity plan in 2010 made an explicit reference to the academic literature emphasizing the different effects of tax hikes versus spending cuts. The European Commission's Ireland Stability Programme Update (December 2009, p.15) explains:

> In framing Budget 2010, the Government focused on curbing spending as expenditure needs to adjust to the revenue base which has been reduced as a result of the overall contraction of the economy and the loss of certain income streams. In addition, in formulating policy the Government took on board evidence from international organizations, such as the EU Commission, the OECD and the IMF, as well as the relevant economic literature which indicates that consolidation driven by cuts in expenditure is more successful in reducing deficits than consolidation based on tax increases. Past Irish experience also supports this view and suggests that confidence is more quickly restored when adjustment is achieved by cutting expenditure rather than by tax increases.

Although our work focused on 16 high-income OECD economies, many of our results may apply to other countries. Gunter, Riera-Crichton, Vegh, and Vuletin (2018) show that this is indeed the case for Latin American countries with a relatively large government sector. They find that fiscal adjustments carried out mainly through tax increases might be heavily recessionary in Latin American countries with high levels of taxation, such as Argentina and Uruguay (similarly to our OECD countries), but be fairly innocuous where initial taxation is low. They also find that the output costs of spending cuts are lower the more gradual is the fiscal adjustment.

In the remainder of this section, we address various questions which we suspect may have already occurred to the readers: whether the effects of expenditure- and tax-based austerity might differ in expansions versus recessions or at the zero lower bound, as well as whether the milder effects of expenditure-based, compared to tax-based austerity, might be more likely when austerity is implemented in combination with structural reforms or accommodative monetary policies. None of these possibly confounding factors alters our central finding that expenditure-based austerity plans are less costly to the economy and more effective in reducing the debt/GDP ratio.

Austerity during Expansions and Recessions

Government spending is likely to have larger expansionary effects in recessions than in expansions because, when the economy has slack, an increase in government spending is less likely to crowd out private demand. It might seem intuitively obvious that, conversely, a cut in government spending should have a larger effect in recessions than in expansions. But while this argument seems intuitive, testing it proves to be difficult. A central problem is that recessions and booms evolve dynamically. An economy in a recession may already have in place a dynamic recovery mechanism, or a growing economy may already be sputtering.

Auerbach and Gorodnichenko (2012, 2013) allow for the effects of shifts in fiscal policy to differ depending on whether they are introduced during an expansion or a recession, using a version of the model of taxes, government spending, and output by Blanchard and Perotti (2002) that allows for the estimated parameters to be different in expansions and recessions. They find very different tax and expenditure multipliers in recessions and in expansions. These authors, however, when simulating a shift in fiscal policy do not allow the economy to change state during recessions and booms: they assume that the state of the economy is constant for at least the 20 quarters over which multipliers are computed. Ramey and Zubairy (forthcoming) note that this is not a reasonable assumption for recessions, which in their sample have a mean duration of only 3.3 quarters. Ramey and Zubairy, instead, compute multipliers allowing the state of the economy to evolve during their simulation. Using quarterly US data covering wars and deep recessions (1889–2015), they find that government spending multipliers are less than one both in recessions and in booms.

Those papers consider both fiscal expansions and contractions. In our own work, we instead look only at periods of austerity (Alesina, Favero, and Giavazzi 2019). We find that, on average, expenditure-based adjustments have consistently much lower costs than tax-based ones, and that the costs of the former are close to zero regardless of the state of the economy. If austerity begins in a recession, it does look a bit more costly than if it starts in a boom, but the difference is small, and it does not affect the comparison between expenditure- and tax-based plans. When we use the methodology of Auerbach and Gorodnichenko (2012, 2013), we find larger negative effects of austerity during recessions, but the difference between expenditure- and tax-based episodes remains clear.

Whether expansionary fiscal policies and austerity measures have perfectly symmetrical effects with opposite signs remains an open question and a topic for future research. (For some evidence of an asymmetric effect of positive and negative fiscal measures, Barnichon and Matthes (2016) is a useful starting point.) However, there is a problem in that the narrative identification of exogenous expansionary episodes is difficult. Fiscal expansions typically occur during an economic downturn and are motivated by the state of the cycle, thus they are endogenous. Conversely, the narrative identification strategy that we adopt tends to exclude austerity plans beginning in a boom, because they could be confused with stabilization policies. Thus, our estimates of the cost of austerity should be considered an upper bound, because austerity starting in booms may be less costly. In any case, these considerations do not affect the comparisons between expenditure- and tax-based plans, with the former being more likely to be expansionary when started in a boom.

Austerity at the Zero Lower Bound

Do our results on the comparisons between expenditure- and tax-based plans also hold when the policy interest rate set by the central bank is at the zero lower bound? This question is difficult to answer because cases of austerity at the zero lower bound are essentially those that occurred in Europe in 2010–2014, plus a few episodes in Japan. There were many other factors at play in these episodes: the cases of European austerity started in the middle of very large recessions and occurred at the same time in many highly integrated economies, with some countries also facing major banking problems, like Ireland and Spain. Thus, it is hard to identify what caused what, given that so many factors were at play at the same time and the relatively few data points we have.

In order to shed some tentative light on this issue, we split our data into two subsamples: 1) euro area countries (Austria, Belgium, France, Finland, Germany, Ireland, Italy, Portugal, and Spain) from 1999, when the euro was introduced, onwards; and 2) non-euro-area countries (Australia, Denmark, UK, Japan, Sweden, United States, and Canada) together with euro area countries before 1999. We do this because, as in the case of the zero lower bound, the presence of a common currency prevents monetary policy from responding to fiscal developments in a specific country, while the presence of year fixed effects allows us to control for the fact that the European Central Bank might have responded to fiscal consolidations implemented in a large number of countries at the same time. Obviously, our test is imperfect, but the result is that we do not find a large difference between consolidations at or away from the zero lower bound.

European Austerity in 2010–2014

Did the recent episodes of austerity that occurred after the financial crisis—mostly in Europe in the aftermath of the euro crisis—differ from previous cases? Blanchard and Leigh (2014) answer "yes" to this question, considering the results of an ordinary least squares regression on a cross-section of 27 advanced economies. The dependent variable is the difference between actual cumulated real GDP growth (year-over-year) during 2010–11 (based on the latest available data) and the forecast prepared for the April 2010 IMF *World Economic Outlook*. The explanatory variable is the forecasted change, over the same period, of the general government cyclically adjusted fiscal balance measured in percent of potential GDP. They interpret the significant coefficient (- 1.09) on the regressor as evidence that fiscal multipliers generated by the fiscal adjustments in 2011 were higher than those predicted by forecasters.

In our opinion, these results should be interpreted extremely cautiously. To begin, we discussed above the limitations of the cyclically adjusted budget balance as a measure of fiscal stance. In addition, one-third of the fiscal adjustments considered in Blanchard and Leigh (2014) were fiscal expansions, not contractions, and so asymmetries between the effects of expansions and contractions could invalidate the result. Finally, as we show in Alesina, Favero, and Giovazzi (2019), the fiscal adjustments expected as of April 2010 were in fact correlated with the change in long-term interest rates: the estimated coefficient in the regression run by Blanchard and Leigh (2014) could thus simply measure the recessionary effect of the contemporaneous rise in the long-term interest rates.

In other words, suppose that harsher austerity was implemented in those European countries that were fiscally weaker and more exposed to a sudden increase in the cost of financing the debt because of the high level of debt and because of the "doom loop"—in which government debt becomes riskier, bank balance sheets become weaker, the government bails out the banks with additional debt, which in turn weakens bank balance sheets further, and so on. As a consequence of the Greek crisis, eventually the worst-case scenario materialized: not only the feared hike in interest rates, but also the amplification via the "doom loop" (Brunnermeier et al.

2016). The contractionary effect of the increase of long-term rates was amplified by an unusual contraction in lending and caused weakening of banks' balance sheets, which were heavily invested in government bonds. In this case, the stronger recessionary impact of austerity on growth could simply measure the unusual strength of the effect of the hike in long-term rates caused by this "doom loop." If the doom loop channel is not included in the model used to produce the forecasts, then the forecast error for output is correlated with the shock to interest rates, for which the fiscal correction works as an instrument.

If Blanchard and Leigh (2014) are taken at face-value, they would suggest that the multipliers assumed by the IMF models, as by the models used by other international organizations, were "too small," around 0.5. However, at least in the case of tax-based plans, those multipliers are much smaller than the multipliers revealed by our plan-based reduced-form empirical evidence, which hover around 2.

When we investigate in detail these European episodes of austerity, we conclude that one cannot reject the hypothesis that their effects on output were not statistically different from previous cases (Alesina, Favero, and Giavazzi 2019). The analysis of these episodes also confirms that countries that chose tax-based austerity suffered deeper recessions compared to those that decided to adopt expenditure-based plans. The very large size of recessions in some countries (Greece, Spain, Portugal, Italy) are consistent with the large "tax multipliers" which we found for previous periods, given the size and composition of some plans adopted in 2010–14. Most of these plans included large tax increases. The two countries that adopted expenditure cuts almost exclusively (Ireland) or adopted mostly expenditure cuts (the United Kingdom) had much smaller and shorter recessions, and, in the case of Ireland, this was despite a massive banking problem. The United Kingdom, which had kept the pound rather than switching to the euro, was also helped by an exchange rate devaluation.

Accompanying Policies

If expenditure-based plans were systematically accompanied by more accommodative policies, then the difference between expenditure- and tax-based plans would result from these other policies and have nothing to do with different fiscal multipliers. Guajardo et al. (2014) suggest that the stance of monetary policy may explain the difference between expenditure- and tax-based measures. In Alesina, Favero, and Giavazzi (2015), we show that only a very small fraction of the different effect on output of expenditure- and tax-based adjustments can be ascribed to monetary policy. We come to this conclusion by augmenting the model used throughout with a monetary policy channel.[7] When this channel is closed, in a counterfactual that prevents monetary policy from reacting to fiscal adjustments, an important

[7] This channel produces estimates of the impact of monetary policy on output which lie in between the typical response obtained on US data (see for example, Christiano, Eichenbaum, and Evans (1999) and that obtained on euro area data, which is smaller than that observed for the United States (for example, see Peersman and Smets 2001).

heterogeneity between expenditure- and tax-based austerity plans is still observed. Finally, note that the response of monetary policy may be endogenous in the sense that the central bank may on purpose react differently to more or less credible austerity plans, and credibility of a fiscal plan may be related to its composition.

The same line of analysis applies to the behavior of the exchange rate. If an expenditure-based austerity plan reduces interest rates and leads to a devaluation, it may in turn sustain output growth. However, a devaluation occurring before the beginning of an expenditure-based austerity plan (ignoring the possible significant effects of expectations of a plan on exchange rates) may lead to a spurious attribution of low output cost to the plan, since the benefits of the devaluation are overlooked. In Alesina, Favero, and Giavazzi (2019), we find no systematic difference in the behavior of the exchange rate before expenditure- or tax-based austerity plans. We also exclude, in a variety of different ways, plans that were preceded by significant (or even small) devaluations, and show that our results are robust. Also, if the main explanation for the difference between the output effects of expenditure- and tax-based plans were the exchange rate, then the component of aggregate demand that should reflect it would be net exports. This is not the case.

Periods of austerity are sometimes accompanied by structural reforms, which may include product and/or labor market liberalizations. The latter may stimulate growth and, if they were systematically occurring along with expenditure-based austerity plans, that could explain our findings. But the answer to this conjecture is also "no." In fact, structural reforms do not occur systematically during periods of spending cuts according to our findings. Although, Perotti (2013), Alesina, Perotti, Tavares, Obstfeld, and Eichengreen (1998), and Alesina and Ardagna (2013) show that amongst all fiscal adjustments, the least costly were those accompanied by supply-side reforms and by wage moderation, this is not inconsistent with our results. Our robustness check is different: we checked whether the choice of expenditure- or tax-based austerity plans can be explained by supply-side reforms, and we found that it cannot.

Further Thoughts

In this final section, we briefly discuss a few additional issues regarding the trade-offs between expenditure-based and tax-based austerity. First, redistributive effects may differ between austerity achieved through tax hikes or spending cuts. There is a common, though untested, perception that spending cuts may raise inequality more than tax increases, but clearly this conclusion depends on which taxes are raised and which spending items are cut. In the context of high-income OECD countries, total government spending is close to 50 percent of GDP. It seems quite plausible then that budget cuts of the magnitudes needed to reduce a deficit by, say, 3–4 percent of GDP can be achieved without affecting the welfare of the really poor. In fact, much of the modern welfare state supports the middle class and in some cases even the upper-middle class, which often enjoys almost free health care regardless of income levels, heavily subsidized university education, and (especially in continental Europe) subsidized services like transportation. In addition, issues of redistribution arise not just at a point in time, but also across generations. For example, an increase in the mandatory retirement age may lead to a more equitable distribution of resources across generations. Cuts in current public investment, rather than current transfers, also have important redistributive consequences across generations. Passing a large debt burden to future generations will have consequences for intergenerational redistribution. Of course, the question of how different spending or tax changes might affect redistribution is separate from an analysis of whether austerity plans adopted in this or that country had these goals in mind.

Second, what are the electoral consequences of austerity? A common view amongst commentators is that deficit reduction polices are the kiss of death for the governments which implement them. However, the electoral effects of austerity are not clear-cut or easy to predict (Alesina, Favero, and Giavazzi 2019). Several governments (and not only in Germany) have extended their time in office during periods of austerity.

Third, although a great deal has been written about the experience of Greece during 2010–2014, and the many errors, confusion, messy choices made during that time, we have not discussed it specifically here. In our book we discuss Greece in more detail (Alesina, Favero, and Giavazzi 2019); also see the excellent work by Gourinchas, Philippon, and Vayanos (2017) and Ardagna and Caselli (2014). As the Greek situation unfolded, the "Troika" (the European Commission, the European Central Bank, and the International Monetary Fund) paid very little attention to the composition of austerity plans, whether in Greece or anywhere else, and demanded an extraordinarily heavy dose of both tax increases and spending cuts, which were then implemented in a very unclear and hard-to-measure way. Using the (admittedly rough) data available for Greece, we used our model of fiscal adjustments, developed in Alesina, Favero, and Giavazzi (2019), to simulate the effects of the Greek austerity plans. This exercise predicts the Greek recession quite well. It is baffling that the Troika seemed surprised by the size of the Greek recession. Indeed, the "surprise" of the Troika seems to be a hypocritical attempt at deflecting responsibility.

Fourth, was 2010 too soon to start austerity plans in some European countries? It is obviously impossible to know what would have happened if countries across Europe had continued to expand their borrowing beyond 2010 and for several years afterwards. We suspect that the rosy scenario painted by the anti-austerity side is too optimistic. However, our analysis suggests that the effects of austerity would have been lighter if it had been focused mostly on the spending side: Ireland and (in part) the United Kingdom, did exactly that and had much smaller and shorter recessions than Italy, Portugal, and Spain, where a large portion of austerity was on the tax side. Spending-based austerity plans that were less front-loaded, but credible, would probably have worked better, leading to smaller recessions and debt stabilization. An earlier intervention by the European Central Bank would have been a welcome help, too.

Finally, one may wonder: if spending cuts are so much less contractionary than tax increases, why didn't policymakers incorporate this knowledge into their decisions? They typically did not, with the exception of the Irish government in 2010 as we showed above. One possibility is that when the IMF and others argued for austerity, the advice failed to distinguish between the expenditure- and tax-based policies, implicitly sending the message that it did not much matter how deficits were reduced. In addition, policymakers may find it harder to cut spending than to raise taxes. Tax hikes are faster to implement and bring revenue more rapidly than cuts in government spending programs, and so policymakers might adopt them even if they suspect that they may be more recessionary. Moreover, spending cuts often affect specific groups, like retirees, students, and public sector unions, who are often organized and able to oppose spending cuts with strikes, protests, cuts of campaign contributions, and other political activities which go above and beyond voting. By comparison, taxpayers as a group are less politically organized. This is a vivid example of a situation where the concentrated costs of blocking specific spending cuts may loom larger to politicians than generalized costs of higher taxes.

■ *We thank the board of editors of the Journal of Economic Perspectives for guidance and comments. Pierfrancesco Mei provided excellent research assistantship.*

References

Alesina, Alberto, and Silvia Ardagna. 1998. "Tales of Fiscal Adjustment." Scholarly Articles 2579822, Harvard University Department of Economics.

Alesina, Alberto, and Silvia Ardagna. 2010. "Large Changes in Fiscal Policy: Taxes versus Spending." Chap. 2 in *Tax Policy and the Economy*, vol. 24. University of Chicago Press.

Alesina, Alberto, and Silvia Ardagna. 2013. "The Design of Fiscal Adjustments." *Tax Policy and the Economy* vol. 27, pp. 19–68. University of Chicago Press.

Alesina, Alberto, Omar Barbiero, Carlo Favero, Francesco Giavazzi, and Matteo Paradisi. 2017. "The Effects of Fiscal Consolidations: Theory and Evidence." NBER Working Paper 23385, May.

Alesina, Alberto, and Allan Drazen. 1991. "Why Are Stabilizations Delayed?" *American Economic Review* 81(5): 1170–88.

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2015. "The Output Effect of Fiscal Consolidation Plans." *Journal of International Economics* 96(S1): S19–S42.

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019. *Austerity: When It Works and When It Doesn't.* Princeton University Press.

Alesina, Alberto, and Andrea Passalacqua. 2016. "The Political Economy of Government Debt." Chap. 33 in *Handbook of Macroeconomics*, vol. 2, edited by John Taylor and Harold Uhlig, 2599–2651. Elsevier.

Alesina, Alberto, and Roberto Perotti. 1996. "Fiscal Adjustments in OECD Countries: Composition and Macroeconomic Effects." NBER Working Paper 5730, August.

Alesina, Alberto, Roberto Perotti, José Tavares, Maurice Obstfeld, and Barry Eichengreen. 1998. "The Political Economy of Fiscal Adjustments."

Brookings Papers on Economic Activity no. 1, pp. 197–266.

Ardagna, Silvia, and Francesco Caselli. 2014. "The Political Economy of the Greek Debt Crisis: A Tale of Two Bailouts." *American Economic Journal: Macroeconomics* 6(4): 291–323.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2013. "Fiscal Multipliers in Recession and Expansion." Chap. 2 in *Fiscal Policy after the Financial Crisis,* edited by Alberto Alesina and Francesco Giavazzi. University of Chicago Press.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2016. "Effects of Fiscal Shocks in a Globalized World." *IMF Economic Review* 64(1): 177–215.

Barnichon, Régis, and Christian Matthes. 2016. "Understanding the Size of the Government Spending Multiplier: It's in the Sign." CEPR Discussion Papers 11373, July.

Blanchard, Olivier J. 1990. "Comments on Giavazzi and Pagano." In *NBER Macroeconomics Annual,* 5: 110–117.

Blanchard, Olivier J., and Daniel Leigh. 2014. "Learning about Fiscal Multipliers from Growth Forecast Errors." *IMF Economic Review* 62(2): 179–212.

Blanchard, Olivier J., and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117(4): 1329–68.

Brunnermeier, Marcus K., Luis Garicano, Philip R. Lane, Marci Pagano, Ricardo Reis, Tano Santos, David Thesmar, Stijn Van Nieuwerburgh, and Dimitri Vayanos. 2016. "The Sovereign-Bank Diabolic Loop and ESBies." *American Economic Review* 106(5): 508–12.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary Policy Shocks: What Have We Learned and To What End?" Chap. 2 in *Handbook of Macroeconomics* vol. 1, edited by J. B. Taylor and M. Woodford. Elsevier.

Croce, M. Max, Howard Kung, Thien T. Nguyen, and Lukas Schmid. 2012. "Fiscal Policies and Asset Prices." *Review of Financial Studies* 25(9): 2635–72.

Devries, Pete, Jamie Guajardo, Daniel Leigh, and Andrea Pescatori. 2011. "A New Action-based Dataset of Fiscal Consolidation." IMF Working Papers WP/11/128, June.

European Commission. 2009. "Ireland—Stability Programme Update: December 2009." http://ec.europa.eu/economy_finance/economic_governance/sgp/pdf/20_scps/2009-10/01_programme/ie_2009-12-09_sp_en.pdf.

Favero, Carlo, and Francesco Giavazzi. 2012. "Measuring Tax Multipliers: The Narrative Method in Fiscal VARs." *American Economic Journal: Economic Policy* 4(2): 69–94.

Fisher, Jonas D. M., and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *Economic Journal* 120(544): 414–36.

Giavazzi, Francesco, and Marco Pagano. 1990. "Can Severe Fiscal Contractions Be Expansionary? Tales of Two Small European Countries." In *NBER Macroeconomics Annual* 5: 75–122.

Gourinchas, Pierre-Olivier, Thomas Philippon, and Dimitri Vayanos. 2017. "The Analytics of the Greek Crisis." *NBER Macroeconomics Annual* 31: 1–81.

Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. "Expansionary Austerity? International Evidence." *Journal of the European Economic Association* 12(4): 949–68.

Gunter, Samara, Daniel Riera-Crichton, Carlos Vegh, and Guillermo Vuletin. 2018. "Non-Linear Effects of Tax Changes on Output: The Role of the Initial Level of Taxation." World Bank Policy Research Working Paper, WPS8668.

International Monetary Fund (IMF). 2010. *World Economic Outlook: April 2010: Rebalancing Growth.* World Economic and Financial Surveys. International Monetary Fund.

Jordá, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95(1): 161–82.

Jordá, Òscar, and Alan M. Taylor. 2016. "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy." *Economic Journal* 126(590): 219–55.

Lambertini, Luisa, and Jose A. Tavares. 2005. "Exchange Rates and Fiscal Adjustments: Evidence from the OECD and Implications for the EMU." *Contributions in Macroeconomics* 5(1): 1–28.

McDermott, C. John, and Robert F. Wescott. 1996. "An Empirical Analysis of Fiscal Adjustments." *International Monetary Fund Staff Papers* 43(4): 725 –54.

Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103(4): 1212–47.

Mountford, Andrew, and Harald Uhlig. 2009. "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics* 24(6): 960–92.

Peersman, Gert, and Frank Smets. 2001. "The Monetary Transmission Mechanism in the Euro Area: More Evidence from VAR Analysis." Available at SSRN: https://ssrn.com/abstract=356269.

Perotti, Roberto. 1999. "Fiscal Policy in Good Times and Bad." *Quarterly Journal of Economics* 114(4): 1399–1436.

Perotti, Roberto. 2013. "The 'Austerity Myth': Gain without Pain?" Chap. 8 in *Fiscal Policy after* *the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi. Elsevier.

Ramey, Valerie A., and Sarah Zubairy. Forthcoming. "Government Spending Multipliers in Good Times and in Bad: Evidence From U.S. Historical Data." *Journal of Political Economy*

Riera-Crichton, Daniel, Carlos A. Vegh, and Guillermo Vuletin. 2016. "Tax Multipliers: Pitfalls in Measurement and Identification." *Journal of Monetary Economics* 79: 30–48.

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100(3): 763–801.

Uhlig, Harald. 2010. "Some Fiscal Calculus." *American Economic Review* 100(2): 30–34.

Woodford, Michael. 2011. "Simple Analytics of the Government Expenditure Multiplier." *American Economic Journal: Macroeconomics* 3(1): 1–35.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 163–190

The Declining Labor Market Prospects of Less-Educated Men

Ariel J. Binder and John Bound

During the last 50 years, labor market outcomes for men without a college education in the United States worsened considerably. Between 1973 and 2015, real hourly earnings for the typical 25–54 year-old man with only a high school degree declined by 18.2 percent,[1] while real hourly earnings for college-educated men increased substantially. Over the same period, labor-force participation by men without a college education plummeted. In the late 1960s, nearly all 25–54 year-old men with only a high school degree participated in the labor force; by 2015, such men participated at a rate of 85.3 percent.

In this article, we examine secular change in the US labor market since the 1960s. We have two distinct but related objectives. First, we assemble an overview of developments in the wage structure, focusing on the dramatic rise in the college wage premium. Second, we examine possible explanations for the decline in labor-force participation among less-educated men. One hypothesis has been that declining labor market activity is connected with declining wages in this population. While such a connection indicates a reduction in labor demand, we point out that the canonical neoclassical framework, which emphasizes a labor demand curve shifting inward across a stable labor supply curve, does not reasonably account for

[1] Hourly earnings declined from $21.40 to $17.50. Throughout this article, we adjust reported wages to 2017 prices using the Personal Consumption Expenditure deflator.

■ *Ariel Binder is a PhD candidate in Economics and Pre-Doctoral Trainee at the Population Studies Center, University of Michigan, Ann Arbor, Michigan. John Bound is the George E. Johnson Collegiate Professor of Economics, a Research Professor at the Population Studies Center, University of Michigan, Ann Arbor, Michigan, and a Research Associate of the National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are ajbinder@umich.edu and jbound@umich.edu, respectively.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.163 doi=10.1257/jep.33.2.163

this development. This is because wages have not declined consistently over the sample period, while labor-force participation has. Moreover, the uncompensated elasticity of labor supply necessary to align wage changes with participation changes, during periods when both were declining, is implausibly large.

We then examine two oft-discussed developments outside of the labor market: rising access to Social Security Disability Insurance (DI), and the growing share of less-educated men with a prison record. Rising DI program participation can account for a nontrivial share of declining labor-force participation among men aged 45–54, but appears largely irrelevant to declining participation in the 25–44 year-old group. Additionally, we document that most nonparticipating men support themselves primarily on the income of other family members, with a distinct minority depending primarily on their own disability benefits. The literature has not progressed far enough to admit a reasonable quantification of the impact of rising exposure to prison on the labor-force participation rate, but recent estimates suggest that sizable effects are possible. We flag this as an important area for further research.

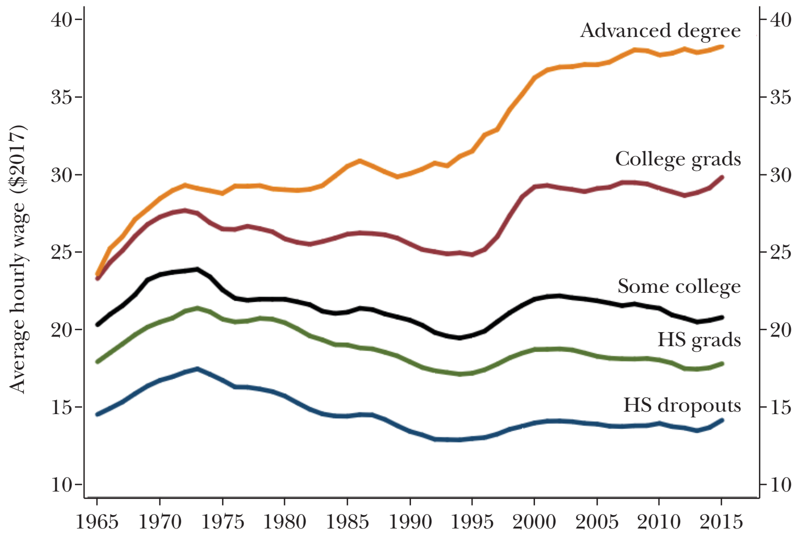
The existing literature, in our view, has not satisfactorily explained the decline in less-educated male labor-force participation. This leads us to develop a new explanation. As others have documented, family structure in the United States has changed dramatically since the 1960s, featuring a tremendous decline in the share of less-educated men forming and maintaining stable marriages. We additionally show an increase in the share of less-educated men living with their parents or other relatives. Providing for a new family plausibly provides a man with incentives to engage in labor market activity: conversely, a reduction in the prospects of forming and maintaining a stable family removes an important labor supply incentive. At the same time, the possibility of drawing income support from existing relatives creates a feasible labor-force exit. We suspect that changing family structure shifts male labor supply incentives independently of labor market conditions, and that, in addition, changing family structure may moderate the effect of a male labor demand shock on labor-force participation. Because male earning potential is an important determinant of new marriage formation, a persistent labor demand shock that reduces male earning potential could impact male labor-force participation through its effects on the marriage market.

Much prior research has addressed US labor market trends over the last half century, including several recent reviews of male employment (Moffitt 2012; Council of Economic Advisors 2016; Abraham and Kearney 2018). Our aim is not to review the literature, but rather to point out where we think consensus has developed and where we think important questions remain unanswered. In the synthesis that emerges, the phenomenon of declining prime-age male labor-force participation is not coherently explained by a series of causal factors acting separately. A more reasonable interpretation, we argue, involves complex feedbacks between labor demand, family structure, and other factors that have disproportionately affected less-educated men.

The US Wage Structure since 1970

Prime-age men of all education levels experienced robust wage growth in the mid-1960s and early 1970s, as shown in Figure 1, which plots trends in (geometric)

Figure 1
Real Hourly Earnings by Education Status, Men Aged 25–54, 1965–2016



Source: Authors' calculations based on the March Supplement to the Current Population Survey.
Note: This figure plots trends in geometric average hourly earnings by education group. Annual wage and salary income is adjusted for top-coding and converted to 2017 dollars using the Personal Consumption Expenditures deflator. Hourly earnings are computed by dividing annual earnings by total hours worked (the product of weeks worked and usual hours worked per week). Before the 1976 survey, weeks worked and usual hours worked per week are imputed using demographic information in conjunction with information on bracketed weeks worked and hours worked last week. Within each education status, a reweighting procedure is employed to hold the age distribution constant across each year. See the online Appendix for further details. Geometric average hourly wages are computed as the exponential of the average of log hourly wages. The graph presents three-year centered moving averages. "HS" is high school.

average hourly earnings by education group.[2] This growth ceased after 1973, with hourly earnings for all but those with advanced degrees falling for the next 20 years. After rising again for a decade starting around 1994, hourly earnings stagnated for the next decade (again, apart from those with advanced degrees). Over the last few years, hourly earnings have begun to grow again for men of all education levels. It is too early to tell whether this growth is a blip in the long-run trend or the beginning of a sustained increase. Nonetheless, for groups without a college degree, real hourly earnings were substantially lower in 2015 than they were in 1973.[3]

[2] We use the March supplement to the Current Population Survey throughout this paper unless otherwise specified. Following the standard definition, we consider prime-age men to be between the ages of 25 and 54. Within each education group, we implement a reweighting method to hold constant the age distribution across time. We compute geometric averages by applying the exponential function to the average of log hourly earnings. An online Appendix, available with this paper at the journal website includes details regarding data processing as well as additional figures and tables.
[3] The hourly earnings numbers we report represent geometric average hourly earnings for those working in the reference year. Conceptually, we might prefer the typical wage that a man might expect (including those not working). Using techniques developed by Juhn, Murphy, and Topel (1991) we imputed wages

Substantial changes in wage dispersion are also apparent. During the 1970s, wages of college graduates fell relative to those of high school graduates; after 1980, the college wage premium increased dramatically. For example, in 1980, the average college graduate earned 1.26 times as much per hour as the average high school graduate. By 2015, this differential had widened to 1.68. The wage differential for advanced degree holders relative to high school graduates grew more sharply, from 1.41 to 2.17. Note that we are using the *geometric* average for hourly wages, which is less sensitive than the arithmetic average to outlier earners: if hourly earnings are log-normally distributed, the geometric average is the median of the distribution. Moreover, we have removed outlier earners from our calculations (see online Appendix for details). Even with these adjustments to the data, we still measure a large secular increase in male wage dispersion.

A breakdown of the full sample into different demographic groups reveals similar patterns (illustrated in online Appendix Figure A2). For example, whites and blacks of all ages and education levels experienced strong wage growth in the mid-1960s and early 1970s. Thereafter, whites and blacks of all ages experienced substantial wage decline—especially the high school dropout populations. This was followed by robust growth for all demographic groups from 1994–2002 and then modest decline throughout the 21st century. Since the early 1990s, cumulative wage growth appears to be higher for high school dropouts than for high school graduates or those with some college education. One modest exception to these overall patterns is that secular wage fluctuations appear relatively muted for the 45–54 age group. Card and Lemieux (2001) have shown how these patterns are consistent with cohort-specific changes in relative supply, though an alternative explanation is that older workers are more insulated from labor market change than younger workers.[4]

When analyzing secular change in labor market outcomes by education group, an important question is whether the underlying skill composition of education groups has remained constant over time. This may be especially relevant for the less-educated groups of interest. For example, in the late 1960s, nearly 40 percent of prime-age men had not completed high school; by the 2010s, this share had plummeted to around 10 percent. The share of prime-age men with only a high school degree has remained relatively constant over time as high school completion rates, but also college participation rates, have risen. It is plausible that at least a portion of observed secular declines in wages and employment among less-educated men stem from this population becoming increasingly negatively selected on labor market skills.

To assess this possibility, we compiled data from the National Longitudinal Surveys of Youth (NLSY) containing well-validated measures of cognitive and noncognitive skills. Results and further discussion comparing the 1959–65 birth cohorts (drawn from the NLSY79) to the 1980–84 birth cohorts (drawn from the NLSY97)

for those who did not work throughout the entire reference year. Including these imputations in our calculation of wages mildly affects the trends, but does not substantially alter the broad patterns across time. See online Appendix Figure A1 available with this paper at the journal website.

[4]While the focus of this paper is on men, it is worth mentioning that wages for women have followed similar trends (Autor 2014), though with more overall wage growth since the late 1970s, commensurate with the narrowing of the gender wage gap (Blau and Kahn 2017).

appear in the online Appendix. Perhaps surprisingly, we document little evidence that average skill levels in the high-school-dropout and high-school-graduate populations declined between the two cohorts. Our observations are relatively consistent with the work of Altonji, Bharadwaj, and Lange (2012), documenting population improvements in various skill levels between the NLSY79 and NLSY97 cohorts.

Explaining Wage Dispersion

Labor economists have drawn on the neoclassical supply-demand framework to interpret secular changes in wage dispersion. For example, the relative size of the college-educated workforce grew rapidly during the 1970s, inducing the relative wages of college graduates to fall (Freeman 1975). Since then, the relative supply of college-educated workers grew at a slower rate and their relative wages rose dramatically. These developments are consistent with an outward shift in relative demand for college-educated workers (Katz and Murphy 1992; Murphy and Welch 1992; Bound and Johnson 1992; for a longer-term perspective, Goldin and Katz 2008). Autor, Katz, and Kearney (2008) attributed secular growth in the college wage premium over the period from the 1970s to the early 2000s to an outward shift in relative demand, which proceeded at a constant pace until the early 1990s and then slowed somewhat.

Since 1990, a voluminous literature has analyzed potential causes of shifts in relative demand. Some initially observed that rising trade with countries with abundant supplies of less-skilled workers should put downward pressure on the relative wages of less-skilled US workers (for example, Murphy and Welch 1991). Economists during the 1990s produced a range of empirical estimates of this effect, generally finding that the magnitudes were not nearly large enough to explain the entire observed decline in relative wages. According to the 1997 survey of William Cline, "a reasonable estimate … would be that the international influences [have] contributed about 20 percent of the rising wage inequality in the U.S." Even this modest estimate was high relative to other contemporary surveys (for example, the 1995 "Symposium on Inequality and Trade" issue of this journal).

Over the last 25 years, the North American Free Trade Agreement was implemented, China joined the World Trade Organization, and the volume of trade between high and middle-income countries increased dramatically. Nonetheless, more recent studies reached the same basic conclusion, with Katz (2008) suggesting that rising trade accounted for less than 20 percent of the increase in the college wage premium between 1980 and 2006 (see also Krugman 2008; Bivens 2007). More recently, Autor, Dorn, and Hanson (2013) leveraged commuting-zone-level variation to estimate that rising import competition from China could explain roughly 25 percent of the decline in US manufacturing employment between 1990 and 2007. Their estimates indicate large effects of trade on highly exposed communities, but are consistent with modest overall effects of trade on relative wages.

What explains the other 80 percent of the growth in dispersion in the US wage structure? During the 1990s, a consensus arose that skill-biased technological change was a primary driver. Bound and Johnson (1992) and Katz and Murphy (1992) showed that labor reallocation across sectors could account for relatively little of the shift in the utilization of skilled labor during the 1980s. Most of the change occurred within narrowly defined sectors, suggesting a broad-based shift in demand for skilled labor unrelated to trade forces.[5] Berman, Bound, and Griliches (1994) and Autor, Katz, and Krueger (1998) reported similar findings, while also showing that skill upgrading was most rapid within industries that invested most in computer technologies.

The "skill-based" framework succeeds in explaining the dramatic growth in wage dispersion of the 1980s. However, since the 1990s, wages of high school dropouts relative to high school graduates and those with some college have not continued to decline—if anything they have increased. Accordingly, Autor, Levy, and Murnane (2002, 2003) envisioned a framework in which the production process involves several types of tasks: for example, manual, routine, and abstract. Computers substitute for labor in routine tasks but complement labor in abstract tasks, resulting in increased relative demand for labor with high cognitive skill. Autor, Katz, and Kearney (2006, 2008) have also hypothesized that computers do not impact low-skill manual tasks: thus, this "task-based" framework can account for the post-1990s polarization of relative wage growth.[6] The task-based framework has become enormously influential in terms of understanding which jobs are vulnerable to displacement (Acemoglu and Autor 2011). Research following Autor, Levy, and Murnane (2002, 2003) has found that routine tasks are likeliest to be moved offshore or automated (for example, Frey and Osborne 2017) and, in general, that skill-biased technological change is fundamentally altering the nature of work (Levy and Murnane 2004; Acemoglu and Restrepo 2018).

Other factors may also have contributed to the growth in wage dispersion since 1980. Some have focused on institutional factors. For example, DiNardo, Fortin, and Lemieux (1996) estimated that the decline in unionization of workers contributed close to 20 percent of the rise in the college wage premium over the 1980s, although these estimates ignore general equilibrium effects of de-unionization. While such effects in theory could cut in either direction, recent work has found positive spillover effects of unions on the wages of comparable nonunion workers (Fortin, Lemieux, and Lloyd 2018). DiNardo, Fortin, and Lemieux also estimated that the falling real minimum wage modestly increased the college wage premium for young men but had negligible effects for more experienced men. Subsequent research has confirmed small effects of changes in the minimum wage on the wages of low-wage workers (Lee 1999; Autor, Manning, and Smith 2016). Other work has discussed the impact of the deregulation of various industries (Hirsch 1988; Rose 1987; in this journal, Fortin and Lemieux 1997).

[5] Feenstra and Hanson (1996, 1999) argued that the outsourcing of intermediate products implied that the framework used by the above authors would underestimate the role played by trade. While this point is well taken, even after accounting for outsourcing, Feenstra and Hanson's estimates suggested that skill-biased technological change was substantially more important than trade in explaining the rise in the relative demand for college-educated labor.

[6] Holzer (2010) has argued that the polarization claim is oversold: while middle-skill jobs have been disappearing in some sectors, they have been stable or growing in others. A different explanation emphasizes feedback effects of the skill-based framework operating through the product market. For example, an increase in relative demand for skilled labor may raise skilled households' demand for service-oriented products, which are intensive in unskilled labor (Mazzolari and Ragusa 2013; Murphy 2016).

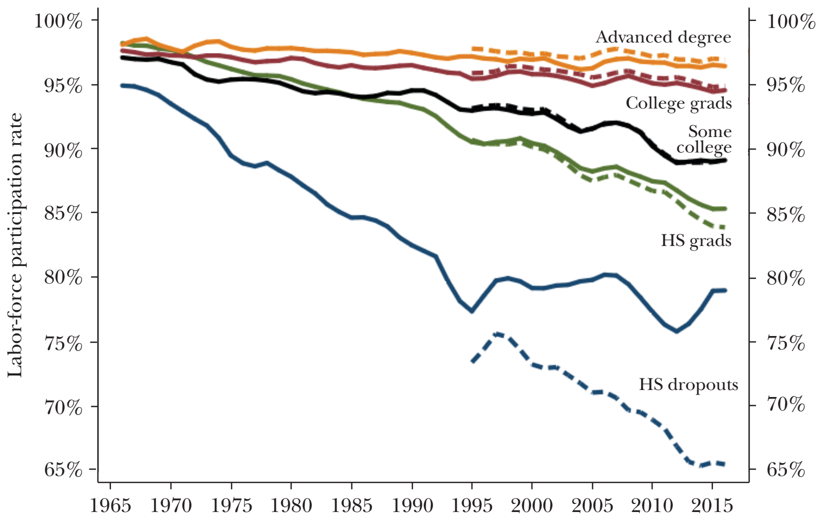
Another factor that has received considerable attention is the assimilation of massive inflows of women and immigrant men into the US labor market. As illustrated in online Appendix Figure A1, women and immigrant men of all education statuses and age levels have increasingly entered the workforce since the 1960s. However, the entry of women has been skill-biased. For example, among 35–44 year-olds, the female share of the college-educated workforce more than doubled, from 25 percent in 1960 to 51 percent in 2016. On the other hand, the female share of the high school graduate workforce increased only slightly, from 38 percent in 1960 to 41 percent in 2016. It is therefore unlikely that the entry of women into the workforce has increased the college wage premium: if anything, such a development should have exerted downward pressure.

The effect of immigration on the male wage distribution is more complicated to evaluate. The increase in competition from immigrant men has been modest and relatively uniform across education statuses—except for high school dropouts, where the share of immigrant men increased from roughly 5 percent to roughly 40 percent between 1960 and 2016. The impact of this influx on native (and immigrant) labor market opportunities remains a hotly debated topic (Card 2009; Borjas 2016; in this journal, Peri 2016). While it might seem that the large inflow of workers with less than a high school education would significantly depress high school dropouts' relative wages, there are several reasons to suspect smaller effects. First, if dropouts and high school graduates are close to perfect substitutes (as Card 2009 argued), the negative impact of this inflow will be diffused across all those with no more than a high school education. Second, if immigrants tend to specialize in certain occupations, this may increase opportunities for workers in other occupations due to production complementarities (Peri and Sparber 2009). Third, the endogenous responses of natives to an influx of immigrants is likely to dilute the economic effects of immigration (for a fuller discussion of each of these issues, see Peri 2016). It does seem likely, however, that immigrants depress relative wages locally in occupations into which they cluster (Cortes 2008; Burstein, Hanson, Tian, and Vogel 2017).

Secular Stagnation

If average wages for prime-age men had been trending upward strongly, then greater wage dispersion could have been accompanied by higher wages for all, even the less-educated. But instead, greater wage dispersion happened against a backdrop of stagnant average wage growth. As reported by the Bureau of Labor Statistics, labor productivity grew at a brisk pace of 2.8 percent per year between 1948 and 1972. Thereafter, between 1973 and 2016, productivity growth averaged only 1.4 percent per year (apart from a temporary boom from 1995 to 2004). Thus, even if real hourly earnings had kept pace with productivity, there would have been a slowdown in average wage growth. But additionally, in the last 20 years, labor's share in national income has steadily dropped (for possible explanations, see Elsby, Hobijn, and Şahin 2013, Karabarbounis and Neiman 2014, Autor, Dorn, Katz, Patterson, and Van Reenen 2017; for illustration, see online Appendix Figure A3), leading real hourly earnings to grow even more slowly than labor productivity. In the face of stagnant overall wage growth created by these two developments, greater wage dispersion has translated into a modest reduction in wages paid to less-skilled men.

Figure 2
Labor-Force Participation Rates by Education Status, Males Aged 25–54, 1965–2016



Source: Authors' calculations based on the March supplement to the Current Population Survey.
Note: This figure plots the 1965–2016 evolution of the labor-force participation rate by education status, using data from the March supplement to the Current Population Survey based on individuals' reported labor-force statuses in the survey week. Dotted lines exclude foreign-born (the CPS begins tracking birthplace in 1994). Within each education status, a re-weighting procedure is employed to hold the age distribution constant across each year. See the online Appendix for further details. Graph presents 3-year centered moving averages. "HS" is high school.

The Secular Decline in Male Labor-Force Participation

As reported by the Bureau of Labor Statistics, the labor-force participation rate of prime-age American men has decreased in a near-continuous fashion from 97.2 percent in 1960 to 88.2 percent in 2015—a cumulative decline of 9 percentage points. Considering that the population of prime-age men in 2015 was around 61.4 million, the secular decline in participation implies a cumulative loss of 5.53 million men from the prime-age workforce.[7]

In Figure 2, we plot the 1965–2016 evolution of the labor-force participation rate by education status, using data from the March supplement to the Current Population Survey based on individuals' reported labor-force statuses in the survey week. The solid series include all prime-age men, while the dashed series exclude

[7] Although we focus here on the labor-force participation rate, the literature has sometimes focused on the employment-to-population ratio. This statistic has also exhibited a long-run decline, but the pattern is far more cyclical (at shorter as well as slightly longer-run frequencies). Business cycle dynamics are not a focus of this paper, although it should be acknowledged that the participation rate exhibits mild cyclicality as well. This was particularly evident in the early 2000s boom (in this journal, Charles, Hurst, and Notowidigo 2016), Great Recession, and subsequent recovery. At the long-run frequency of interest, cyclicality does not affect the participation rate.

foreign-born immigrants. (The Current Population Survey does not record birthplace until 1994.) This exclusion lowers measured participation for high school graduates and dramatically affects the downward trend for those without a high school diploma, reflecting increasing representation of immigrants in this segment of the population combined with their higher participation rates (Borjas 2017). The hourly earnings series presented earlier in Figure 1 are not nearly as sensitive to the exclusion of immigrants.

As in the case of earnings, we observe a dramatic rise in labor-force participation dispersion across education levels. In 1965, the difference in the participation rate between advanced degree holders and high school dropouts was 3.2 percentage points; by 2015, this difference had widened to 17.6 percentage points. Within the US-born population, the difference in the participation rate between advanced degree holders and high school dropouts was a tremendous 31.5 percentage points in 2015. Clearly, most of the secular decline in prime-age male labor-force participation can be attributed to those without a college degree; we focus on this group for the remainder of the paper.

Breaking the full sample up into demographic subgroups by race, age, and education confirms that the same overall pattern of decline in prime-age male labor-force participation holds true within each subgroup (as detailed in online Appendix Figure A4). However, within a given education status, blacks experienced larger declines than whites at all age levels. This is especially true among high school dropouts, where black participation rates tumbled by 30–40 percentage points. We also note some heterogeneity in participation trends by age: while all age groups have withdrawn from the labor force consistently over the sample period, the rate of withdrawal for young workers aged 25–34 relative to older workers increased slightly after the mid-1990s. (This is apart from high school dropouts aged 45–54, who have experienced alarmingly high rates of withdrawal from the labor force.)

Is Nonparticipation Temporary or Permanent?

Does the point-in-time labor-force participation rate reflect brief periods of nonparticipation experienced by a sizable share of the population, or perpetual nonparticipation experienced by a small share of the population? An account of how nonparticipation is distributed in the population provides information about the appropriate theoretical framework for examining change. For example, the canonical neoclassical labor supply framework conceives of the amount of time spent working, given an hourly wage rate, as an optimal tradeoff between income and substitution effects (Deaton and Muellbauer 1980). This framework cannot readily shed light on the decision not to spend any time working, which represents a corner solution. However, if most men who are not participating at a given point in time will participate in the future, then the standard framework remains a useful tool to analyze (changes in) male participation behavior.

Using retrospective reports from the March Current Population Survey, Juhn, Murphy, and Topel (1991, 2002) argued that the rise in male joblessness since the 1960s was almost entirely the result of a growing number of men withdrawing permanently from the workforce. Using contemporaneous reports of participation, Coglianese (2018) recently challenged this contention, showing that retrospective reports result in an undercount of brief nonparticipation spells. We examine this issue by leveraging two longitudinal sources of labor-force participation data. The data come from a US Census Bureau data product that links Survey of Income and Program Participation (SIPP) respondents to their earnings information from the Social Security Administration (SSA).

Using the data from the Survey of Income and Program Participation, we define participation as working at least one week in the given survey month, and report in the top panel of Table 1 nonparticipation to participation (N → P) transition probabilities for currently nonparticipating subgroups of men with high school education or less. We consider two such groups: men who experienced a transition from participation to nonparticipation within the first 12 sample months, and men who began the sample period as nonparticipants. We observe substantial short-run spells of nonparticipation. Among white high school graduates (without college) who experienced a transition to nonparticipation, 51 percent had returned to the labor force after 3 months, and 77 percent had returned to the labor force after 12 months. Among white high school graduates who began the sample period as nonparticipants, 25 percent were in the labor force after 3 months and 49 percent were in the labor force after 12 months. While there are mild education and racial gradients to these estimates, we note striking consistency across demographic groups.

Using earnings histories from the Social Security Administration, which exist on a consistently coded basis for the years 1978–2011, we are also able to construct nonparticipation to participation transition probabilities over a longer-run horizon. These are shown in the second panel of Table 1. Following Coglianese (2018), we define yearly participation as a situation in which total administrative earnings for the year exceed the threshold of one-half of the federal minimum wage times 40 hours per week times 13 weeks. We construct two similar groups of nonparticipants: men who experienced a transition to yearly nonparticipation within the first 15 years observed, and men who were nonparticipants in the first year observed. (Because the definition of participation differs between these short- and long-run analyses, the twelve-month probabilities reported in the top panel of Table 1 differ from the one-year probabilities reported in the bottom panel.) Among white high school graduates who experienced a transition to full-year nonparticipation, 66 percent had achieved annual participation two years later, and 83 percent had achieved annual participation five years later. Among white high school graduates who entered the sample as full-year nonparticipants, 54 percent achieved annual participation after two years, and 76 percent achieved annual participation after five years.

These findings indicate substantial churn into and out of the labor force at long-run as well as short-run frequencies.

How Do Jobless Men Survive?

Another important aspect of labor market nonparticipation is an account of how nonparticipating men obtain resources. Table 2 records annual household income statistics for prime-age non-college-educated men with low labor-force

Table 1

Male Workforce Attachment over the Short and Long Run for Those with High School Education or Less: Evidence from SIPP–SSA Longitudinal Data, 1978–2011

A: Short-run N→P transition probabilities

Group	*Race*	*Education*	*1 month*	*3 months*	*6 months*	*12 months*
Experienced an P→N transition in first 12 panel months	Whites	Dropouts	0.18	0.45	0.63	0.72
		HS grads	0.19	0.51	0.69	0.77
	Blacks	Dropouts	0.16	0.33	0.50	0.59
		HS grads	0.15	0.38	0.57	0.66
Nonparticipant in first panel month	Whites	Dropouts	0.08	0.24	0.38	0.46
		HS grads	0.09	0.25	0.40	0.49
	Blacks	Dropouts	0.05	0.16	0.24	0.30
		HS grads	0.07	0.18	0.29	0.36

B: Long-run N→P transition probabilities

Group	*Race*	*Education*	*1 year*	*2 years*	*5 years*	*10 years*
Experienced an P→N transition in first 15 years observed	Whites	Dropouts	0.42	0.58	0.76	0.83
		HS grads	0.50	0.66	0.83	0.89
	Blacks	Dropouts	0.34	0.50	0.69	0.77
		HS grads	0.40	0.57	0.76	0.83
Non-participant in first year observed	Whites	Dropouts	0.27	0.39	0.54	0.62
		HS grads	0.43	0.54	0.72	0.81
	Blacks	Dropouts	0.29	0.38	0.51	0.57
		HS grads	0.34	0.47	0.63	0.74

Source: Authors' calculations based on the 1984–2008 Survey of Income and Program Participation (SIPP) panels (top panel) and SIPP panels linked to Social Security Administration (SSA) earnings records (bottom panel).

Note: Sample consists of men with high school education or less with 0–30 years of potential experience and not enrolled in school at the beginning of the SIPP panel window. "Dropouts" are high school dropouts. "HS grads" are high school graduates without college. Panel A reports the share of men who transitioned out of nonparticipation by the given month—1, 3, 6, or 12 months after the initial experience of nonparticipation. Two groups of men are considered: those who experienced a transition from participation to nonparticipation (a "P → N" transition) within the first 12 panel months, and those who were nonparticipants at the start of the panel window. Panel B considers the SSA earnings records of all SIPP panel respondents over the subset of years 1978–2011 when the respondents had between 0 and 30 years of potential experience. It reports the share of men who transitioned out of nonparticipation by the given year—1, 2, 5, or 10 years after the initial experience of nonparticipation. Yearly participation is defined as having total administrative earnings for the year above a minimum threshold. Two groups of men are considered: those who experienced a transition to nonparticipation within the first 15 years of observation, and those who were nonparticipants in the first year of observation. See main text and the online Appendix available with this paper at the journal website for further details.

attachment. (We define low labor-force attachment as no more than 13 weeks of employment during the reference year.) The March supplement to the Current Population Survey does not fully distinguish among relevant sources of income until 1992; thus, we consider all years from 1992 to 2017. For each education status, the top panel reports average levels of annual income, rounded to the nearest hundred dollars and broken down by income source. This illustrates how much income the *average* man has access to and where it comes from. The bottom panel records the

Table 2
Household Income Characteristics of Men with Low Labor-Force Attachment by Race, Education, and Age, 1992–2017

	Whites			*Blacks*		
A: High School Dropouts	*25–34*	*35–44*	*45–54*	*25–34*	*35–44*	*45–54*
Average Annual Income ($)						
Own earnings	900	1,100	500	500	400	300
Total unearned income	35,000	28,300	28,400	29,200	28,300	23,500
Own disability-related benefits	3,300	5,700	7,600	2,400	4,200	5,400
Own other unearned income	900	1,100	1,500	400	700	900
Cohabitants' total earnings	21,800	13,100	12,500	18,600	15,800	10,100
Cohabitants' total unearned income	8,000	7,400	6,100	6,600	6,800	6,300
Household food stamps income	1,000	1,000	700	1,200	800	800
Maximal Source of Income (%)						
Own earnings	3	2	2	2	2	1
Own disability-related benefits	12	23	31	10	18	27
Earnings OR unearned income from:						
Parents	30	21	11	38	22	16
Spouse	15	20	20	5	12	12
Other HH members	25	20	21	31	31	26
HH food stamps income	4	3	3	4	3	4
Other source or tie	5	6	7	4	5	6
None (living on < $4 per day)	6	5	5	6	7	8
	Whites			*Blacks*		
B: High School Graduates (no college)	*25–34*	*35–44*	*45–54*	*25–34*	*35–44*	*45–54*
Average Annual Income ($)						
Own earnings	1,300	1,300	1,300	600	700	600
Total unearned income	43,000	35,900	36,200	36,800	29,200	27,800
Own disability-related benefits	3,100	5,300	7,200	2,400	3,600	5,300
Own other unearned income	1,600	2,100	3,600	800	1,300	2,200
Cohabitants' total earnings	30,200	19,900	18,000	26,100	17,000	14,100
Cohabitants' total unearned income	7,400	7,900	6,900	6,600	6,500	5,500
Household food stamps income	700	700	500	900	800	700
Maximal Source of Income (%)						
Own earnings	3	3	2	2	2	2
Own disability-related benefits	9	18	25	7	14	24
Earnings OR unearned income from:						
Parents	35	21	12	41	23	11
Spouse	14	22	24	7	17	19
Other HH members	25	19	19	29	25	23
HH food stamps income	3	3	2	2	3	3
Other source or tie	6	7	10	5	6	8
None (living on < $4 per day)	5	7	6	7	10	10

Source: Authors' calculations based on the March Supplement to the Current Population Survey.
Note: "HH" is household. Sample consist of all households in which at least one prime-age, non-college-educated man with low labor-force attachment resided. Low-labor force attachment is defined as no more than 13 weeks of employment in the reference year. Households with imputed sources of income are excluded. Disability-related benefits are not fully identifiable until 1988; food stamps benefits are not identifiable until 1992; as a result, we consider the years 1992–2017. The top subpanels record average levels of the man's household's yearly income in 2017 dollars, rounded to the nearest hundred and broken down by income source. The bottom subpanels record the frequency with which each source of earnings accounts for the largest share of total household income. Extremely poor households, which subsist on less than $4 per day (with a square-root equivalence scale to adjust for household size), are classified as having no maximal source of income. See main text and the online Appendix available with this paper at the journal website for additional information on data processing and some additional tabulations.

frequency with which each source of income accounts for the largest share of total household income, thereby illustrating *heterogeneity* in income receipt.

Low-participating men earn very little income: average earnings across all demographic groups are at or below $1,300—an extremely low amount to live on for an entire year (top panels of parts A and B of Table 2). Accordingly, only 2–3 percent of the time does the man's own earnings account for the largest share of total household income (bottom panels). Among 45–54 year old men, own-disability-related benefits are relatively important, amounting to around $7,300 for whites and $5,300 for blacks. Own disability benefits constitute the largest share of total household income in 25–30 percent of these cases. While this number is not small, it still represents a distinct minority; and for younger groups, own disability benefits are relatively unimportant. Across all demographic groups, income from other household members appears to be the dominant income source. This is especially true among 25–34 year-old men, where cohabitants' income accounts for 82–86 percent of total household income on average, and some source of cohabitants' income accounts for the largest share of household income 70–75 percent of the time. Even among 45–54 year-old men, some source of cohabitants' income accounts for the largest share of household income in the majority of cases.

In tabulations not shown here, we found that around 40 percent of white high school dropouts, 30 percent of white high school graduates, 53 percent of black high school dropouts, and 44 percent of black high school graduates lived in households with total income below the federal poverty line. Much smaller shares, however—5 to 7 percent of whites and 7 to 10 percent of blacks—lived in extreme poverty (last rows of parts A and B of Table 2). We define extreme poverty as a situation in which household members subsist on less than $4 per day at 2017 prices. (This figure is chosen by implementing a square-root equivalence scale to adjust for household size. For example, the extreme poverty threshold for a four-person household is $4 per day times 365 days times the square root of 4 = $2,920.) Even smaller shares of men—around 3 percent across all groups—lived above the extreme poverty line but depended primarily on food stamps assistance (last rows of parts A and B of Table 2). Thus, while living standards may not be high for these men, the income of other household members plays an effective role in keeping most of them out of extreme poverty.

It is important to reiterate that the above numbers are based on survey data. There is ample reason to believe that the reported earned income figures are underestimates, because episodic sources of income tend to be underreported (Mathiowetz, Brown, and Bound 2002) or even unreported (Edin and Lein 1997; Meyer and Sullivan 2003, in this journal 2012; Bollinger, Hirsch, Hokayem, and Ziliak forthcoming). Recent work using administrative data has demonstrated that survey data understates the incomes of low-income men and overstates the incidence of extreme poverty (Meyer and Mittag 2015; Meyer, Wu, Mooers, and Medalia 2018). However, this work tends to confirm the notion that men with a weak attachment to the workforce largely persist on the incomes of other household members.

Explaining the Decline

Less-educated men have experienced downward trends in wages, employment, and participation since the early 1970s. Therefore, in attempting to explain the latter trends, a reasonable starting point is to ask if changes in the returns to labor market activity can account for the observed decline in labor market activity.

This interpretation of the data gained currency with the work of Juhn, Murphy, and Topel (1991, 2002; see also Juhn 1992). Focusing on the period of the late 1960s through the late 1980s, these authors estimated labor supply curves from repeated cross-sectional and cross-regional data. In both cases, for less-skilled men, they found strong associations between declines in wages and declines in labor market activity, implying an uncompensated labor supply elasticity of around 0.4 for the lowest skilled group.[8] They interpreted their findings as reflecting labor demand shifts against a stable and upward-sloping supply curve.

We are uncomfortable with this interpretation for two reasons. First, it belies the conventional wisdom that the (uncompensated) labor supply curve for men is inelastic. This conventional wisdom comes from the fact that tremendous wage increases in the United States, Canada, the United Kingdom, and Germany between 1900 and 1970 have all been associated with *small declines* in the labor-force participation of prime-age men (for international evidence, see Pencavel 1986; for US evidence, see Ruggles 2015). This notion makes economic sense: if leisure is a superior good, we would expect leisure's share in total household consumption to rise as wages rise. Second, although wages and participation for less-educated men trended downward in tandem throughout the 1970s and 1980s, participation continued to decline after 1995, when wages were comparatively stable (recall Figure 1).

More recent work has also attributed secular decline in less-educated male wages and employment to declining labor demand (Autor 2011; Moffitt 2012; Council of Economic Advisors 2016). For example, Abraham and Kearney (2018) write: "Our review of the evidence leads us to conclude that labor demand factors are the primary drivers of the secular decline in employment over the 1999 to 2016 period." The growth of support for this connection has come from recent studies of local labor demand shocks. As mentioned earlier, Autor, Dorn, and Hanson (2013) found that commuting zones with rising exposure to manufactured imports from China saw substantive declines in manufacturing employment and wages. Similarly, Charles, Hurst, and Schwartz (2018) estimated that a 10 percent decline in manufacturing employment since 2000 led to an 18 percent decline in wages, a 7.9 percent decline in hours worked, and a 4.6 percentage point decline in employment of less-educated men. These estimates imply an uncompensated labor supply elasticity of above 0.4.

[8] Interpreting this estimate as an uncompensated (Marshallian) labor supply elasticity, rather than as an inter-temporal (Frisch) elasticity, implies assuming that workers anticipated the wage changes they experienced to be permanent rather than temporary. Given the apparent protracted nature of local labor demand shocks in low- and medium-skill male sectors (to be discussed below), this assumption seems appropriate.

It is important to reconcile such estimates with the conventional wisdom of a small, or zero, uncompensated elasticity of labor supply. One possible reconciliation recognizes that the labor demand shocks in our period of study not only involve wage cuts, but also job losses, including mass layoffs or the closing of establishments. Displaced male workers take some time to return to the workforce (Jacobson, LaLonde, and Sullivan 1993; Autor, Dorn, Hanson, and Song 2014), in part because current local job losses are often associated with subsequent local job losses. For example, Dix-Carneiro and Kovak (2017) found that initial shocks (albeit in Brazil, not the United States) played out over an extended period partly because of the slow adjustment of capital to the initial shocks. In general, several researchers have documented the slowness of local economies to recover from economic shocks (Bartik 1991; Yagan 2017; Austin, Glaeser, and Summers 2018).

After losing work, individuals may not easily transition into new jobs. For instance, Foote and Ryan (2015) showed that medium-skill displaced male workers tend to face few attractive alternatives within their skill class and sector. Jaimovich and Siu (2018) captured this phenomenon in the context of a search model, in which those losing their jobs in declining sectors of the economy must consider whether to search for new work in the declining sector or invest in new skills. Such a framework can capture business cycle patterns but also secular shifts of the Beveridge curve. A parallel literature has illustrated a reluctance of workers to migrate to better labor markets (Kennan and Walker 2011; Dao, Furceri, and Loungani 2017; Zabek 2018), underscoring the potential importance of migration as well as search frictions.

These dimensions of individual and local labor market adjustment following shocks strike us as key factors underlying the strong secular relationship between wages and participation seen in the 1970s, 1980s, and since 2002. Such adjustment frictions are also consistent with the fact that many spells of nonparticipation are not permanent (recall Table 1)—though it may take some time, most displaced individuals do return to the workforce. A labor supply framework that takes these factors seriously may help resolve the apparent inconsistency between the male labor supply behavior revealed by the historical record and that of the more recent period.

The Expansion of Disability Insurance

A parallel explanation for the large increase in less-educated male joblessness despite a comparatively small decrease in the wages of this group rests on factors operating outside of the labor market. An oft-examined factor is the dramatic increase in the availability and generosity of income maintenance programs targeted at people with disabilities. The largest such program is the Social Security Disability Insurance program (DI), of which we provide a brief history in the online Appendix. The DI rolls expanded considerably between the 1960s and early 1990s and have grown more slowly since (for discussion in this journal, see Liebman 2015; online Appendix Figure A5 graphs trends in the population share of men receiving DI benefits by age group).

To assess the extent to which Disability Insurance expansions have impacted the male labor-force participation rate, Bound (1989) proposed the "rejected-applicant" method, which involves comparing participation behavior of those receiving DI benefits to those who applied for benefits but failed to pass the medical screening necessary to be granted benefits. In a sophisticated application of this method, Maestas, Mullen, and Strand (2013) used random variation in stringency across DI case examiners to estimate that prime-age men on the margin short of DI receipt have a labor-force participation rate of 40 percentage points higher than marginal beneficiaries.

Two issues arise when using this estimate to draw broader inferences about how the Social Security Disability Insurance program has impacted prime-age male labor-force participation. First, if the marginal beneficiary is healthier than the average beneficiary, then the marginal beneficiary's work propensity may be more sensitive to benefit receipt than that of the average beneficiary. Second, estimates based on the rejected-applicant method are *conditional on having applied for benefits*. Since nonparticipation is necessary for eligibility, application for DI benefits itself may reduce participation. It may also affect rejected applicants' subsequent participation behavior (Bound 1989, 1991; Parsons 1991): rejected applicants may reapply, time spent out of work may affect rejected applicants' job prospects, and the act of applying may permanently affect individuals' identities.

Bound, Lindner, and Waidmann (2014) attempted to circumvent these issues by arguing that the participation rate of individuals who report having a health-related work limitation but who never applied for Disability Insurance benefits should be a conservative upper bound on the participation rate of both rejected and accepted applicants. Using SIPP–SSA data, they reported an employment rate differential between prime-age men with work limitations who never applied for DI benefits and those who were awarded benefits of roughly 50 percentage points. The reported employment differential between rejected applicants and those awarded benefits was roughly 30 percentage points.

Pulling these considerations together, let us conservatively assume that the disincentive for labor-force participation induced by applying for disability benefits outweighs the fact that the marginal beneficiary may be healthier than the average beneficiary. Under this assumption, Maestas, Mullen, and Strand's (2013) estimate mentioned above provides a lower bound to the effect of the availability of DI benefits on the labor-force nonparticipation rate of 0.40 times the DI participation rate. Additionally, Bound, Lindner, and Waidmann's (2014) logic yields an upper bound of 0.50 + 0.30 = 0.80 times the DI participation rate under the assumption that the rejected applicant pool is the same size as the beneficiary pool (a slight exaggeration relative to what they report).

Table 3 uses data from the March Current Population Survey to compare prime-age male Disability Insurance and labor-force participation rates, by age and education, between 1975–84 and 2008–2017.[9] (Decadal averages are used to reduce sampling error.) The last row of the table implements the bounding exercise

[9] Published data on the number of individuals receiving Disability Insurance benefits do not report breakdowns by education. However, responses to the March Current Population Survey about income receipt seem to track closely the published administrative statistics on DI receipt.

Table 3
Assessing the Effect of Greater Receipt of Social Security Disability Insurance (SSDI) on the Secular Decline in Male Labor-Force Participation

		Dropouts			*HS graduates*		
		25–34	*35–44*	*45–54*	*25–34*	*35–44*	*45–54*
SSDI participation rate	1975–84	0.030	0.046	0.076	0.009	0.014	0.027
	2008–17	0.022	0.039	0.088	0.022	0.030	0.059
	Change	–0.008	–0.007	0.012	0.013	0.016	0.032
Labor-force nonparticipation rate	1975–84	0.102	0.106	0.153	0.036	0.034	0.070
	2008–17	0.182	0.175	0.304	0.120	0.122	0.171
	Change	0.080	0.069	0.151	0.084	0.088	0.101
SSDI change/OLF change[a]		–0.100	–0.101	0.079	0.155	0.182	0.317
Conservative upper-bound contribution of SSDI growth to nonparticipation growth		–4%	–4%	6%	12%	15%	25%

Source: March Supplement to the Current Population Survey.
Note: "Dropouts" are high school dropouts. "HS grads" are high school graduates without college. Decadal averages are used to reduce sampling error. (Highly) conservative upper bounds equal 0.4 times the ratio of SSDI growth to nonparticipation growth when SSDI growth is negative, and 0.8 times the ratio of SSDI growth to nonparticipation growth when SSDI growth is positive. See main text for a discussion of these upper-bound multipliers.
[a]OLF is "Outside the Labor Force."

just described, reporting a conservative upper bound estimate of the contribution of increased DI program participation to the concomitant rise in nonparticipation. Increased DI participation explains virtually none of the rise in nonparticipation among high school dropouts and little of the rise in nonparticipation among high school graduates (without college) below age 45. On the other hand, it could explain up to 25 percent of the rise in nonparticipation among 45–54 year-old high school graduates (without college). In tabulations not shown, we found similar results for the "some college" group.

In addition to the roughness of this bounding exercise, we note that the growth in Disability Insurance participation may be endogenous to falling labor demand. Previous work has demonstrated that adverse local labor demand shocks expand the DI rolls (Black, Daniel, and Sanders 2002; Autor and Duggan 2003; Charles, Li, and Stephens 2018). Rising DI participation may thus underlie some of the adjustment behavior noted above: rather than search for new work, displaced workers in relatively poor health may simply apply for DI.

Mass Incarceration

The size of the US prison population rose dramatically between 1970 and 2000, from roughly 0.1 percent to 0.5 percent of the total resident population (Raphael and Stoll 2013). We have been studying the civilian, noninstitutionalized male population in this paper—so those incarcerated are excluded. However, as incarceration rates have grown, so has the previously incarcerated share of the currently noninstitutionalized population. Bucknor and Barber (2016) estimated that in 2014, between 6.0 and 7.7 percent of all prime-age men had previously been incarcerated. For black men, the rates were between 19.4 and 21.9 percent; for high school dropout men, the rates were between 26.6 and 30.1 percent. Bucknor and Barber do not report cross-tabulations by education and race, but their figures suggest that potentially a substantial fraction of black male dropouts have prior prison records.

Survey data show that men with criminal records are less likely to be employed than observationally similar individuals without records. Employers report a reluctance to hire men with criminal convictions, while audit studies have found that those with criminal records are less likely to be called back for interviews (Holzer, Raphael, and Stoll 2006a, b; Pager 2003, 2007; Pager, Bonikowski, and Western 2009). The most credible estimates of the effect of prior incarceration on employment leverage random assignment of defendants to courtrooms, judges, and prosecutors. Taking this design to data from Harris Country, Texas, Mueller-Smith (2015) found that each additional year behind bars reduced post-release employment propensity by 3.6 percentage points. For felony defendants with stable pre-charge earnings, incarceration for one or more years reduced employment propensity by much larger amounts. Moreover, Mueller-Smith found prior incarceration to increase subsequent criminal activity. Harding, Morenoff, Nguyen, and Bushway (2018) recently found comparable results using data from Michigan.

These estimates strongly support the notion that incarcerating men is likely to have significant effects on their future labor market outcomes (see also Raphael 2014). That said, there is no straightforward way to make inferences from such estimates about the impact of mass incarceration on noninstitutionalized male labor-force participation. First, the estimates vary considerably across different subpopulations. Second, the statistical design delivers causal effects for the marginal, not the average, man incarcerated. Third, the estimates do not capture spillover effects of mass incarceration on local labor markets and communities.[10] Thus, while we do not feel comfortable using available information to estimate even crudely the effect of mass incarceration on the decline in male labor-force participation since the early 1990s, it seems likely that such effects exist and could be of sizable magnitude, especially among populations who have experienced particularly large exposure to prison (high school dropouts and blacks without any college education). We flag this as an important area for further research.

[10] On spillover effects, two points are in order. First, recent research has found evidence of employer statistical discrimination on expected prior criminal history (Doleac and Hansen 2016; Agan and Starr 2018). These findings suggest that, due to low-skill employers' hiring responses, rising incarceration may impact the labor market prospects of those who were not previously incarcerated but have similar observable characteristics to those who were. Second, men with outstanding warrants may be reticent to seek formal employment, as described in the ethnographic work of Goffman (2014) and supported in statistical analysis by Brayne (2014).

Feedback from the Marriage Market to Male Labor Supply

The above discussions frame a puzzle. On its own, falling labor demand does not sufficiently explain the secular decline in less-educated male labor-force participation—at least, not without allowing for substantial adjustment frictions in the long run as well as the short run. Rising access to Disability Insurance is at most a partial explanation for the 45–54 year-old group and matters quite little for younger men and for high school dropouts. Rising exposure to prison may be a significant factor for dropouts and for blacks without college education, but labor-force participation for these groups began declining decades before prison populations skyrocketed. Certainly no single explanation can sufficiently explain the decline, and even in combination, the explanations appear insufficient.

We suspect that there is another factor at play. We will argue that the prospect of forming and providing for a new family constitutes an important male labor supply incentive; and thus, that developments within the marriage market can influence male labor-force participation. A decline in the formation of stable families produces a situation in which fewer men are actively involved in family provision or can expect to be involved in the future. This removes a labor supply incentive; and the possibility of drawing support from one's existing family, as shown earlier in Table 2, creates a feasible labor-force exit.

The Retreat from Marriage and Onset of Parental Cohabitation

American family structure has undergone dramatic change since the 1960s, featuring a reduction in the incidence of stable two-parent households that has been concentrated in the non-college-educated population (Cherlin 2014; in this journal, Lundberg, Pollak, and Stearns 2016). We summarize these changes in Table 4. Across all demographic groups without college education, the share of men currently married fell dramatically between 1970 and 2015. Currently-marrieds now make up a distinct minority of the black population without college degree. As of 2015, only white high school graduates above age 35 were married a majority of the time, but even these groups experienced a 30-percentage-point drop in currently-married rates since 1970.[11] Similar changes have occurred for the some-college population.

Over the same period, the share of men living with at least one parent has risen in equally dramatic fashion. While parental co-residence was a rare event in 1970, by 2015 over one-quarter of whites and 40 percent of blacks aged 25–34 lived with a parent. Older groups experienced a smaller but still substantial rise in parental co-residence, especially considering the extremely low base rates. For example, white

[11] Some of the measured decline in marriage has been offset by a rise in cohabitation with an unmarried partner. However, for the less-educated, cohabitation is often a transient and unstable situation. Copen, Daniels, and Mosher (2013) report a median duration of cohabitations for less-educated individuals of 22–24 months, based on 2006–2010 data from the National Survey of Family Growth. Ishizuka (2018) reports similar findings in SIPP data, and also finds that cohabitations are less likely to transition into marriages for less-educated individuals. Since our underlying concept of interest is stable family formation, the focus on marriage seems warranted for this population of men.

Table 4
Males Living in Selected Family Arrangements by Race, Education and Age, 1970–2015
(shares)

			Currently married				*Living with parent*			
Race	*Education*	*Age*	*1970*	*1985*	*2000*	*2015*	*1970*	*1985*	*2000*	*2015*
Whites	Dropouts	25–34	0.81	0.61	0.48	0.37	0.12	0.18	0.22	0.28
		35–44	0.86	0.75	0.59	0.51	0.06	0.09	0.14	0.17
		45–54	0.86	0.79	0.66	0.49	0.04	0.05	0.06	0.12
	HS grads	25–34	0.83	0.63	0.53	0.38	0.10	0.14	0.16	0.25
		35–44	0.90	0.79	0.65	0.58	0.04	0.05	0.10	0.13
		45–54	0.89	0.84	0.74	0.61	0.03	0.03	0.05	0.09
Blacks	Dropouts	25–34	0.65	0.31	0.23	0.14	0.17	0.36	0.40	0.47
		35–44	0.66	0.50	0.32	0.27	0.09	0.16	0.25	0.29
		45–54	0.70	0.58	0.44	0.31	0.02	0.07	0.13	0.20
	HS grads	25–34	0.70	0.42	0.32	0.17	0.13	0.26	0.31	0.41
		35–44	0.70	0.58	0.45	0.38	0.08	0.10	0.18	0.22
		45–54	0.75	0.67	0.53	0.46	0.06	0.06	0.11	0.12

Source: Authors' calculations based on the March Supplement to the Current Population Survey.
Note: "Dropouts" are high school dropouts. "HS grads" are high school graduates without college. Statistics were calculated in 5-year windows around the specified years: thus 1970 refers to 1968–72; 1985 refers to 1983–87; 2000 refers to 1998–2002; 2015 refers to 2013–17.

high school graduates aged 35–44 experienced a greater-than-threefold increase in the rate of parental co-residence, from 4 percent in 1970 to 13 percent in 2015. In tabulations not shown, we documented similar increases (though starting from higher base rates) in the shares of men cohabiting with parents *or* other adult relatives.

In this journal, Stevenson and Wolfers (2007) emphasized the contribution of a variety of important developments to the secular decline in marriage, including greater access to contraception, liberalization of family law, changes in home production technology which have reduced gains from task specialization within the household, and changes to how prospective partners match (for example, the rise of online dating). Falling stigmatism of out-of-wedlock childbearing and single motherhood are also likely relevant (Akerlof, Yellen, and Katz 1996).

Others have argued that male earning potential is central to stable marriage formation. Becker (1981) popularized the "specialization-and-exchange" theory of marriage, which predicts that gains from marriage are an increasing function of the gender wage gap. Other work has suggested the influence of "male breadwinner norms": even after accounting for specialization incentives, the marriage rate appears to decline when a local labor market shock makes men less likely to out-earn women (Bertrand, Kamenica, and Pan 2015). A different paradigm contends that marriage occurs when the level of resources exceeds a standard associated with successful family pursuits (Easterlin 1966, 1987). Ruggles (2015) argued for the importance of declining less-educated male earnings prospects relative to those of their fathers. A variant marriage threshold is the typical living standard within a peer reference group (Gibson-Davis, Edin, and McLanahan 2005; Watson and

McLanahan 2011; Ishizuka 2018). In this view, rising income inequality harms the marriage prospects of below-median individuals.

Marriage as a Social Institution for Male Employment

Previous work has associated marriage with a decline in irresponsible male behavior (Akerlof 1998), such as crime (Edlund, Li, Yi, and Zhang 2013) and excessive drug and alcohol use (Duncan, Wilkerson, and England 2006). In the words of Lundberg, Pollak, and Stearns (in this journal, 2016): “If social and economic changes have reduced the value of marriage to noncollege graduates, these changes may also be responsible for a further causal … effect on men’s behavior.”

Male participation in the labor force may also be a socially responsible activity that, like the avoidance of pathological behaviors, is intertwined with stable marriage. To the extent that the gains from marriage depend on male earnings, married men face an additional incentive to find and maintain a job. Indeed, the securing of gainful employment may even be stipulated by men’s (explicit or implicit) marital contracts with their wives. This mechanism has dynamic implications: with the expectation that they will one day marry and provide for a family, single men are incentivized to invest in their future productivity by working today. Doing so also improves their positions in the marriage market, raising their probabilities of matching with high-quality spouses. As marriage rates and attendant marital expectations decrease, so do these labor supply incentives.

We offer an informal assessment of this marriage market mechanism of labor supply in Table 5. We divide the population of interest into three mutually exclusive statuses: men living with at least one parent, unmarried men not living with a parent, and married men not living with a parent. We compute the decline in observed labor-force participation attributable to changes in population shares *between* these statuses versus changes in labor-force participation rates *within* each status during the period from 1970–2015. Our proposed mechanism emphasizes the contribution of shifts *between* statuses (as married men work more than unmarried men and marriage rates have declined), and declining labor-force participation *within* unmarried statuses (as expectations of stable family formation have declined). The latter process should be especially relevant for younger groups, where future marital expectations have plausibly changed the most.

As reported in the last column of Table 5, the decline in labor-force participation among those currently married accounts for a minority of the total observed decline in all demographic groups. This is especially true among blacks: the within-married component accounts for less than one-third of the total decline across all black subgroups. We also observe the predicted age gradient: for groups of men aged 25–34, the within-married component accounts for only about one-quarter of the total change, meaning three-quarters came from the unmarried groups and change in status among the groups. Moreover, the 25–34 age groups have experienced almost as large a decline in labor-force participation as the older groups.

This mechanism offers an interpretation of the puzzle of the recent and disproportionate decline in labor-force participation among young males, identified by Aguiar, Bils, Charles, and Hurst (2017). We repeated the above decomposition for

Table 5
Family Structure Decompositions of Changes in the Male Labor-Force Participation Rate by Race, Education, and Age between 1970 and 2015
(percentage point changes, except last column)

					Change within			
Race	*Education*	*Age*	*Total change in labor-force participation rate*	*Change between*	*Unmarried with parent*	*Unmarried without parent*	*Married*	*Within married contribution to total change*
All Men	Dropouts	25–34	−13.5	−4.7	−3.9	−1.4	−3.5	26%
		35–44	−11.0	−3.7	−2.1	−1.8	−3.5	32%
		45–54	−21.0	−5.5	−2.4	−4.7	−8.4	40%
	HS grads	25–34	−11.2	−3.5	−3.6	−1.3	−2.8	25%
		35–44	−11.0	−3.3	−2.2	−1.7	−3.9	35%
		45–54	−14.8	−3.6	−1.7	−3.8	−5.7	39%
Whites	Dropouts	25–34	−21.2	−6.3	−5.3	−3.7	−6.0	28%
		35–44	−23.3	−5.8	−3.5	−4.5	−9.5	41%
		45–54	−33.1	−6.5	−3.1	−8.6	−14.9	45%
	HS grads	25–34	−10.0	−2.8	−3.1	−1.5	−2.7	27%
		35–44	−10.8	−2.9	−2.4	−1.6	−3.9	36%
		45–54	−13.6	−3.2	−1.6	−3.5	−5.3	39%
Blacks	Dropouts	25–34	−31.6	−11.8	−9.9	−6.2	−3.6	11%
		35–44	−27.5	−8.1	−8.1	−6.5	−4.8	17%
		45–54	−35.5	−10.0	−4.4	−11.3	−9.8	28%
	HS grads	25–34	−18.8	−5.3	−7.0	−2.5	−4.0	21%
		35–44	−15.2	−4.2	−2.8	−3.4	−4.8	32%
		45–54	−20.5	−4.8	−2.2	−7.5	−5.9	29%

Source: Authors' calculations based on the March Supplement to the Current Population Survey.
Note: "Dropouts" are high school dropouts. "HS grads" are high school graduates without college. See online Appendix for detail on how the within/between decomposition was executed. LFP rates were measured in 5-year windows around the beginning and endpoints: thus 1970 refers to 1968–72; 2015 refers to 2013–2017.

the population of men with 0–10 years of potential experience and not enrolled in school, considering changes from 1997–2015. For this group, observed declines in labor-force participation are largely explained by the between-status component and the within-living-with-parents component (as shown in online Appendix Table A4). Today's less-educated labor market entrants face increasingly small probabilities of forming their own stable families. Thus, while Aguiar et al. (2017) attribute their disproportionate withdrawal from the labor force to the rising value of leisure (stemming from innovations in video gaming technology), we suggest that a marriage-market-based fall in the value of work could be an alternative explanation.

Implications for the Relationship between Labor Demand and Labor Supply

The marriage market mechanism we propose adds complexity to the relationship between male labor demand and labor supply. To the extent that male earning potential positively influences marriage, a decline in male labor demand results in fewer males forming and heading their own stable families. Thus, a male labor demand shock produces—through the marriage market—an indirect effect on male labor supply incentives and resultant labor-force participation.

The recent findings of Autor, Dorn, and Hanson (2018) indicate that this indirect channel may be important. These authors found that rising local exposure to import competition from China led to local declines in the marriage rate and in the share of children living with their fathers. Thus, the large employment effects of local labor demand shocks may embed family-related effects in addition to the other adjustment frictions we have discussed. Family processes may also interact with these other adjustment frictions: for example, by making men more dependent on their adult relatives and less amenable to finding a new occupation or labor market in which to seek employment. In addition, others have argued that mass incarceration has disrupted family formation (for example, Charles and Luoh 2010; Schneider, Harknett, and Stimpson 2018). These relationships further complicate the separate quantification of the various forces driving the decline in male labor-force participation.

Conclusion

During the last 50 years, the earnings of prime-age men in the United States have stagnated and dispersed across the education distribution. At the same time, the labor-force participation rates of men without a college education have steadily declined. While wage and participation trends are often linked for this population, we have argued that this connection cannot solely be the result of an inward labor demand shift across a stable and elastic labor supply curve. The uncompensated labor supply elasticities implied by the twin declines of wages and participation during the 1970s, 1980s, and 2000s appear too large to be plausible. Moreover, labor-force participation continued to decrease in the 1990s while wages were rising. While the increasing availability of disability benefits and the increase in the fraction of the population with prior incarceration exposure may help explain some of the participation decline, we doubt either factor can explain the bulk of the decline.

We have argued that more plausible explanations for the observed patterns involve feedbacks from male labor demand shocks, which often involve substantial job displacement, to worker adjustment frictions and to family structure. Marriage rates, and corresponding male labor supply incentives, have also fallen for reasons other than changing labor demand. Moreover, we have noted interactions between labor demand and disability benefit take-up, and between mass incarceration and family structure. These factors have all converged to reduce the feasibility and desirability of stable employment, leading affected men—who may not often be eligible for disability or other benefits—to participate sporadically in the labor market and depend primarily on family members for income support. In sum, our observations lead us to be skeptical of attempts to attribute the secular decline in male labor-force participation to a series of causal factors acting separately. We prefer an interpretation in which falling labor demand and other interconnected factors have led to substantially lower participation rates among men with less than a college education.

■ *We thank Charlie Brown, Gordon Hanson, Enrico Moretti, Mel Stephens, and Timothy Taylor for helpful comments. Binder acknowledges research support from an NICHD training grant to the Population Studies Center at the University of Michigan (T32 HD007339). Tyler Radler provided excellent research assistance.*

References

Abraham, Katharine G., and Melissa S. Kearney. 2018. "Explaining the Decline in the U.S. Employment-to-Population Ratio: A Review of the Evidence." NBER Working Paper 24333.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Chap. 12 in *Handbook of Labor Economics*, edited by David Card and Orley Ashenfelter, vol. 4, Part B, pp. 1043–1171. Elsevier.

Acemoglu, Daron, and Pascual Restrepo. 2018. "Automation and New Tasks: The Implications of the Task Content of Production for Labor Demand." November, 6. https://economics.mit.edu/files/16338.

Agan, Amanda, and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *Quarterly Journal of Economics* 133(1): 191–235.

Aguiar, Mark, Mark Bils, Kerwin Kofi Charles, and Erik Hurst. 2017. "Leisure Luxuries and the Labor Supply of Young Men." NBER Working Paper 23552.

Akerlof, George A. 1998. "Men Without Children." *Economic Journal* 108(447): 287–309.

Akerlof, George A., Janet L. Yellen, and Michael L. Katz. 1996. "An Analysis of Out-of-Wedlock Childbearing in the United States." *Quarterly Journal of Economics* 111(2): 277–317.

Altonji, Joseph G., Prashant Bharadwaj, and Fabian Lange. 2012. "Changes in the Characteristics of American Youth: Implications for Adult Outcomes." *Journal of Labor Economics* 30(4): 783–828.

Austin, Benjamin, Edward Glaeser, and Lawrence H. Summers. 2018. "Saving the Heartland: Place-Based Policies in 21st Century America." *Brookings Papers on Economic Activity,* Spring.

Autor, David H. 2014. "Skills, Education, and the Rise of Earnings Inequality among the 'Other 99 Percent.'" *Science* 344(6186): 843–51.

Autor, David H. 2011. "The Polarization of Job Opportunities in the U.S. Labor Market: Implications for Employment and Earnings." *Community Investments* 23(2): 11–16.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.

Autor, David H., David Dorn, and Gordon Hanson. 2018. "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men." Discussion Paper 11465, IZA Institute of Labor Economics.

Autor, David H., David Dorn, Gordon H. Hanson, and Jae Song. 2014. "Trade Adjustment: Worker-Level Evidence." *Quarterly Journal of Economics* 129(4): 1799–1860.

Autor, David H., David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. "Concentrating on the Fall of the Labor Share." *American Economic Review* 107(5): 180–85.

Autor, David H., and Mark G. Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118(1): 157–206.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the U.S. Labor Market." *American Economic Review* 96(2): 189–94.

Autor, David H, Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90(2): 300–323.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113(4): 1169–1213.

Autor, David H., Frank Levy, and Richard J. Murnane. 2002. "Upstairs, Downstairs: Computers and Skills on Two Floors of a Large Bank." *ILR Review* 55(3): 432–47.

Autor, David H., Frank Levy, and Richard J.

Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118(4): 1279–1333.

Autor, David H., Alan Manning, and Christopher L. Smith. 2016. "The Contribution of the Minimum Wage to US Wage Inequality over Three Decades: A Reassessment." *American Economic Journal: Applied Economics* 8(1): 58–99.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: Upjohn Press.

Becker, Gary S. 1981. *A Treatise on the Family.* Cambridge, MA: Harvard University Press.

Berman, Eli, John Bound, and Zvi Griliches. 1994. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics* 109(2): 367–97.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130(2): 571–614.

Bivens, L. Josh. 2007. "The Gains from Trade: How Big and Who Gets Them?" EPI Working Paper 280, Economic Policy Institute.

Black, Dan, Kermit Daniel, and Seth Sanders. 2002. "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust." *American Economic Review* 92(1): 27–50.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55(3): 789–865.

Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. Forthcoming. "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch." *Journal of Political Economy.*

Borjas, George J. 2016. *We Wanted Workers: Unraveling the Immigration Narrative.* New York, NY: W. W. Norton & Company.

Borjas, George J. 2017. "The Labor Supply of Undocumented Immigrants." *Labour Economics* 46: 1–13.

Bound, John. 1989. "The Health and Earnings of Rejected Disability Insurance Applicants." *American Economic Review* 79(3): 482–503.

Bound, John. 1991. "The Health and Earnings of Rejected Disability Insurance Applicants: Reply." *American Economic Review* 81(5): 1427–34.

Bound, John, and George E. Johnson. 1992. "Changes in the Structure of Wages During the 1980's: An Evaluation of Alternative Explanations." *American Economic Review* 82(3): 371–92.

Bound, John, Stephan Lindner, and Timothy Waidmann. 2014. "Reconciling Findings on the Employment Effect of Disability Insurance." *IZA Journal of Labor Policy* 3: Article 11.

Brayne, Sarah. 2014. "Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment." *American Sociological Review* 79(3): 367–91.

Bucknor, Cherrie, and Alan Barber. 2016. "The Price We Pay: Economic Costs of Barriers to Employment for Former Prisoners and People Convicted of Felonies." CEPR Report, Center for Economic and Policy Research. June. http://cepr.net/publications/reports/the-price-we-pay-economic-costs-of-barriers-to-employment-for-former-prisoners-and-people-convicted-of-felonies.

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel. 2017. "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence from the U.S." NBER Working Paper 23330.

Card, David. 2009. "Immigration and Inequality." *American Economic Review* 99(2): 1–21.

Card, David, and Thomas Lemieux. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis." *Quarterly Journal of Economics* 116(2): 705–46.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2016. "The Masking of the Decline in Manufacturing Employment by the Housing Bubble." *Journal of Economic Perspectives* 30(2): 179–200.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2018. "The Transformation of Manufacturing and the Decline in U.S. Employment." *NBER Macroeconomics Annual* vol. 33. http://www.nber.org/chapters/c14081.

Charles, Kerwin Kofi, Yiming Li, and Melvin Stephens Jr. 2018. "Disability Benefit Take-Up and Local Labor Market Conditions." *Review of Economics and Statistics* 100(3): 416–23.

Charles, Kerwin Kofi, and Ming Ching Luoh. 2010. "Male Incarceration, the Marriage Market, and Female Outcomes." *Review of Economics and Statistics* 92(3): 614–27.

Cherlin, Andrew J. 2014. *Labor's Love Lost: The Rise and Fall of the Working-Class Family in America.* New York: Russell Sage Foundation.

Cline, William R. 1997. *Trade and Income Distribution.* Washington, DC: Peterson Institute for International Economics.

Coglianese, John. 2018. "The Rise of In-and-Outs: Declining Labor Force Participation of Prime Age Men." https://scholar.harvard.edu/files/coglianese/files/coglianese_2017_in-and-outs.pdf.

Copen, Casey E., Kimberly Daniels, and William D. Mosher. 2013. "First Premarital Cohabitation in the United States: 2006–2010 National Survey of Family Growth." *National Health Statistics Reports* 64: 1–15.

Cortes, Patricia. 2008. "The Effect of Low-Skilled

Immigration on U.S. Prices: Evidence from CPI Data." *Journal of Political Economy* 116(3): 381–422.

Council of Economic Advisers. 2016. "The Long-Term Decline in Prime-Age Male Labor-Force Participation." June. https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160620_cea_primeage_male_lfp.pdf.

Dao, Mai, Davide Furceri, and Prakash Loungani. 2017. "Regional Labor Market Adjustment in the United States: Trend and Cycle." *Review of Economics and Statistics* 99(2): 243–57.

Deaton, Angus, and John Muellbauer. 1980. *Economics and Consumer Behavior.* New York, NY: Cambridge University Press.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach." *Econometrica* 64(5): 1001–44.

Dix-Carneiro, Rafael, and Brian K. Kovak. 2017. "Trade Liberalization and Regional Dynamics." *American Economic Review* 107(10): 2908–46.

Doleac, Jennifer L., and Benjamin Hansen. 2016. "Does 'Ban the Box' Help or Hurt Low-Skilled Workers? Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden." NBER Working Paper 22469.

Duncan, Greg J., Bessie Wilkerson, and Paula England. 2006. "Cleaning Up Their Act: The Effects of Marriage and Cohabitation on Licit and Illicit Drug Use." *Demography* 43(4): 691–710.

Easterlin, Richard A. 1987. *Birth and Fortune*, 2nd ed. University of Chicago Press.

Easterlin, Richard A. 1966. "On the Relation of Economic Factors to Recent and Projected Fertility Changes." *Demography* 3(1): 131–53.

Edin, Kathryn J., and Laura Lein. 1997. *Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work.* New York, NY: Russell Sage Foundation.

Edlund, Lena, Hongbin Li, Junjian Yi, and Junsen Zhang. 2013. "Sex Ratios and Crime: Evidence from China." *Review of Economics and Statistics* 95(5): 1520–34.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity* no. 2: 1–63.

Feenstra, Robert C., and Gordon H. Hanson. 1996. "Globalization, Outsourcing, and Wage Inequality." *American Economic Review* 86(2): 240–45.

Feenstra, Robert C., and Gordon H. Hanson. 1999. "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979–1990." *Quarterly Journal of Economics* 114(3): 907–40.

Foote, Christopher L., and Richard W. Ryan. 2015. "Labor-Market Polarization over the Business Cycle." *NBER Macroeconomics Annual* 29(1): 371–413.

Fortin, Nicole M., and Thomas Lemieux. 1997. "Institutional Changes and Rising Wage Inequality: Is There a Linkage?" *Journal of Economic Perspectives* 11(2): 75–96.

Fortin, Nicole M., Thomas Lemieux, and Neil Lloyd. 2018. "Labor Market Institutions and the Distribution of Wages: The Role of Spillover Effects." September. https://econ2017.sites.olt.ubc.ca/files/2018/12/pdf_paper_thomas-lemieux-spillovers.pdf.

Freeman, Richard B. 1975. "Overinvestment in College Training?" *Journal of Human Resources* 10(3): 287–311.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254–80.

Gibson-Davis, Christina M., Kathryn Edin, and Sara McLanahan. 2005. "High Hopes but Even Higher Expectations: The Retreat from Marriage among Low-Income Couples." *Journal of Marriage and Family* 67(5): 1301–12.

Goffman, Alice. 2014. *On the Run: Fugitive Life in an American City.* University of Chicago Press.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology.* Cambridge, Massachusetts: Belknap Press.

Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway. 2018. "Imprisonment and Labor Market Outcomes: Evidence from a Natural Experiment." *American Journal of Sociology* 124(1): 49–110.

Hirsch, Barry T. 1988. "Trucking Regulation, Unionization, and Labor Earnings: 1973–85." *Journal of Human Resources* 23(3): 296–319.

Holzer, Harry J. 2010. "Is the Middle of the U.S. Job Market Really Disappearing? A Comment on the 'Polarization' Hypothesis." May 13, Center for American Progress. https://www.americanprogress.org/issues/economy/reports/2010/05/13/7843/is-the-middle-of-the-us-job-market-really-disappearing/.

Holzer, Harry J., Steven Raphael, and Michael A. Stoll. 2006a. "How Do Employer Perceptions of Crime and Incarceration Affect the Employment Prospects of Less-Educated Young Black Men?" Chap. 4 in *Black Males Left Behind*, edited by Ronald B. Mincy. Washington, DC: The Urban Institute Press.

Holzer, Harry J., Steven Raphael, and Michael A. Stoll. 2006b. "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers." *Journal of Law and Economics* 49(2): 451–80.

Ishizuka, Patrick. 2018. "The Economic Foundations of Cohabiting Couples' Union Transitions." *Demography* 55(2): 535–57.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83(4): 685–709.

Jaimovich, Nir, and Henry E. Siu. 2018. "Job Polarization and Jobless Recoveries." NBER Working Paper 18334.

Juhn, Chinhui. 1992. "Decline of Male Labor Market Participation: The Role of Declining Market Opportunities." *Quarterly Journal of Economics* 107(1): 79–121.

Juhn, Chinhui, Kevin M. Murphy, and Robert H. Topel. 1991. "Why Has the Natural Rate of Unemployment Increased over Time?" *Brookings Papers on Economic Activity*, no. 2, pp. 75–142.

Juhn, Chinhui, Kevin M. Murphy, and Robert H. Topel. 2002. "Current Unemployment, Historically Contemplated." *Brookings Papers on Economic Activity* no. 1, pp. 79–116.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129(1): 61–103.

Katz, Lawrence F. 2008. Comment on "Trade and Wages, Reconsidered" by Paul Krugman. *Brookings Papers on Economic Activity*, no. 1: 143–49.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107(1): 35–78.

Kennan, John, and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79(1): 211–51.

Krugman, Paul R. 2008. "Trade and Wages, Reconsidered." *Brookings Papers on Economic Activity* no. 1, pp. 103–54.

Lee, David S. 1999. "Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?" *Quarterly Journal of Economics* 114(3): 977–1023.

Levy, Frank, and Richard J. Murnane. 2004. *The New Division of Labor: How Computers Are Creating the Next Job Market.* Princeton University Press; Russell Sage Foundation.

Liebman, Jeffrey B. 2015. "Understanding the Increase in Disability Insurance Benefit Receipt in the United States." *Journal of Economic Perspectives* 29(2): 123–50.

Lundberg, Shelly, Robert A. Pollak, and Jenna Stearns. 2016. "Family Inequality: Diverging Patterns in Marriage, Cohabitation, and Childbearing." *Journal of Economic Perspectives* 30(2): 79–102.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103(5): 1797–1829.

Mathiowetz, Nancy A., Charles Brown, and John Bound. 2002. "Measurement Error in Surveys of the Low-Income Population." Chap. in *Studies of Welfare Populations: Data Collection and Research Issues.* National Academies Press.

Mazzolari, Francesca, and Giuseppe Ragusa. 2013. "Spillovers from High-Skill Consumption to Low-Skill Labor Markets." *Review of Economics and Statistics* 95(1): 74–86.

Meyer, Bruce D., and Nikolas Mittag. 2015. "Using Linked Survey Data and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net." NBER Working Paper 21676.

Meyer, Bruce D., and James X. Sullivan. 2003. "Measuring the Well-Being of the Poor Using Income and Consumption." *Journal of Human Resources* 38(Special Issue on Income Volatility and Implications for Food Assistance Programs): 1180–1220.

Meyer, Bruce D., and James X. Sullivan. 2012. "Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure." *Journal of Economic Perspectives* 26(3): 111–36.

Meyer, Bruce D., Derek Wu, Victoria Mooers, and Carla Medalia. 2018. "The Use and Misuse of Income Data and the Rarity of Extreme Poverty in the United States." Working paper.

Moffitt, Robert A. 2012. "The Reversal of the Employment-Population Ratio in the 2000s: Facts and Explanations." *Brookings Papers on Economic Activity*, Fall, pp. 201–64.

Mueller-Smith, Michael. 2015. "The Criminal and Labor Market Impacts of Incarceration." https://sites.lsa.umich.edu/mgms/wp-content/uploads/sites/283/2015/09/incar.pdf.

Murphy, Daniel. 2016. "Welfare Consequences of Asymmetric Growth." *Journal of Economic Behavior & Organization* 126(Part A): 1–17.

Murphy, Kevin M., and Finis Welch. 1991. "The Role of International Trade in Wage Differentials." Chap. 2 in *Workers and Their Wages Changing Patterns in the United States,* edited by Marvin H. Kosters. Washington, DC: AEI Press.

Murphy, Kevin M., and Finis Welch. 1992. "The Structure of Wages." *Quarterly Journal of Economics* 107(1): 285–326.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108(5): 937–75.

Pager, Devah. 2007. *Marked: Race, Crime, and Finding Work in an Era of Mass Incarceration.* University of Chicago Press.

Pager, Devah, Bart Bonikowski, and Bruce Western. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American*

Sociological Review 74(5): 777–99.

Parsons, Donald O. 1991. "The Health and Earnings of Rejected Disability Insurance Applicants: Comment." *American Economic Review* 81(5): 1419–26.

Pencavel, John. 1986. "Labor Supply of Men: A Survey." Chap. 1 in *Handbook of Labor Economics*, vo1. 1, edited by Orely C. Ashenfelter and Richard Layard, 3–102. Elsevier.

Peri, Giovanni. 2016. "Immigrants, Productivity, and Labor Markets." *Journal of Economic Perspectives* 30(4): 3–30.

Peri, Giovanni, and Chad Sparber. 2009. "Task Specialization, Immigration, and Wages." *American Economic Journal: Applied Economics* 1(3): 135–69.

Raphael, Steven. 2014. *The New Scarlet Letter? Negotiating the U.S. Labor Market with a Criminal Record.* Kalamazoo, MI: W.E. Upjohn Institute.

Raphael, Steven, and Michael A. Stoll. 2013. *Why Are So Many Americans in Prison?* Russell Sage Foundation.

Rose, Nancy L. 1987. "Labor Rent Sharing and Regulation: Evidence from the Trucking Industry." *Journal of Political Economy* 95(6): 1146–78.

Ruggles, Steven. 2015. "Patriarchy, Power and Pay: The Transformation of American Families, 1800–2015." *Demography* 52(6): 1797–1823.

Schneider, Daniel, Kristen Harknett, and Matthew Stimpson. 2018. "What Explains the Decline in First Marriage in the United States? Evidence from the Panel Study of Income Dynamics, 1969 to 2013." *Journal of Marriage and Family* 80(4): 791–811.

Stevenson, Betsey, and Justin Wolfers. 2007. "Marriage and Divorce: Changes and Their Driving Forces." *Journal of Economic Perspectives* 21(2): 27–52.

Watson, Tara, and Sara McLanahan. 2011. "Marriage Meets the Joneses: Relative Income, Identity, and Marital Status." *Journal of Human Resources* 46(3): 482–517.

Yagan, Danny. 2017. "Employment Hysteresis from the Great Recession." NBER Working Paper 23844.

Zabek, Mike. 2018. "Local Ties in Spatial Equilibrium." https://mikezabek.com/pdf/LocalTies.pdf.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 191–210

When Labor's Lost: Health, Family Life, Incarceration, and Education in a Time of Declining Economic Opportunity for Low-Skilled Men

Courtney C. Coile and Mark G. Duggan

The economic progress of US men has stagnated in recent decades. The labor force participation rate of men ages 25–54 peaked at 97 percent in the mid-1960s and has declined by roughly eight percentage points since then (according to the Bureau of Labor Statistics), while men's real median earnings have been flat since the early 1970s (Fontenot, Semega, and Kollar 2018, figure 2). These population averages mask considerably larger declines in participation among less-educated and non-white men (as discussed in this symposium by Binder and Bound) as well as substantial increases in wage inequality (Autor, Katz, and Kearney 2008). The decline in economic opportunities for low-skilled men and the possible negative effects of this trend on their well-being is a matter of increasingly urgent concern for policymakers and the general public.

In this paper, we seek to illuminate the broader context in which prime-age men are experiencing economic stagnation. We explore changes for prime-age men over time in education, mortality, morbidity, disability program receipt, family structure, and incarceration rates. We focus on prime-age men, namely those ages 25–54, and on the years 1980–2016 (or 2017 when possible), encompassing much of the period of reduced economic progress for low-skilled men. Where possible, we examine trends by education, and in some cases, draw comparisons between

■ *Courtney C. Coile is Professor of Economics and Director of the Knapp Social Science Center, Wellesley College, Wellesley, Massachusetts. Mark G. Duggan is the Wayne and Jodi Cooperman Professor of Economics and the Trione Director of the Stanford Institute for Economic Policy Research, Stanford University, Stanford, California. Their email addresses are ccoile@wellesley.edu and mgduggan@stanford.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.191 doi=10.1257/jep.33.2.191

men and women or highlight trends by race and ethnicity or geography. In the concluding discussion, we explore the relevance of these trends in the context of men's economic stagnation. Some of the key indicators that we discuss may be affected by men's sluggish economic progress or play a role in explaining it, or both. While establishing causality for such a wide range of health and other outcomes is inherently difficult, we will discuss some of the clues provided by recent research.

Our approach is consistent with Case and Deaton's (2017) theory of "cumulative disadvantage." While their approach was motivated by a rise in "deaths of despair" from drug poisonings, suicide, and alcohol-related liver disease, particularly among less-educated, non-Hispanic whites, they posit that worsening labor market opportunities for successive cohorts of less-educated whites affect and are affected by a cluster of factors including health, education, and marriage and family outcomes.

We build on their findings while considering additional measures of well-being among prime-age men and also highlighting some positive developments during the latter half of our analysis period. Interestingly, these more recent developments have tended to benefit white prime-age men much less than other men in the 25–54 age range.

Educational Attainment

Gains in the educational attainment of prime-age males have slowed over time. As illustrated in Table 1, between 1980 and 2000, men ages 45–54 (of all races) experienced a 22 percentage point decline in the share that were high school dropouts, as well as a 15-point increase in the share with some college and a 12-point increase in the share with a college degree. This reflects the fact that men in the first half of the Baby Boom cohort (those born between 1946 and 1955, who were ages 45–54 in 2000) had very different levels of educational attainment than those born 20 years earlier. By contrast, changes between 2000 and 2017 were minimal, as men born nearly two decades later (between 1963 and 1972, or roughly in the first half of the Generation X cohort) made similar educational choices as had the earlier cohorts. Recently, there has been a more modest increase in men's education, as the share of men ages 25–34 with a college degree grew by 5 percentage points between 2000 and 2017, indicating that birth cohorts from the Millenial generation are slightly more likely to seek higher education than were Gen Xers.

Gains in educational attainment for men have also lagged behind gains for women in recent years. In 1980, the share of men age 25–34 with a college degree exceeded the share of women with such a degree by 7 percentage points (28 percent for men versus 21 percent for women). But in the mid-1990s, the share of women ages 25–34 with a college degree surpassed that of men, and by 2017, this gender gap had grown to 7 percentage points in favor of women (34 for men versus 41 percent for women). Jacob (2002) finds that among a cohort making these decisions in the

Table 1
Educational Attainment, Men Ages 25–54, by Race, 1980 to 2017

Age group	*Education level*	*1980*	*2000*	*2017*	*Change, 1980–2000 (in percentage points)*	*Change, 2000–2017 (in percentage points)*
All Men						
25–34	< High school	14%	13%	9%	–1	–4
	High school	35%	32%	29%	–3	–3
	Some college	23%	26%	28%	+3	+2
	College	28%	29%	34%	+1	+5
35–44	< High school	22%	12%	11%	–10	–1
	High school	37%	35%	28%	–2	–7
	Some college	17%	26%	25%	+9	–1
	College	25%	27%	36%	+2	+9
45–54	< High school	33%	12%	11%	–22	–1
	High school	34%	29%	32%	–6	+4
	Some college	12%	27%	24%	+15	–3
	College	20%	32%	33%	+12	0
Black Men						
25–34	< High school	25%	12%	10%	–12	–3
	High school	39%	41%	34%	+2	–8
	Some college	24%	28%	35%	+4	+7
	College	12%	18%	22%	+6	+3
35–44	< High school	37%	12%	10%	–26	–1
	High school	41%	42%	35%	0	–7
	Some college	14%	30%	29%	+15	0
	College	7%	17%	26%	+10	+8
45–54	< High school	60%	19%	11%	–41	–8
	High school	25%	34%	40%	+8	+6
	Some college	8%	28%	28%	+20	–1
	College	7%	19%	22%	+12	+2

Source: Authors using data from the Current Population Survey.
Note: Changes listed may differ slightly from implied changes due to rounding.

mid-1990s, differences in the return to attending college and in noncognitive skills[1] accounted for the vast majority of the female advantage in college attendance.

There have also been differential gains in educational attainment by race. As shown in Table 1, the share of black men ages 45–54 who had not completed high school fell by nearly 50 percentage points between 1980 and 2017, more than twice the decline for all men. Similarly, black men in each of the three age groups had larger increases in the share with some college or with a college degree between

[1] Noncognitive skills include the ability to pay attention in class, work with and seek help from others, and organize.

1980 and 2017 than did US men as a whole. While there remains a racial gap in educational attainment, its magnitude shrank considerably during this era.[2] Trends for native-born Hispanics during this time period were similar to those for blacks (Ryan and Bauman 2016).

Given the steady increase in the return to a college degree during our study period (Card and Lemieux 2001), one might have expected prime-age men to respond—as women appear to have done (Jacob 2002)—by attending college and community college in greater numbers. The much slower increase in educational attainment among men during the 1980–2017 period partially explains why the earnings of full-time male workers actually fell by 1 percent (in real terms) during this nearly 40-year period while earnings of full-time female workers increased by 32 percent (Fontenot, Semega, and Kollar 2018). The fact that men ages 25–34 experienced larger gains in educational attainment in the period since 2000 than between 1980 and 2000 is consistent with more recent cohorts of men responding to increasing (or increasingly evident) returns to education, but further research is needed to establish this pathway.

Mortality

According to mortality rates, which are probably the most widely used measure of health, the health of prime-age men steadily improved during the 1980s and 1990s. As shown in Table 2, between 1980 and 2000, mortality rates among men ages 25–34, 35–44, and 45–54 declined at annual rates of 1.7 percent, 0.8 percent, and 1.7 percent, respectively. Weighting each of these three age groups equally to reduce sensitivity to changes in the age distribution over time, the mortality rate among men ages 25–54 fell at an annual rate of 1.5 percent, from 421 per 100,000 in 1980 to 312 per 100,000 in 2000, a 26 percent decline.[3] This was larger than the corresponding drop of 20 percent among prime-age women during the same period. Despite this differential improvement, prime-age men still had an 80 percent higher mortality rate than prime-age women in 2000.

The substantial reduction in mortality among prime-age men did not continue in the subsequent years, with the mortality rate falling by less than 2 percent (from 312 to 307 per 100,000) from 2000 to 2016. This overall change masks substantial heterogeneity among the three age groups. Perhaps most strikingly, the mortality rate among men ages 25–34 increased by 28 percent (from 139 to 178 per 100,000) during this 16-year period, corresponding to an annual growth rate of 1.6 percent.

[2] Western and Pettit (2000) caution that the exclusion of incarcerated persons from the Current Population Survey may lead to an overestimate of the rise in black men's education, as incarceration rates for black men have risen over time and incarcerated black males tend to have lower levels of education than non-incarcerated black males.

[3] Gelman and Auerbach (2016) argue for using even finer age adjustments when assessing mortality trends, as a population shift within a given age group—for example, having fewer 45 year-olds and more 54 year-olds—can affect the mortality rate within even a narrowly-defined group.

Table 2
Male Mortality Rates by Age and Cause, Ages 25 to 54, 1980 to 2016

	Annual Mortality Rates per 100,000			*Annual % change*	
	1980	*2000*	*2016*	*1980–2000*	*2000–2016*
By Age					
25 to 34	196	139	178	–1.7%	+1.6%
35 to 44	299	255	244	–0.8%	–0.3%
45 to 54	767	543	498	–1.7%	–0.5%
By Race[a]					
Black	859	577	429	–2.0%	–1.8%
White	376	285	301	–1.4%	+0.3%
Hispanic[b]	–	257	212	–	–1.2%
Non-Hispanic White[b]	–	285	320	–	+0.7%
By Cause[a]					
Heart disease	121	64	53	–3.2%	–1.2%
Cancer	82	58	42	–1.8%	–2.0%
Accidents	65	48	76	–1.4%	+2.9%
Suicides	24	22	27	–0.4%	+1.5%
Homicides	25	11	14	–3.9%	+1.2%
HIV/AIDS	0	15	4	–	–8.3%
All other	105	94	91	–0.5%	–0.2%
Total	421	312	307	–1.5%	–0.1%

Source: Authors using data from Centers for Disease Control and Prevention, National Center for Health Statistics at https://www.cdc.gov/nchs/hus/contents2017.htm.
[a] The mortality rate for men ages 25–54 is computed as a simple average of the rate for men ages 25–34, 35–44, and 45–54, to minimize the effect of changing age distribution of the population over time.
[b] Mortality data is not available by Hispanic ethnicity status in 1980.

While mortality rates did fall for men ages 35–44 and 45–54, the annual rate of decline was much slower than it was from 1980 through 2000.

A closer examination of the changes in prime-age mortality rates by cause, as reported in Table 2, helps to explain these patterns. From 1980 to 2000, more than half of the improvement for prime-age men was driven by falling heart disease mortality, while nearly one-fourth was the result of an impressive reduction in cancer mortality. The rest was explained primarily by large reductions in homicides and in accidental (primarily motor vehicle) deaths. The only notable increase in cause-specific mortality during this time period was for HIV/AIDS, which accounted for more deaths than homicides by 2000.

From 2000 to 2016, the pace of improvement in heart disease mortality slowed considerably, with the annual mortality rate falling by just 11 deaths per 100,000 between 2000 and 2016 (versus a drop of 57 deaths per 100,000 between 1980 and 2000). The death rate from accidents rose substantially, with this entirely driven by an increase in drug overdose deaths, which nearly quadrupled—from 12 to 45 per 100,000—between 2000 and 2016, even as mortality from motor vehicle and other accidents fell modestly. In addition, there was an increase in both the suicide rate and in the homicide rate, both of which had declined between 1980 and 2000. Offsetting these increases were declines in cancer mortality, which continued at a pace similar to that seen in the earlier period, as well as a remarkable 75 percent decline in the mortality rate from HIV/AIDS, which went from being the fifth most common cause of death among prime-age men in 2000 to the tenth most common in 2016.

These trends in mortality by cause of death also explain why young men fared substantially worse in the more recent period. Mortality rates from drug overdose and suicide as well as the increase in these rates between 2000 and 2016 are very similar across all age groups. By contrast, men ages 45–54 die from heart disease and cancer at rates more than ten times those of men 25–34. Thus, declines since 2000 in heart disease mortality (albeit smaller than those seen in earlier periods) and in cancer mortality have been sufficient to outweigh increases in drug overdose and suicide deaths for men 35–54, but not for men 25–34.

Mortality trends since 2000 have varied substantially by race and ethnicity, as highlighted by Case and Deaton (2015, 2017). For example, the mortality rate among black prime-age men declined at a similar rate before and after 2000 (2.0 percent annually from 1980 to 2000 versus 1.8 percent annually from 2000 to 2016). By contrast, the mortality rate among white prime-age men increased by 0.3 percent annually after 2000 versus a 1.4 percent annual decline in the preceding 20 years. One of the most important drivers of this difference was the differential benefit from declining HIV/AIDS mortality. Black prime-age men were seven times more likely than white prime-age men in 2000 to die from HIV/AIDS (65 versus 9 per 100,000), and thus benefitted far more from the subsequent plunge in the HIV/AIDS mortality rate.

In addition, black men saw much smaller increases in the suicide rate (7 percent versus 32 percent for white men) and were much less likely to commit suicide initially. As a result, white men are now more than twice as likely as black men to commit suicide (31 versus 13 per 100,000). Finally, the death rates from (primarily drug- and alcohol-induced) accidents increased by five times as much among white prime-age men as among black men of the same age. Thus, even though black men have been more affected by the increase in the homicide rate since 2000, this change is dwarfed by trends in the combination of HIV/AIDS, drug- and alcohol-induced accidents, and suicides.

For Hispanic or Latino prime-age males, mortality rates fell by 1.2 percent annually from 2000 to 2016, while the corresponding rate among their non-Hispanic white counterparts rose by 0.7 percent annually.[4] As a result of these changes, mortality rates among white, non-Hispanic prime-age males are now more than 50 percent higher than among Hispanic prime-age males (320 versus 211 per 100,000).

The adverse mortality trend for white non-Hispanic men is stronger among those with low levels of education, Case and Deaton (2017) find that mortality rates have risen over the past two decades among prime-age white non-Hispanic men without a college degree, while holding steady or declining for their counterparts with a college degree. One challenge for assessing how mortality changes have

[4] The Centers for Disease Control did not collect data on ethnicity in 1980, so it is not possible to compare the 1980–2000 trends with the 2000–2016 trends for these two groups.

Table 3
Male Mortality Rates among White, Non-Hispanic Males Aged 25–54, by State Education Quartile, 2000 and 2016

State education quartile	*% without HS degree*	*Annual mortality rates per 100,000*		
		2000	*2016*	*% Change*
1	10%	257	265	+2.8%
2	13%	262	296	+12.9%
3	17%	275	329	+19.7%
4	23%	334	400	+19.6%

Source: Authors using data from the US Census, CDC Wonder.
Note: Changes listed may differ slightly from implied changes due to rounding.

differed by educational attainment is that the fraction of men who are high school dropouts has fallen substantially over time (as shown earlier in Table 1). Differential changes in mortality by education in recent decades thus could be driven by changes in the composition of individuals in each education group—that is, high school dropouts becoming more adversely selected—rather than by differential changes in each group's mortality rate. To address this issue, Bound, Geronimus, Rodriguez, and Waidmann (2014) categorize individuals by their rank in each year's educational attainment distribution and show that from 1990 to 2010 the life expectancy at age 25 of non-Hispanic white males in the bottom quartile of educational attainment rose by three years less than in the top three quartiles (a six-year increase versus a three-year increase). These findings imply that low-skilled men have fallen further behind high-skilled men with respect to this key measure of health.[5]

An alternative way to examine the relationship between education and mortality changes is to compare trends in geographic areas with lower levels of educational attainment with those with greater attainment. Consistent with the evidence from Bound et al. (2014), mortality rates among prime-age men rose by substantially more from 2000 to 2016 in states with low levels of education. Table 3 groups all 50 states and the District of Columbia into four quartiles based on the share of non-Hispanic white men aged 25–54 who were without a high school degree in 2000. As Table 3 shows, mortality rates were already higher in states with the most low-skilled men in 2000. This disparity increased significantly in the subsequent years, with mortality rising by just 3 percent in the state with higher levels of education such as California and Washington, DC, compared with a 20 percent increase in the states with lower levels of education such as West Virginia and Kentucky.[6]

Establishing whether there is a link between the rise in mortality for low-skilled men—particularly the increase in "deaths of despair"—and economic conditions

[5] Consistent with this finding, Chetty et al. (2016) find that life expectancy of 40-year old males in the highest income quartile increased by 2.6 years from 2001 to 2014 versus only 1.0 years among 40 year-old men in the lowest income quartile.
[6] For an in-depth analysis of mortality trends by state, see US Burden of Disease Collaborators (2018).

is an active area of research. Pierce and Schott (2016) find that the lowering of trade barriers with China around 2000 led to substantial increases in both suicide and opioid overdose mortality rates, with larger effects in trade-exposed counties with a higher share of less-educated men, reflecting the policy's larger employment effects on this group. This is consistent with Case and Deaton (2017), who find that increases in reports of pain (which is often treated with opioids) since the mid-1990s have occurred exclusively among the less educated. Ruhm (2018) argues that the rise in opioid use is more strongly related to drug access (the "drug environment") than to economic conditions. Krueger's (2017) surprising finding that nearly half of prime-age men who are out of the labor force take pain medication on a daily basis does not establish the direction of causality between employment and opioid use, but does underscore the importance of future research that might do so.

Morbidity, Self-Reported Health Measures, and Disability Insurance

Health measures other than mortality may capture health issues that are more prevalent among prime-age men and potentially more pertinent to labor force participation. On Table 4, we report values for a variety of measures frequently used in the literature on health and disability trends, including self-reported health, work-limiting disability, physical limitations, limitations in Activities of Daily Living (ADLs) or in Instrumental Activities of Daily Living (IADLs), and obesity, using data from the National Health Interview Survey (NHIS).

Health problems rise with age, as expected, although their incidence depends on the measure used. In 2000, the share of men without a high school degree who report themselves to be in fair or poor health triples from 6 percent at ages 25–34, to 18 percent at ages 45–54; the share reporting a work-limiting disability is similar. The same age pattern exists for the other health measures, but the share of men without a high school degree reporting ADL or IADL problems (1 to 3 percent) is much lower and the share reporting physical limitations (14 to 34 percent) or obesity (22 to 29 percent) is much higher.

There is a steep health gradient with respect to education—within each age group, the share in fair or poor health is roughly 2.5 times as large for men with a high school education or less than for men with some college or more. Men with less education are similarly more likely to report having a work-limiting disability, limitations in physical activity or ADLs/IADLs, and obesity, although the relative differences are somewhat smaller for physical limitations and obesity (about 1.3 times as large). Interestingly, for ADL/IADL difficulty, there is some indication that the education gradient is larger (on a relative basis) for young men than for older men.

Men's health, as captured by these measures, is getting worse over time. As seen in Table 4, the fraction of men reporting a health problem is higher in 2015 than in 2000 in nearly every case. On a relative basis, increases are greatest for ADL and IADL difficulties, which had the lowest values initially but saw increases of about

Table 4
Health Measures, Men Ages 25 to 54, 2000 and 2015

Health measure	*Age group*	*Education group*	*Share with condition*		*Change 2000–2015 (percentage points)*
			2000	2015	
Fair/Poor health	25–34	HS or less	5.8%	8.4%	2.6
		>HS	2.3%	3.7%	1.4
	35–44	HS or less	9.4%	12.0%	2.5
		>HS	3.7%	5.0%	1.3
	45–54	HS or less	17.7%	18.6%	0.9
		>HS	6.5%	7.7%	1.2
Work-limiting disability	25–34	HS or less	5.9%	8.3%	2.4
		>HS	2.6%	2.9%	0.3
	35–44	HS or less	8.9%	9.7%	0.7
		>HS	4.7%	3.8%	–0.9
	45–54	HS or less	15.1%	16.0%	0.9
		>HS	7.4%	7.5%	0.1
Physical limitations	25–34	HS or less	13.9%	18.8%	4.8
		>HS	11.2%	11.4%	0.2
	35–44	HS or less	21.3%	26.0%	4.7
		>HS	17.4%	18.4%	1.0
	45–54	HS or less	33.5%	38.0%	4.5
		>HS	24.5%	27.4%	2.9
ADL Difficulties	25–34	HS or less	0.6%	1.3%	0.7
		>HS	0.2%	0.3%	0.2
	35–44	HS or less	0.9%	1.6%	0.7
		>HS	0.4%	0.4%	0.0
	45-54	HS or less	1.4%	1.7%	0.3
		>HS	0.5%	0.9%	0.5
IADL Difficulties	25–34	HS or less	1.3%	2.4%	1.0
		>HS	0.3%	0.8%	0.5
	35–44	HS or less	1.7%	2.9%	1.2
		>HS	0.7%	0.8%	0.1
	45–54	HS or less	3.0%	3.7%	0.7
		>HS	1.1%	1.7%	0.6
Obesity	25–34	HS or less	21.8%	27.7%	5.9
		>HS	16.8%	22.6%	5.9
	35–44	HS or less	24.5%	35.3%	10.8
		>HS	19.5%	29.1%	9.7
	45–54	HS or less	28.7%	36.6%	7.9
		>HS	21.4%	32.5%	11.1

Source: Authors using data from the National Health Interview Survey (NHIS).
Note: HS is "high school." All measures rely on self-reported data. Data is reported for 2000 and 2015 only due to a NHIS redesign in 1997. Data are aggregated over a 3-year period (1999–2001 or 2014–2016) to minimize sampling variation; data are weighted to reflect population values. Fair/poor health is based on self-reported health. Work limitations, physical limitations, and ADL/IADL difficulties refer to the share reporting any level of difficulty with work, physical activity, or Activities of Daily Living/Instrumental Activities of Daily Living (ADL/IADL). For physical limitations, the nine physical activities include: walking a quarter mile, climbing ten steps, standing two hours, sitting two hours, stooping/bending/kneeling, reaching over one's head, grasping small objects, carrying ten pounds, and moving large objects. For ADLs, the six activities include: bathing, dressing, eating, toileting, getting around the house, and getting in and out of a bed or chair. The IADL measure is based on a single question about difficulty with routine needs such as such as everyday household chores, doing necessary business, shopping, or getting around for other purposes. Obesity is defined as having a Body Mass Index (BMI) over 30. Changes listed may differ slightly from implied changes due to rounding.

70 percent relative to baseline. For the other health measures, the increases are about 10 percent for work and physical limits and 30 to 40 percent for self-reported health and obesity. Using an earlier version of this data, Duggan and Imberman (2009) find a small increase in work and activity limitations among prime-age men from 1984 to 1996, while Martin and Schoeni (2014) show that the increase in limitations is partly explained by the rise in body mass index.

Although the recent changes represent sizeable increases over a relatively short period, the absolute share of the population reporting the most serious health problems, such as ADL difficulties, generally remains low. In comparing changes in mortality and morbidity, it is interesting to note that mortality improved from 1980 to 2000 but the trend reversed after 2000 for younger men and non-Hispanic less-educated whites, while morbidity has worsened continuously since 1980 and more so for the less educated.

The increases over time in reports of health problems are generally larger in absolute terms for less-educated men. On a relative basis, the pattern is less uniform. Even so, the larger absolute increases for less-educated men could contribute to divergences in labor force participation, although the changes in health reported here are significantly smaller than the changes in labor force participation (discussed in Binder and Bound in this symposium).

Receipt of disability benefits is also of interest, because it may be related to health through the medical eligibility requirement, yet also potentially subject to influence by economic factors and by political decisions about eligibility standards. Approximately 3 million prime-age men (4.5 percent of the total) currently receive Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) benefits, which provide cash benefits to individuals unable to engage in substantial gainful activity. Average monthly benefits among nonelderly adult males enrolled in the SSDI and SSI programs are approximately $1,300 and $600, respectively. SSDI and SSI recipients also typically qualify for health insurance through the federal Medicare and federal-state Medicaid programs, respectively. Individuals must have worked in at least five of the ten most recent years to be potentially eligible for SSDI benefits, while no work history is required for the means-tested SSI program.

The fraction of prime-age men receiving benefits from the Social Security Disability Insurance and Supplemental Security Income programs declined somewhat in the early 1980s and then rose steadily for decades (in this journal, Liebman 2015). For example, Table 5 shows that the fraction of men 45–54 receiving SSDI benefits fell from 4.1 percent to 3.4 percent in the first few years of the 1980s and then rose steadily to 5.7 percent by 2010. SSI enrollment followed a similar pattern over this period, with SSI receipt higher than SSDI receipt among those ages 25–34 and lower among those ages 35–54.

One driver of these changes in enrollment is changes in the medical eligibility criteria. The criteria for both Social Security Disability Insurance and Supplemental Security Income became much more stringent in the late 1970s and early 1980s and then much more lenient beginning in 1984. Before 1984, the most common conditions with which individuals qualified for SSDI benefits were circulatory conditions

Table 5
Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) Enrollment, Men Ages 25 to 54, 1980 to 2017

		Share of men enrolled				
Program	*Age group*	*1980*	*1984*	*2000*	*2010*	*2017*
SSDI	25–34	0.8%	0.8%	0.9%	1.2%	0.9%
	35–44	1.7%	1.8%	2.4%	2.5%	2.2%
	45–54	4.1%	3.4%	4.8%	5.7%	5.0%
SSI	25–34	0.9%	0.9%	1.6%	2.5%	2.7%
	35–44	0.8%	0.8%	1.7%	1.8%	1.8%
	45–54	1.4%	1.5%	2.0%	2.3%	2.5%

Source: Authors using data from Social Security Administration (for SSDI and SSI enrollment) and U.S. Bureau of the Census (for population for denominators of the rates).

(for example, heart attacks and stroke) and cancer. Then, with the 1984 changes, it became easier for individuals with relatively subjective conditions such as back pain and depression to qualify for the program and the award rates for these conditions increased substantially.

Economic factors also affect disability program enrollment. Over the past several decades, increases in wage inequality interacted with both programs' benefit formulas led to substantial increases in the share of low-skilled workers' wages that would be replaced by these programs (in this journal, Autor and Duggan 2006). Applications and awards for disability insurance fluctuate with the business cycle (Autor and Duggan 2003; Maestas, Mullen, and Strand 2018), suggesting that more permanent declines in economic opportunities for low-skilled men are likely to encourage more men in marginal health to apply for benefits.

Following 30 years of steady expansion, disability insurance enrollment started to decline in 2014, with the share of men receiving disability insurance in all three age groups lower in 2017 than seven years earlier. This decline is to some extent surprising, given the trends in mortality described above. The steadily improving economy and the tightening of the program's medical eligibility criteria appear to be the key factors driving this reduction in program enrollment (Li 2018). Related research has shown that SSDI benefit income reduces mortality among beneficiaries, suggesting that the tightening eligibility for this program may be contributing to the recent mortality increases described above (Gelber, Moore, and Strand 2018).

Enrollment in Social Security Disability Insurance and Supplemental Security Income varies substantially across and within states. Perhaps not surprisingly, disability enrollment is highly correlated with the state-level mortality rates among prime-age men mentioned above and has risen significantly more in those parts of the country hit harder by adverse economic shocks in recent decades (Autor, Dorn, and Hansen 2016). It is also substantially greater among those with less education. An examination of data from the March 2018 Current Population Survey reveals that prime-age men with only a high school degree or less are more than five times more likely to receive SSDI or SSI disability benefits than their counterparts with a four-year college degree (7.0 percent versus 1.3 percent). Those with some college or an associate's degree are between these two extremes (at 3.4 percent).[7]

Turning to the link between economic conditions and health, recent research suggests that employment can have a substantial positive effect on men's health. Fitzpatrick and Moore (2017) show that there is a substantial increase in male mortality rates at the age of 62, when more than one-third of men claim Social Security retirement benefits and many leave the labor force. The authors conclude that the decline in employment is a key contributor to the mortality increase and note that the relationship for women is much smaller. This suggests that employment declines among low-skilled men likely contributed to the recent increases in mortality, and potentially also to the increases in health issues and disability program enrollment that emerged even earlier.

Related research has shown that income has a significant positive effect on health, and that these effects are much larger among lower-income beneficiaries. Gelber, Moore, and Strand (2018) use features of the Social Security Disability Insurance formula to estimate the elasticity of mortality with respect to income among new SSDI recipients. Their findings demonstrate that mortality rates fall substantially for lower-income individuals who receive higher SSDI benefits, but that the corresponding effects for high-income individuals are much smaller. This strongly suggests that the decline in earnings among lower-skilled men has contributed to the increases in mortality highlighted above.[8] It also suggests there is a causal connection between the well-documented rise in earnings inequality in recent decades and the rising inequality in life expectancy documented by Chetty et al. (2016).

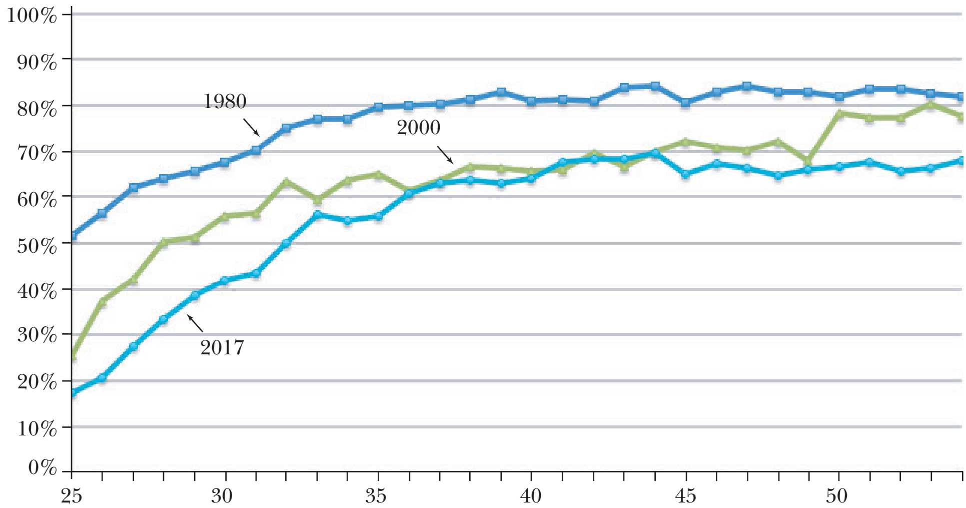
Marital and Family Status

The marital patterns of men were remarkably stable during the century leading up to 1980 (as discussed in this journal by Stevenson and Wolfers 2007). In the decades since, however, men have been marrying later, as seen in Figure 1. While half of 25 year-old men were married in 1980, less than 20 percent of this group was married in 2017. Figure 1 reports the share of men *currently* married, which is lower than the share of men *ever* married because it excludes currently unmarried men who were previously married. However, separate tabulations of data from the Current Population Survey confirm that the share that has never been married has

[7] Autor and Duggan (2003) demonstrate that the increase in enrollment in Social Security Disability Insurance during the 1980s and 1990s was much greater among those with only a high school degree or less.

[8] Evans and Moore (2011) find that additional income from transfer programs can actually lead to increases in mortality. However, these authors are looking at very short-term changes in income, comparing mortality changes in the days after cash benefit receipt or after receiving tax rebates. The Social Security Disability Insurance evidence is more relevant for the trends highlighted in this paper since it represents an increase in permanent income rather than simply a short-term increase in mortality in the few days after receiving a check.

Figure 1
Share of Men Married by Age and Year



Source: Authors using data from the Current Population Survey (CPS).

also risen over time: for example, among men ages 45–54, this share rose from 6 percent for men in 1980 to 16 percent in 2017.

Table 6 examines how men's marital status has changed over time across race and education groups, focusing on the share married at ages 40–44. Marriage rates for whites in 1980 were much higher than for blacks, but whites experienced a somewhat larger decline over time; Hispanics had the highest marriage rates initially and also the largest drop. The differences by education are striking. Marriage rates were nearly identical across all education groups in 1980, but by 2017 had dropped by roughly 20 percentage points among all groups with less than a college education, while dropping only slightly for college graduates. Men with less education are now less likely to ever get married, more likely to get divorced, and less likely to remarry than their counterparts with more education (Aughinbaugh, Robles, and Sun 2013).

Many of the theories that have been put forth to explain the changes in marriage rates focus on women: for example, greater access to contraception, greater opportunities for women in the labor market, and a rise in welfare support for single mothers (Loughran and Zissimopoulos 2009). However, several hypotheses relate to men's economic status, including rising wage inequality (Loughran 2002) and a decline in the availability of marriageable men (Brien 1997). The latter may be important in explaining black–white differences, since black men face a higher risk of incarceration and unemployment, among other differences. Being married may also affect men's earnings (Ahituv and Lerman 2007), complicating efforts to estimate how men's economic status affects marriage decisions.

Trends in having children mirror those in marriage, with men today having children later and being less likely ever to have them as compared to earlier cohorts.

Table 6
Marriage Rates at Ages 40 to 44, 1980 to 2017

	Share of men married at ages 40 to 44				
Group	*1980*	*2000*	*2017*	*Change 1980–2000 (percentage points)*	*Change 2000–2017 (percentage points)*
By Race					
White	84%	69%	69%	–15	0
Black	63%	52%	51%	–11	–1
Hispanic	87%	74%	64%	–13	–10
By Education					
<HS	80%	65%	60%	–16	–4
HS graduate	82%	62%	59%	–19	–4
Some college	82%	67%	64%	–16	–3
College	85%	77%	79%	–8	+2

Source: Authors using data from the Current Population Survey (CPS).
Note: Changes listed may differ slightly from implied changes due to rounding.

At ages 25–34, the share of men living with children (including own and stepchildren) was about 50 percent in 1980 and 30 percent in 2017. While the difference over time is smaller at older ages—for example, about 10 percentage points for men in the 45–54 age bracket—men at every age are less likely to be living with children in 2017 than they were in 1980.

Young adults are also increasingly likely to be living with their parents, with nearly 20 percent of adults ages 25–34 doing so in 2015 (Vesta 2017). Relative to other young adults, those living at home were more likely to be male and less likely to be employed or to have a college degree. There are substantial geographic differences in living with parents that appear related at least in part to differences in cost of living, with particularly high rates of living at home in high-cost states such as New York and Connecticut.

Some recent research suggests a link between changing economic conditions and family structure. Autor, Dorn, and Hanson (2018) show that areas with trade-induced declines in manufacturing employment experienced increases in idleness among prime-age men and that these men were also less likely to marry or to have children. Correspondingly, these same areas experienced increases in the proportion of children living with just one parent and in the fraction of children living below the poverty line. Similarly, Gould (2018) finds that the decline in manufacturing employment led to lower marriage rates as well as a larger gap in marriage rates by race and education.

Incarceration

Changes in a group's incarceration rate are likely to have current and future effects on their earnings, health, education, and family formation. After all, a person in jail or prison is typically unable to have a job, to pursue educational opportunities, or to spend time with family members or friends. Even after release from prison or jail, there may be long-term effects that reduce earnings potential or health.

The US incarcerated population grew rapidly in the 1980s and 1990s, with the number of people in federal or state prison or in county or city jails jumping from about 500,000 in 1980 to nearly 2 million in 2000 (US Bureau of Justice Statistics 2018). Many factors contributed to the increase, including rising crime rates. However, rising arrest rates (especially for drug-related crimes), increased probabilities of incarceration conditional on arrest, and longer sentence lengths were far more important than the rise in crime (National Research Council 2014). Additionally the reduction in the capacity of institutions for those with mental illness explained 4 to 7 percent of the increase in the prison population from 1980 to 2000 (Raphael and Stoll 2013).

Prime-age men accounted for more than 70 percent of all incarcerated individuals during this period. In general, prime-age men account for a larger share of the incarcerated population than of all criminals, because those who commit crimes in their late teens or early 20s may remain incarcerated for many years after that. For example, while men 25–54 accounted for 74 percent of male prisoners in 2000, they accounted for just 46 percent of males arrested for murder and 55 percent of male murder victims (Beck and Karberg 2001; FBI 2001). In contrast, males 18–24 accounted for 37 percent of murder offenders in that same year but just 19 percent of incarcerated males.

As shown in Table 7, the rise in the incarceration rate was especially high among younger men from 1980 to 2000, with the fraction of men aged 25–34 in prison or jail rising from 1.3 percent in 1980 to 3.5 percent by 2000. Men in the 35–44 and 45–54 age ranges also became much more likely to be incarcerated during this same 20-year period. Weighting each of the three age groups equally (to avoid any effects from a shifting age distribution), the incarceration rate of prime-age men increased by approximately 160 percent over 1980 to 2000, from 0.9 percent to 2.3 percent.

Following this period of rapid growth, the incarcerated population grew much more slowly starting in the late 1990s, peaked in 2008, and has declined modestly over the past decade. This changing trend was primarily driven by the decline in crime that began in the mid-1990s and continued through 2014: for example, the nation's violent crime rate fell by almost 50 percent (from 714 to 362 violent crimes per 100,000 residents) over this period (FBI 2015). The resulting decline in incarceration was concentrated among younger adults. The fraction of men aged 25–34 in prison or jail fell from 3.5 percent in 2000 to 2.8 percent by 2016 (Table 7). In contrast, incarceration rates continued to increase for older men in the 35–44 and 45–54 age ranges. This difference likely reflects the fact that prisoners in their 40s and early 50s were more likely to have committed their crimes as young adults, before crime rates started to fall.

The importance of recent changes in the incarceration rate differed substantially by race. Most notably, black men in the 25–34 age-range saw their incarceration rate fall from 12.8 percent in 2000 to 7.4 percent by 2016, following

Table 7
Male Incarceration Rate by Race and Ethnicity, Ages 25 to 54, 1980 to 2016

		Incarceration rate per 100,000 men				
Race/ Ethnicity	Age group	1980	2000	2016	Change 1980–2000 (percentage points)	Change 2000–2016 (percentage points)
All	25–34	1.3%	3.5%	2.8%	2.2	–0.7
	35–44	0.9%	2.3%	2.6%	1.4	0.4
	45–54	0.4%	1.0%	1.8%	0.6	0.8
	25–54[a]	0.9%	2.3%	2.4%	1.4	0.1
White	25–34	0.7%	1.7%	1.6%	1.0	–0.2
	35–44	0.5%	1.3%	1.6%	0.7	0.3
	45–54	0.3%	0.6%	1.1%	0.4	0.5
	25–54[a]	0.5%	1.2%	1.4%	0.7	0.2
Black	25–34	5.5%	12.8%	7.4%	7.2	–5.4
	35–44	3.7%	8.6%	7.4%	4.9	–1.2
	45–54	1.6%	3.7%	5.0%	2.1	1.3
	25–54[a]	3.6%	8.4%	6.6%	4.7	–1.8
Hispanic	25–34	–	3.9%	3.1%	–	–0.9
	35–44	–	2.9%	2.8%	–	–0.1
	45–54	–	1.6%	1.9%	–	0.3
	25–54[a]	–	2.8%	2.6%	–	–0.2

Source: Authors using data from Bureau of Justice Statistics and US Bureau of the Census. See Data Appendix for details.
Note: Changes listed may differ slightly from implied changes due to rounding.
[a] The incarceration rate for men ages 25–54 is computed as a simple average of the rate for men ages 25–34, 35–44, and 45–54, in order to minimize the effect of changing age distribution of the population over time.

an even larger (in magnitude) increase from 1980 to 2000. The corresponding reduction from 2000 to 2016 among young white men was minimal. Incarceration rates also fell for black men ages 35–44, while rising for older black men. As was the case for falling HIV/AIDS mortality, black men stood to gain more from the declining incarceration rate in recent years because of their higher baseline rate of incarceration. Finally, while data for Hispanic men are not available for 1980, the incarceration rates for this group followed a generally similar pattern, with large reductions for younger men and increases for older men since 2000.

The connection between economic factors and incarceration is not simple. On the one hand, neither the rise nor the drop in incarceration rate appears to be primarily driven by economic factors. As noted, the rise in incarceration was largely due to changes in criminal justice policy and the fall to declining crime rates. Levitt (2004) does not point to economic conditions as a key factor in explaining the decline in crime, though this does not rule out the possibility of some relationship—for example, Raphael and Winter-Ember (2001) find a link between unemployment and property crime. But falling incarceration rates may affect men's future economic outcomes.

Kling (2006) leverages plausibly exogenous variation across judges in the stringency of their sentencing and—perhaps surprisingly—finds no substantial evidence that longer prison sentences have a negative effect on employment or earnings after release. More recent evidence using a similar methodology, however, suggests that there is a significant negative effect of incarceration on future employment (Dobbie, Golden, and Yang 2018). Furthermore, Doleac and Hanson (2016) find that "ban the box" policies, which limit employers' ability to ask about criminal background checks in the hiring process, decrease employment for young, low-skilled black and Hispanic men, suggesting that employers may prefer not to hire ex-offenders (who are disproportionately represented among these groups). As young black and Hispanic men have historically experienced higher incarceration rates, they face larger potential employment gains from the recent decline in incarceration; by contrast, effects for white men would be expected to be smaller.

Discussion

The lives of prime-age men have changed in important ways in recent decades. A number of explanations have been put forward to explain the declining labor force participation and wages of men, trends that are much stronger among those with less education (Council of Economic Advisors 2016). Possible causes include demand-side factors like skill-biased technological change (Acemoglu and Autor 2010) and globalization (Autor, Dorn, and Hanson 2013), as well as supply-side factors such as rising spousal employment, greater use of the Social Security Disability Insurance program (Autor and Duggan 2003), and rising utility of leisure due to improvements in video game technology (Aguiar, Bils, Charles, and Hurst 2017).

In this essay, we have focused on a number of dimensions that affect the well-being of prime-age males in a direct sense—mortality and morbidity, marriage and children, education, and incarceration rates. The labor market has not been kind to low-skilled men in the last few decades, and patterns in the outcome measures that we examine suggest that these men have suffered from a cluster of other problems as well.

We document slowing mortality gains and rising morbidity and disability program enrollment that are stronger for less-educated males and in states with lower levels of education. Declines in marriage are also concentrated among non-college-educated men. Examining differences by race, we find bigger declines in both mortality and incarceration since 2000 for blacks and Hispanics, groups that have also experienced more rapid educational gains. The narrowing racial gap in outcomes—for example, the fact that blacks did not experience the same increase in suicides and overdose deaths as whites—is consistent with a beneficial effect of education in a weakening economic climate, although these trends also reflect the effect of other factors (such as advancements in HIV treatment and declining crime rates) that disproportionately benefitted these groups. Importantly, a growing number of studies offer compelling evidence that rising import competition and other economic events that reduced opportunities for low-skilled men had direct adverse effects on their health and well-being.

As researchers continue to explore the causes and consequences of the trends highlighted in this paper, we call attention to perhaps the most significant change among prime-age men in recent decades. In 1980, fully 45 percent of prime-age men reported in the Bureau of Labor Statistics' monthly Current Population Survey said that they had previously served in the military. This number steadily declined during the next 36 years and stood at just 10 percent by 2016 in this same survey. Much of the economics literature has examined the effect of military service by using plausibly exogenous variation in the likelihood of service driven by one's draft lottery number (Angrist 1990). This research has tended to find quite modest long-term effects of military service on employment, earnings, and health status (for example, Angrist, Chen, and Frandsen 2010; Angrist, Chen, and Song 2011).[9] However, these studies are unable to capture the peer effects or general equilibrium effects of military service. Recent research has suggested substantial gains to cognitive and noncognitive skills stemming from military service (Spiro, Stetterson, and Aldwin 2015) and associated benefits such as the GI bill. Overall, we see a strong need for further work to investigate how changing economic opportunities, declines in military service, and other factors are contributing to or cushioning the problems of low-skilled prime-age men.

■ *We are grateful to Nicole Chen, Sasha Dierauf, Grace Hong, Gina Li, and Olivia Martin for their outstanding research assistance.*

[9] While Angrist (1990) finds that Vietnam-era service has a negative effect on the earnings of white veterans in the short-to-medium term post-service, Angrist (1998) finds a positive effect of post-Vietnam service on black veterans over a comparable time period.

References

Acemoglu, Daron, and David Autor. 2010. "Skills, Tasks, and Technologies: Implications for Employment and Earnings." Chap. 4 in *Handbook of Labor Economics, Volume 4B*, edited by Orley Ashenfelter and David Card. New York: Elsevier.

Aguiar, Mark, Mark Bils, Kerwin Kofi Charles, and Erik Hurst. 2017. "Leisure Luxuries and the Labor Supply of Young Men." NBER Working Paper 23552.

Ahituv, Avner, and Robert I. Lerman. 2007. "How Do Marital Status, Work Effort, and Wage Rates Interact?" *Demography* 44(3): 623–47.

Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80(3): 313–36.

Angrist, Joshua D. 1998. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants." *Econometrica* 66(2): 249–88.

Angrist, Joshua D., Stacey H. Chen, and Brigham R. Frandsen. 2010. "Did Vietnam Veterans Get Sicker in the 1990s? The Complicated Effects of Military Service on Self-Reported Health." *Journal of Public Economics* 94(11–12): 824–37.

Angrist, Joshua D., Stacey H. Chen, and Jae Song. 2011. "Long-Term Consequences of Vietnam-Era Conscription: New Estimates Using Social Security Data." *American Economic Review* 101(3): 334–38.

Aughinbaugh, Alison, Omar Robles, and Hugette Sun. 2013. "Marriage and Divorce: Patterns by Gender, Race, and Educational Attainment." *Monthly Labor Review*, October.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.

Autor, David H., David Dorn, and Gordon H. Hanson. 2016. "The China Shock: Learning from Labor Market Adjustment to Large Changes in Trade." *Annual Review of Economics* 8: 205–40.

Autor, David H., David Dorn, and Gordon H. Hanson. 2018. "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Men." CESifo Working Paper 7010.

Autor, David H., and Mark G. Duggan. 2006. "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding." *Journal of Economic Perspectives* 20(3): 71–96.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90(2): 300–23.

Beck, Allen J., and Jennifer C. Karberg. 2001. "Prison and Jail Inmates at Midyear 2000." Bureau of Justice Statistics Bulletin, US Department of Justice, March. NCJ 185989. https://www.bjs.gov/content/pub/pdf/pjim00.pdf.

Bound, John, Arline Geronimus, Javier Rodriguez, and Timothy Waidmann. 2014. "The Implications of Differential Mortality among the Aged by Socioeconomic Status." Prepared for the 16th Annual Joint Conference of the Retirement Research Consortium.

Brien, Michael J. 1997. "Racial Differences in Marriage and the Role of Marriage Markets." *Journal of Human Resources* 32(4): 741–78.

Card, David, and Thomas Lemieux. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis." *Quarterly Journal of Economics* 116(2): 705–746.

Case, Anne, and Angus Deaton. 2015. "Rising Morbidity and Mortality in Midlife among White Non-Hispanic Americans in the 21st Century." *PNAS* 112(49): 15078–083.

Case, Anne, and Angus Deaton. 2017. "Mortality and Morbidity in the 21st Century." *Brookings Papers on Economic Activity*, Spring, pp. 397–443.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergerson, and David Cutler. 2016. "The Association between Income and Life Expectancy in the United States, 2001–2014." *JAMA* 315(16): 1750–66.

Council of Economics Advisors. 2016. "The Long-Term Decline in Prime-Age Male Labor Force Participation."

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108(2): 201–40.

Doleac, Jennifer L., and Benjamin Hansen. 2016. "Does 'Ban the Box' Help or Hurt Low-Skilled Workers? Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden." NBER Working Paper 22469.

Duggan, Mark, and Scott A. Imberman. 2009. "Why Are the Disability Rolls Skyrocketing?" Chap. 11 in *Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly*. University of Chicago Press.

Evans, William N., and Timothy J. Moore. 2011. "The Short-Term Mortality Consequences of Income Receipt." *Journal of Public Economics* 95(11–12): 1410–24.

Federal Bureau of Investigation (FBI). Various years. Crime in the United States. https://ucr.fbi.gov/crime-in-the-u.s (accessed on December 5, 2018).

Fitzpatrick, Maria D., and Timothy J. Moore. 2017. "The Mortality Effects of Retirement: Evidence from Social Security Eligibility at Age 62." *Journal of Public Economics* 157: 121–37.

Fontenot, Kayla R., Semega, Jessica L., and Melissa A. Kollar. 2018. *Income and Poverty in the United States: 2017.* Current Population Report P60-263, September. https://www.census.gov/content/dam/Census/library/publications/2018/demo/p60-263.pdf.

Gelber, Alexander, Timothy Moore, and Alexander Strand. 2018. "Disability Insurance Income Saves Lives." Unpublished paper.

Gelman, Andrew, and Jonathan Auerbach. 2016. "Age-Aggregation Bias in Mortality Trends." *PNAS* 113(7): E816–17.

Gould, Eric D. 2018. "Torn Apart? The Impact of Manufacturing Employment Decline on Black and White Americans." IZA Discussion Papers 11614.

Jacob, Brian A. 2002. "Where the Boys Aren't: Non-Cognitive Skills, Returns to School, and the Gender Gap in Higher Education." *Economics of Education Review* 21(6): 589–98.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96(3): 863–78.

Krueger, Alan B. 2017. "Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate." Brookings Papers on Economic Activity, Fall.

Levitt, Steven D. 2004. "Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not." *Journal of Economic Perspectives* 18(1): 163–90.

Li, Zhe. 2018. "Trends in Social Security Disability Insurance Enrollement." CRS Report, R45419, Congressional Research Service. https://fas.org/sgp/crs/misc/R45419.pdf.

Liebman, Jeffrey B. 2015. "Understanding the Increase in Disability Insurance Benefit Receipt in the United States." *Journal of Economic Perspectives* 29(2): 123–50.

Loughran, David S. 2002. "The Effect of Rising Male Wage Inequality on Female Age at First Marriage." *Review of Economics and Statistics* 84(2): 237–50.

Loughran, David S., and Julie M. Zissimopoulos. 2009. "Why Wait? The Effect of Marriage and Childbearing on the Wages of Men and Women." *Journal of Human Resources* 44(2): 326–49.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2018. "The Effect of Economic Conditions on the Disability Insurance Program: Evidence from the Great Recession." NBER Working Paper 25338.

Martin, Linda G., and Robert F. Schoeni. 2014. "Trends in Disability and Related Chronic Conditions among the Forty-and-Over Population: 1997–2010." *Disability Health Journal* 7(1 Supplement): S4–S14.

National Research Council. 2014. *The Growth of Incarceration in the United States: Exploring Causes and Consequences.* National Academies Press.

Pierce, Justin R., and Peter K. Schott. 2016. "Trade Liberalization and Mortality: Evidence from U.S. Counties." NBER Working Paper 22849.

Raphael, Steven, and Michael A. Stoll. 2013. "Assessing the Contribution of the Deinstitutionalization of the Mentally Ill to Growth in the U.S. Incarceration Rate." *Journal of Legal Studies* 42(1): 187–222.

Raphael, Steven, and Rudolf Winter-Ember. 2001. "Identifying the Effect of Unemployment on Crime." *Journal of Law and Economics* 44(1): 259–83.

Ruhm, Christopher J. 2018. "Deaths of Despair or Drug Problems?" NBER Working Paper 24188.

Ryan, Camille L., and Kurt Bauman. 2016. "Educational Attainment in the United States: 2015." Current Population Reports P20-578, United States Census Bureau.

Spiro, Avron, III, Richard A. Settersten, and Carolyn M. Aldwin. 2015. "Long-Term Outcomes of Military Service in Aging and the Life Course: A Positive Re-Envisioning." *Gerontologist* 56(1): 5–13.

Stevenson, Betsey, and Justin Wolfers. 2007. "Marriage and Divorce: Changes and Their Driving Forces." *Journal of Economic Perspectives* 21(2): 27–52.

US Burden of Disease Collaborators. 2018. "The State of US Health, 1990–2016: Burden of Diseases, Injuries, and Risk Factors Among US States." *JAMA* 319(14): 1444–72.

US Bureau of Justice Statistics. 2018. Key Statistic: Total Correctional Population. Available at: https://www.bjs.gov/index.cfm?ty=kfdetail&iid=487. Accessed on November 27, 2018.

Vesta, Jonathan. 2017. "The Changing Economics and Demographics of Young Adulthood: 1975–2016." Current Population Reports P20-579, United States Census Bureau.

Western, Bruce, and Becky Pettit. 2000. "Incarceration and Racial Inequality in Men's Employment." *Industrial and Labor Relations Review* 54(1): 3–16.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 211–228

The Tenuous Attachments of Working-Class Men

Kathryn Edin, Timothy Nelson, Andrew Cherlin, and Robert Francis

Work, family, and religion have traditionally played an important role in furnishing working-class Americans with economic resources, moral guidance, and opportunities for civic engagement (Cherlin 2009; McLanahan 2004; Verba, Schlozman, and Brady 1995; Wuthnow 2002). Ongoing attachments to work, family, and religion connected working-class men to social bonds and defined identities that kept them in the formal labor market and forestalled health problems. Conversely, precarious attachments to these key social institutions, we argue, may now dilute their power to shepherd and shift men's trajectories and may place them at risk of a host of negative outcomes. This is in line with sociologist Emile Durkheim's seminal study *Suicide* (1897 [1997]), which argued that "anomie," or normlessness, could explain variations in suicide rates across countries and over time.

In this essay, we explore how working-class men describe their attachments to work, family, and religion. We draw upon in-depth, life history interviews conducted in four metropolitan areas with racially and ethnically diverse groups of working-class men with a high school diploma but no four-year college degree. Between 2000 and 2013, we deployed heterogeneous sampling techniques in the

■ *Kathryn Edin is Professor of Sociology and Public Affairs, Woodrow Wilson School, and Timothy Nelson is Lecturer in the Department of Sociology, both at Princeton University, Princeton, New Jersey. Andrew Cherlin is Benjamin H. Griswold III Professor of Public Policy and Robert Francis is a PhD candidate in Sociology, both at Johns Hopkins University, Baltimore, Maryland. Their email addresses are kedin@princeton.edu, tjnelson@princeton.edu, cherlin@jhu.edu, and rfranc15@jhu.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.211 doi=10.1257/jep.33.2.211

black and white working-class neighborhoods of four metropolitan areas: Boston, Massachusetts; Charleston, South Carolina; Chicago, Illinois; and the Philadelphia/Camden area of Pennsylvania and New Jersey. We placed fliers in public places, sought referrals from a variety of grassroots organizations, and engaged in street sampling (approaching men on commercial streets and transit stops during daylight hours). We invited each of these men to refer up to two men to our study.

Because we were interested in these men's family ties, we screened to ensure that each respondent had at least one minor child. In Charleston and Philadelphia/Camden, we limited our sample to men with at least one child who could potentially have made the respondent subject to a child support order, because he was neither married to the child's mother nor living with her. We interviewed roughly even numbers of black and white men in each site for a total of 107 respondents.

We spoke at length with each respondent at least once, but usually twice. Interviews ranged from 90 minutes to three hours. All conversations were transcribed verbatim and coded using MaxQDA, a software program that is useful for identifying and systematically examining themes in qualitative data. We sorted men's narratives into codes capturing information relevant to prior constructs, as well as themes that emerged inductively from the transcripts.

In the first three sections of this paper, we describe the pattern of tenuous connections we found to key social institutions of work, family, and religion among the working-class men with whom we spoke. Although others have made similar arguments (Putnam 2015; Wilcox, Wolfinger, and Stokes 2015), we provide new evidence. Unlike past research, however, we show that working-class men are not simply reacting to changes in the economy, family norms, or religious organizations. Rather, they are attempting to renegotiate their relationships to these institutions by attempting to construct autonomous, generative selves. For example, these men's desire for autonomy in jobs seems rooted in their rejection of the monotony and limited autonomy that their fathers and grandfathers experienced in the workplace, along with a new ethos of self-expression (Cherlin 2014). Similarly, these working-class men focus on their ties to their children even when they have little relationship with the children's mothers, and they seek spiritual fulfillment even though they disdain organized religion. The drive toward generativity, by which we mean a desire to guide and nurture the next generation (Erikson 1963), is often rooted in past trauma often deriving from their family of origin. Many say that "giving back" in ways that they believe can make the world a better place is a way to redeem their own past as well as protect and nurture the next generation.

In sum, these working-class men show both a detachment from institutions and an engagement with more autonomous forms of work, childrearing, and spirituality, often with an emphasis on generativity. Autonomy refers to independent action in pursuit of personal growth and development. Personal growth has come to be highly valued among middle class Americans (Bellah, Madsen, Sullivan, Swidler, and Tipton 1985), but until recently has not been associated with the working class. The emphasis on activities directed toward personal growth among the working class that we and others (Silva 2013) found surprised us, as past scholarship typically assumed that such forms of action would usually only be found among those so materially comfortable that they needn't spend time worrying about their economic circumstances (Inglehart 1977). Subsequent to Eric Erikson's definition, which emphasized guidance and care of the next generation, researchers have expanded the concept of generativity to include "being a responsible citizen and a contributing member of a community" (McAdams, Hart, and Maruna 1998, p. 7). In other words, generativity is a special type of autonomous action, one directed at encouraging the growth and development not of oneself but of persons one cares about and knows well, such as one's children, as well as those in the community that need care and protection, such as the youth in one's neighborhood.

Our primary goal is to show that in order to comprehend these men's lives, we must consider both the unmaking and remaking aspects of their stories. We then turn to a discussion of the extent to which this autonomous and generative self is also a haphazard self, which may be aligned with counterproductive behaviors. As a secondary aim, we discuss racial and ethnic differences in what have been called "deaths of despair": the recent rise in mortality among whites with no more than a high school diploma due to suicides, drug overdoses, and alcohol-related liver failure (Case and Deaton 2015, 2017; see also Coile and Duggan, this volume). Working-class adults often make comparisons between their own and their parents' standards of living when their parents were their age, as shown in our in-depth interviews. However, drawing both on our interviews and representative survey data, we find that this comparison often leads to more negative assessments among whites than minorities (for both men and women in survey data). The reason is that non-college-educated whites are often comparing themselves to a generation that they feel had more opportunities than they have, whereas many blacks and Hispanics are more often comparing themselves to a generation that, in their view, had fewer opportunities.

Our interview methodology has both weaknesses and strengths. First, we are not drawing on a representative sample. Our interview subjects do not include working-class men without children, or from smaller cities and rural areas, or from the western or south-central regions of the United States. Moreover, men in our sample are more disadvantaged than a simple random sample of men with a high school degree but no college diploma, in part because they were all living in cities where many traditional working-class neighborhoods were in decline. However, an advantage of our approach is that it allows us to explore complex questions in a rich and granular way that allows unanticipated results to emerge. It is also useful for identifying processes and mechanisms that may not be captured in surveys or administrative data sources. Finally, in-depth qualitative interviews allow researchers to situate specific actions and attitudes within the larger context of respondent's lives.

The autonomous, generative identity we describe here can be seen in part as a way in which working-class men have reacted to structural changes in the labor market. Yet the way people describe their perceptions and aspirations will also have further effects on their behavior. We view the hypotheses advanced in this paper, derived from these interviews, as starting points worth further exploration by social scientists, not as definitive evidence.

Work

Studies of the relationship between working-class men and their jobs have traditionally highlighted claims that much of the identity of working-class men is heavily influenced by the nature of the job itself, from the living standard it permitted, and from the authority it allowed them to wield at home. For example, qualitative studies by Michele Lamont (2000) and Paul Willis (1977), as well as Melvin Kohn's surveys (1969) of white working-class men in prior generations, showed that they often claimed identity from their capacity for hard work at demanding, repetitive tasks—what Lamont calls "the disciplined self." In-depth interviews with blue-collar workers by Komarovsky (1964), Rubin (1976), and Halle (1984), conducted in the late 1950s through the early 1980s, call into question the idea that dignity always derived merely from the job itself—especially for those on the lower rungs—rather it may have also stemmed from the level of consumption the job afforded, a family wage that allowed men to be the sole or primary breadwinners. Salaries were usually sufficient for workers to purchase homes plus some luxuries; among those studied by Rubin (1976, p. 199) in the late 1970s, fully one-quarter owned a travel camper or boat. These earnings, in turn, allowed them to exercise considerable authority within the household, a further source of respect.

When describing their own fathers and grandfathers, our respondents often echoed the portrait offered by these older ethnographic studies. In contrast, the identity they themselves drew from their movements in and out of unskilled and semi-skilled jobs was more tenuous (Coglianese 2018). For example, Bernard had earned an HVAC (heating, ventilation, and air-conditioning systems) certification eight years earlier, but hadn't found employment in the field. Since then, he had worked as a security guard in a strip mall "just to pay the bills. I'm not really [into] law enforcement."

The successful performance of the disciplined self requires not only a certain kind of attitude, but also the availability of a certain type of job—manual occupations that provide the opportunity for stable work with good wages (Cherlin 2014). For the current generation of men, such jobs, plus the family wage and familial authority they allowed, have been in short supply (Autor, Katz, and Kearney 2006).

Given the tenuous nature of the work available to them, a good number of the men we interviewed had tried to build expertise, and gain the attendant certifications, in several different occupations, believing that "[I have to] hedge my bets" in order to remain employed. Iraq veteran Demario told us that he aspired to have "at least three options for … income. And if one doesn't work out, then I could have the other two as a fallback." Accordingly, after leaving the Army, he used his Veteran's Administration benefits to enroll in a diesel mechanics program and, more recently, had earned certification as a barber. When we interviewed him, he had just applied to a "visual production" program offered by a nearby state college. This strategy may be far from frivolous, as men sometimes obtain training in specific trades—like Bernard's HVAC certification—yet cannot find employment in that profession. However, hedging bets can also saddle men with thousands of dollars in student loan debt for training they are not using, while finding themselves actually working as security guards (as Bernard did), in retail, or unskilled manual labor.

Several men in our sample were working off-the-books, sometimes for extended periods. When Jeff quit a grueling factory job that required him to work twelve-hour shifts seven days a week and his felony conviction made it hard to find another position, he worked sporadic odd jobs for family and friends and provided childcare for his sister, who had two boys. Since 11th grade, 28-year-old Arthur had worked under-the-table as a handyman for a property manager. Ken had held a similar position with a man he called "my best friend, my landlord, and my boss." For 15 years, he had painted apartments between tenants and served as an unlicensed carpenter, plumber, and electrician on the landlord's properties, all for cash.

A few men recalled periods when they were younger that reflected the current stereotype of the 20-something in his mother's basement playing video games. After high school and before he became a father, Rick went on a "surfing safari" with a friend, traveling across Mexico and California in a VW van for almost four years, subsisting on odd jobs and on funds sent from his well-off suburban family. "We were totally hippied out. Patchouli oil, the whole nine yards," he recalls. In his 20s, Kirk moved back in with his mother, joined a "social club" of owners of classic Pontiac Grand Prix cars, and spent nearly two years "just enjoying the time off, hanging with my buddies." This sojourn ended abruptly when his girlfriend got pregnant. Feeling a new sense of responsibility, he quickly got a job, and secured an apartment for the soon-to-be threesome. None of the men in our sample were currently pursuing leisure full-time, perhaps because, like Rick and Kirk, they were all now fathers.

Other men were in and out of formal sector work while avidly cultivating "side bets"—entrepreneurial hobbies and informal occupations that they hoped to "take commercial" eventually. These took a wide variety of forms. For example, several worked as weekend DJs and hoped to open their own recording studios. Some designed tattoos while planning to forge a career as a tattoo artist, or to open a tattoo parlor. Several penned novels or self-help books and hoped to land a commercial publisher. Gene, who had worked in the past as a private investigator, told us he had interviewed dozens of others in the field about their experiences investigating infidelity. Based on these interviews, he had compiled "a complete profile of the American infidel … Everything that's in that profile is everything you need to survive a marriage." He insists he will "publish it someday." Jeremy, who was inspired to write by a high school internship at *Boston Globe,* self-published a novel when he was 19 with the hope of paying for college with the proceeds. While that first venture had earned little profit, he had sent several additional manuscripts to an Amazon contest for aspiring authors in hopes of attracting the attention of a publisher.[1]

Through these side bets, the men that we interviewed were attempting to renegotiate their attachments to work in ways that provided the autonomy and

[1] We distinguish these activities from "side hustles" (Komarovsky 1964, p. 281), a concept that refers to ways to supplement one's income more immediately. However, we did note quite a few side hustles as well—especially selling drugs while also working a legitimate job.

capacity for creativity and self-expression that they desired. This is reflected in the entrepreneurial nature of many of the side bets, and the emphasis on the creative and performing arts. The desire for autonomy is also evident in the jobs many hoped to obtain eventually. A considerable number dreamed of starting their own businesses, a theme that also appears in studies of prior generations of the working class (Halle 1984). For the men we interviewed, these ranged from opening a barbershop to starting a cannabis farm. A few had even quit steady jobs to start a business or go in on ventures with friends (as one example, a catering firm). Most, though, had had to return to wage work after the businesses failed, either in the wake of the Great Recession or when an injury or family illness sapped their resources.

The attempts of these men to renegotiate their involvements with work not only tended to emphasize autonomy, but generativity.[2] Demario, the barber mentioned earlier, joined a fraternity geared toward those in the "beauty industry." He valued the fact that the fraternity chapters hosted community events focused on boosting the self-esteem of youth and also did charitable work in the community. Blake, who had completed three years of college but was currently training to become a fork-lift operator, worked part time as assistant coach for an AAU basketball team so he could serve as a mentor to neighborhood youth. This desire was rooted in his own childhood, when a next-door neighbor "used to come out and coach [the kids] in the neighborhood basketball games. He saw something in the guys that I was around, and so he took us all in and pretty much adopted us. … And he's the reason that all of us [enrolled in] college." Bernard, the security guard mentioned earlier, dreamed of starting a company "doing landscaping and snow removal … and then I can employ the … kids from the neighborhood. Each kid can have a block, you know? A lot of people … they don't even want to give these kids a chance …, but I'm giving them something to work with first, work for your neighborhood, you know what I'm saying?" Steve had worked on a Charleston fishing trawler for two decades—despite being an alcoholic. Now sober and a convert to Christianity, he spent free time volunteering at a halfway house for recovering addicts. He dreamed of starting his own business because, "That way I'll be able to create jobs for these guys." Larry, a Boston resident hailing from Barbados, had noted how youth in his neighborhood had few resources to learn about Caribbean culture. For several years, he had sewn costumes for Boston's Caribbean American Heritage Festival. When we interviewed him, he was trying to find ways to make his hobby a commercial venture, with the goal of providing neighborhood youth with a source of employment that stoked a sense of cultural pride. Eventually, he wanted to open a community center where young people could learn about Caribbean culture while making and selling ethnic handicrafts.

As these stories illustrate, a desire for generative work—jobs that allow men to "give back" to their communities—is most often voiced when they are asked about

[2] Maruna (2001) has used the concept of generativity to describe the kinds of jobs (such as drug counselor, youth worker, and community volunteer) that were most useful in helping the ex-offenders he studied desist from crime.

the jobs to which they aspire. Aside from the examples above, which involve working with youth, other trades commonly named include substance abuse counselors and first responders. For example, one man who went by the nickname Bear told us that he aspired to be a firefighter "not so much for the pay [but because] it's self-fulfilling, you know? Just to save one baby. To run in and save someone else's child from burning to death." These men often defined a "good" job as one in which they could save someone else from harm.

Yet our analysis of men's life narratives suggests that many are also focused on rescuing themselves or those they see as younger versions of themselves. They tell stories of overcoming past trauma or substance abuse (Silva 2013). For instance, John had the words "my pain" tattooed across his neck. He told the interviewer, "I don't want [my son] to experience the pain I went through [in childhood] of neglect, someone not loving you, not being there." Bob is the child of two heroin addicts. His father died of a drug overdose when he was five and his mother left him home alone for extended periods of time. He told us his goal was to become a "psychologist," saying he wanted to counsel people about their problems. "I have been through everything. And I dealt with so much, [so] I know how to [help others]." In sum, the desire for generative work may represent a wish to help young people avoid the difficulties they perceive themselves to have experienced while growing up. The act of helping others may be a form of self-healing for many of them. Generativity expressed through "intergenerational buffering" has been described by psychologists Kotre and Kotre (1998, p. 367), who write, "although they themselves may bear scars, they say of a sequence of intergenerational damage, 'It stops here. It ends with me.'"

Family

The decline in labor force participation among prime-age men may be driven partly by the retreat from marriage because a lack of family responsibility may decrease men's motivation to work (as discussed in this issue by Binder and Bound). While the men that we interviewed seldom invested strongly in the role of romantic partner, they nonetheless embraced their paternal roles enthusiastically. While some were married or had been divorced, most had postponed marriage, claiming their financial situation wasn't sufficient, or more rarely, because they eschewed marriage altogether. This is in line with marriage trends for the working class as a whole. The proportion of households with children headed by a married couple, for example, declined markedly among those without a college degree since 1980 (Cherlin 2014; see also Binder and Bound, this volume).

For example, Manuel and his girlfriend were parents of two preschoolers, and planned on marrying, but were holding off until they could save money for a "big wedding." Jeff and his fiancé had set a wedding date, but at the time of our interview that date had come and gone. While she was ready to get married, he wanted more of a financial cushion beforehand. "I don't want to get married, have a honeymoon, and now we [are] worried about the mortgage next month. Let's get financially straight. We ain't got to be balling, but just to where the bills are paid," he explained.

As Jeff's narrative suggests, even those in marriage-like relationships (living together with children in common) usually expressed some degree of doubt about the permanency of their relationships (for example, "*If* we're still together in five years, *then* maybe we'll get married"). The married men were generally more confident that their relationships would last, but some were nonetheless hesitant to recommend marriage to friends. Thirty-three year-old Robert, a father of three, had gotten engaged at 18 but had waited to get married until 21 so he could drink at his wedding; "I figured if I was legal to drink—that was the point." He told us, "I have known so many people that have lived together for years, have kids together, things were great. They got married, six months later they're divorced. And I think that piece of paper changes … the way people act [for the worse]."

Yet fatherhood was a highly salient source of meaning and identity for nearly all these men. When asked about parenting his six-year-old, Fred said, "It's definitely great being a father. You just have a little person who looks up to you and [is] trying to emulate some of the stuff that you do. [A child is] someone who you can kind of help shape and form into like a, into a good person." Brian described fatherhood as follows: "To be honest with you it's taught me pure love. I knew pure love with my [ex-]wife, but with my daughter…, you don't get anything [better] than that. You know what I mean?" In contrast, men who have been separated from their children are often emotionally devastated. Bill lived with his twins' mother for nine years. Since they separated, she has not let him see the children. "She wants me completely out of their lives. She has completely slammed the door and made sure that I cannot do anything about it. It's destroyed me. It has destroyed me."

The form of fatherhood these men wish to enact is not modeled on what they observed among their own fathers and grandfathers, who—in their view—were inadequate. Rather, this generation places strong emphasis on nurture and warmth (see also Edin and Nelson 2013). Many derided their own fathers if they "merely" provided financially for the family but didn't provide emotional support. For example, Brian was critical of his father because "he doesn't show emotions. Now don't get me wrong, he's a phenomenal father when it comes to supporting his family and doing what needs to be done to take care of his family. He's always done that. But as far as showing emotion, giving us a hug, he just doesn't do that stuff."

One might question whether the emphasis on nurture and warmth has supplanted men's sense of duty to provide financially. Though the men we interviewed nearly always adhered to the notion that fathers should provide for their children economically, this sense of responsibility could be negated in cases where the custodial parent would not let them see their children, as in Bill's case, or if they felt that she did not need the money, due to her earnings or those of her current partner or spouse. Furthermore, men who did not live with their children often conceived of themselves as "helpers" rather than "providers" financially. Though men did not explicitly say so, the fact that they placed more emphasis on their emotional than their financial role may have weakened their motivation to work.

In sum, nearly all the men we spoke to viewed the father-child tie as central while the partner relationship was more peripheral. As with work, this renegotiation of the family role allows working-class men to exercise autonomy from the constraints of the spousal role while practicing generativity in the parental role. But as Bill's case shows, the success of men's attempts to renegotiate family life are often contingent on the cooperation of the children's mothers.

Religion

Although a small minority of our respondents claimed membership in an established religious tradition—"Irish Catholic" or "Holiness," for example—these identities were only rarely very salient. As Greg told us, "I'm 100 percent Christian at the heart of it. But as far as the practicing part, maybe 2 percent…." Jeremy said, "I treat church just like I treat my girlfriends…. I'll stick around for a while and then I'll go on to the next one."

As recently as the 1970s, white working-class Americans attended religious services as often as white college graduates did, but by 2010 the attendance of high-school-educated whites had fallen substantially, more than that of white college graduates (Wilcox, Cherlin, Uecker, and Messel 2012). Scholarship to date has noted this trend, but has not explained it. Several of the men in our study noted that growing mistrust of religious leaders may have played some role. In addition, ethnographies of prior generations of white working-class Americans referenced earlier in this article suggest that men were often tied to religious institutions through their wives. As marriage declined, men's church attendance might have fallen in tandem. However, it is also true that religious norms and sanctions tie men to the institution of marriage and that these have weakened. Participation in these institutions is intertwined.

Yet the majority of the men that we spoke to asserted the importance of faith. Brian, for example said he was "religious to a certain point…. I don't know if it's necessarily God or something else." Yet he also asserted that "you got to have faith … because why else would you want to be a human being? Why would you want to have kids … or take care of your kids? There's lots of whys [that require faith]." He believed there must be some sort of ultimate meaning because of "some of the things that I've been through in life…. There's a reason that I am here, because I should be dead."

While many of the men we interviewed grew up in households that were at least somewhat religious, most stopped attending services as soon as they were given a choice, generally in late elementary or junior high school. Only a few still identified strongly with an established religious tradition, and even many of those currently questioned the basic tenets of these faiths. Mark told us, "I'm not going to say I'm a whole-hearted believer—there's something far-fetched about him dying and coming back to life …" Blake explained that "… I believe in God, and in Jesus and that he died for our sins [but] I am not quick to say that I'm a Christian because religion itself was manmade and it contradicts itself."

Mistrust of religious leaders was often cited as a reason for eschewing a childhood faith. Some viewed clergy as little more than scam artists, a theme also evident in ethnographies of the white working class of prior generations (Gans 1962; Halle 1984). Of the Catholic Church in which he was raised, one interviewee said, "I mean, it's a beautiful religion, but I see some of the stuff that they've done around the world. You know, here you've got priests drinking out of gold goblets while [in the] villages around them … people are starving to death."

Some of the mistrust among those we interviewed stemmed from the sex abuse scandal in the Catholic Church. Greg, for example, still tried to "mold myself to be like Jesus … even though he may or may not have existed." Yet he rejected the childhood faith in which he was raised "when that whole priest thing came out. I became an agnostic." David hinted that his disbelief stemmed from abuse at the hands of the church; an altar boy who sang in the choir as a child, he said: "[S]omething happened to me when I was a kid that kind of set me in a tailspin, so I kind of gave up on all that shit." Yet he prayed to his dead grandmother twice a day—more if he was really struggling.

As these comments illustrate, almost all the men we spoke with rejected organized religion, yet a substantial minority insisted they were nonetheless deeply engaged in spiritual pursuits. They were attempting to renegotiate their relationship with religion by picking and choosing elements of various religious traditions they found appealing while visiting various congregations and conducting research on the internet or in the public library.

Bernard, for example, believed he heard the voice of God on his 26th birthday, which happened to be a Sunday, February 26th. He recalled that he was sitting in church when he heard a voice, telling him, "If you can do better [than the Holiness faith], then do it." First, he turned to Islam, then to Judaism, and then to various Christian denominations and sects: "Catholics, Pentecostal, and the Baptist and the Methodists and the Latter Day Saints and Jehovah's Witnesses." Bernard believed Martin Luther King was a prophet, just as Moses, Jesus, and Mohammed were. At the time we interviewed him, he was listing to YouTube recordings from a 95-year-old Jehovah's Witness.

Brandon's journey was similar. He was "heavily" involved in a Baptist church for about three years when he was in his early 20s, but then started "hearing a lot of stuff that didn't really sit right with me, [so I started] researching." Brandon said he did not "identify myself with any organized religion… I believe in spiritual energies."

Donald's beliefs were also more spiritual than religious. He followed Protestant televangelist Joel Osteen, whom he believed was also more spiritual than religious. Raised a Catholic, Donald's interest in faith didn't take hold until he joined Alcoholics Anonymous and Narcotics Anonymous. That is when "I finally sat down and took an inventory of the spiritual part and the religious part [of my life]." Donald recited the Serenity Prayer "about 25 times a day" and had recently received a birthday gift—a ring—with the prayer engraved on it. He was strongly opposed to organized religion and "religious people," but nonetheless deeply engaged in his spiritual life, which was grounded in his experience of addiction and pain. He told us: "[S]piritual people have been to hell, religious people are afraid of hell. … People like me who are spiritual [have already been to hell and back]. You can't scare an addict. …You can't scare me with jail, you can't scare me with death. I am not scared of anything. [But] it's religious people who are scared of everything. That's why they pray and need Jesus, because they are afraid."

John believed in the basic unity of all religions and studied them for underlying truths. He researched "all ancient religions" but particularly Christianity. Yet he was now convinced that only by studying the most ancient writing of the Sumerians, which, he claimed, have the "real story of creation," and other civilizations such as "Mayans, Incas, Aztecs," could one find hidden truths that have been obscured by technological distractions and authoritarian suppression. John believed that there are actually twelve planets, not nine, and that this has been known by scientists for some time, but "for some reason the government doesn't want you to know this."

Greg, described earlier, watched "a lot of documentaries [about religion] and [I] read a lot," including daily meditations written by the Dalai Lama, a book of the Tao, and also a book on parenting from a Christian perspective which he found "very helpful." This religious eclecticism was reflected in the tattoos he had gotten over the years, which included a large crucifix, an elven star which he referred to as a "pagan symbol," and the Eye of Horus, which he thought was the Egyptian God of the Sun.

These men's spiritual quests are often so specialized that they struggle to find others with common beliefs. Brian, who had an intense interest in ancient astronaut theories (that Earth has been visited by such beings in the past, even as recently as World War II with the cargo cults) and who was particularly drawn to the Earth Chronicles series by Zecharia Sitchin, told us, "I'm having a hard time with finding friends that I'm able to have conversations about this stuff [with]."

The questing, seeking religious style these men described is one example of autonomous action. But while offering men something they value—independent action that has the capacity for personal growth and development—this form of faith fails to tie men to other societal institutions such as the family. Nor does it promote conformity with behavioral expectations that may be conducive to work, such as being honest, hardworking, or sober. Like most others we talked to, Ed, who prays "to the God of my understanding," explicitly rejected the idea that faith should constrain his behavior. He told us that the God he learned about in Sunday school was "a God with strings telling us how to live. That didn't work for me."

In sum, this approach to religion often lacks the communal aspect of faith that, for centuries, has provided the norms and sanctions that promoted adherence to traditional social roles, plus the legitimation of these arrangements.

Haphazard Lives

Our interviews strongly suggest that the autonomous, generative self that many men described is also a haphazard self. For example, vocational aspirations usually remain nebulous and tentative, rarely taking the form of an explicit strategy. In the meantime, career trajectories are often replaced by a string of random jobs.

These men typically transitioned to parenthood more by accident than design, and in the context of tenuous romantic relationships. Some, like Demario, described earlier, didn't learn they were fathers until after their children were born. Others found out only when their children were several years old. Brandon's suspicions first formed when he viewed a Facebook post from his ex-girlfriend with photographs of her two-year-old twins. A subsequent email from her sister claimed he was the father. Both men demanded DNA tests (Brandon sought the test for both of the twins). Yet in keeping with our argument that fatherhood has become highly salient for working-class men, both were excited to learn that they were fathers, and now delighted in the role. Yet as Bill's case (above) illustrates, their ability to play this role in the future will be contingent on the willingness of their children's mothers to play along. Mothers often limit access to children after a breakup (Claessens 2007; Edin, Tach, and Mincy 2009). When children's mothers move to new partnerships, fathers are especially likely to become disengaged. Frequency of contact also falls when fathers have children with subsequent partners (Tach, Mincy, and Edin 2009).

Religious community and a systemic belief system have been replaced by a patched-together religious identity that holds little sway over behavior, especially as it is divorced from the communal aspects of faith that have adhered working-class men to a set of behavioral norms.

Yet through their attempts to renegotiate work, family, and religious roles, working-class men, whose fathers' and grandfathers' lives were often marked by limited autonomy in the workplace, gender-segregated roles within their family, and religious structures that dictated a set of rigid behavioral norms—these men are showing signs of moving beyond such strictures. Many will likely falter. Yet they are laying claim to a measure of autonomy and generativity in these spheres that were less often available in prior generations.

For their fathers and grandfathers, work, family, and religion created the attachments, investments, involvements, and beliefs (Hirschi 1969) that guided and gave meaning to human activity in specific social domains. In addition, this pattern was broadly shared within the community and successfully reproduced over time (Friedland and Alford 1991). These institutions not only organized social activity into common patterns of behavior, but supplied norms, beliefs, and rituals that legitimated such patterns. If traditional social roles in these domains are now only tenuously embraced, a few may craft lives that are more rewarding than those of prior generations, but the majority will struggle.

Racial and Ethnic Difference in Perceptions of Social Standing

Mortality statistics show sharp differences between racial and ethnic groups in suicide, drug overdoses, and liver-related mortality—"deaths of despair"—which have grown dramatically since the late 1990s among white men without college degrees. Among these men, all five-year age groups between 30–34 and 60–64 saw marked and similar increases in mortality from these causes between 1999 and 2013. For white men closest in age to those we studied, mortality from these causes nearly doubled (Case and Deaton 2015, figure 4). One explanation for these trends may lie in the striking racial and ethnic differences in satisfaction with one's social standing relative to one's parents.

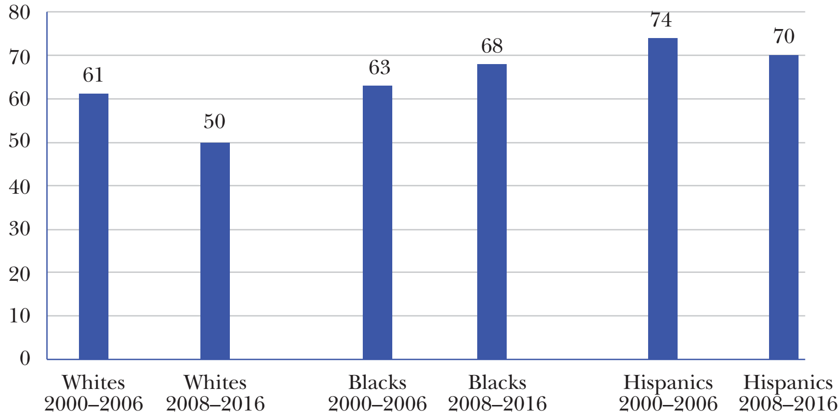
Fathers and grandfathers of today's young working-class men provided a standard of living that many of their adult sons cannot match today. This is particularly true for the whites, who when they look back can remember fathers and grandfathers who were sustained by the booming industrial economy of post-World War II America. African-Americans, however, did not get a fair share of the blue-collar prosperity of the post-World War II period. As a result, they may look back to a time when discrimination deprived their parents of such opportunities. Many Hispanics may look back to the lower standard of living their parents experienced in their countries of origin. Thus, whites are more likely to compare themselves to a reference group that makes them feel worse off, while blacks and Hispanics compare themselves to reference groups that may make them feel better off.

While acknowledging struggles brought on by the Great Recession, Blake, an African American father interviewed in 2012, nonetheless believes his generation is better prepared than his father's cohort. "I feel like our generation is taking steps to become more qualified for certain positions," he said. Greg, also African American, told us, "I think there are better opportunities now because … the color barrier is not as harsh as it was back then."

In contrast, Rick, a 35-year-old white construction worker, said, "It's much harder for me as a grown man than it was for my father." He remembered his father saying that back when he was 35, "'I had a house and I had five kids or four kids.' You know, 'Look where I was at.' And I'm like, 'Well, Dad, things have changed.'" Aaron, also white, explained: "I think that the whole job infrastructure has changed. I think for my dad's generation, there was more jobs in the sense that you could go out and you go learn how to build chairs. Now, I don't think you can learn how to build chairs because there's no work there, because they're not making them here, you know. … You can't do that now. At least not in the United States anyway. … I think that Americans are going to have less and less opportunities unless things change."

The General Social Survey, a biennial survey of the US adult population, has included this question: "Compared to your parents when they were the age you are now, do you think your own standard of living is much better, somewhat better, about the same, somewhat worse, or much worse than theirs was?" Figure 1 shows the percentage responding "much better" or "somewhat better" among people without college degrees between the prime working ages of 25 and 55. We compare responses from 2000–2006 (the years immediately prior to the Great Recession) to 2008–2016, similar to the period in which deaths among less-educated non-Hispanic whites due to overdose, suicide, and liver-related mortality grew so dramatically, as documented by

Figure 1
Percent Saying that Their Standard of Living is Better than Their Parents at the Same Age: Non-College-Graduates Ages 25–55



Source: Data from the General Social Survey, a biennial survey of the US adult population.
Note: Sample includes both men and women.

Case and Deaton (2015).[3] Since these patterns hold for both women and men, we distinguish by race, but not by gender. In both time periods, whites were less likely than blacks or Hispanics to say that they were doing better than their parents. In addition, the proportion judging their living standards to be better than their parents declined among whites from the earlier to the later period, whereas it rose among blacks and remained relatively high among Hispanics. This pessimism among whites may have been a motivating factor in the rise in suicide, and in the increasing use of opioids and alcohol that has led to racial disparities in mortality trends.

Conclusion

One of the central questions this analysis raises is whether valuing autonomy and generativity is a response to poor labor market options or a reflection of broader cultural trends. In our view, these economic and cultural forces are both at work, and mutually reinforcing (Zelizer 2002). When the extrinsic rewards of work (such as wages or job security) decline, the salience of intrinsic rewards may grow. Yet economic forces are insufficient to explain the specific form in which certain aspects of work that have become valued, namely the emphasis on independent action that has the potential for growth and development, not

[3] The General Social Survey does not indicate whether respondents were of Hispanic origin until 2000. Prior to that time, respondents were simply asked to describe their race/ethnicity, and were coded as black, white, or "other."

only for oneself and one's loved ones, but for others in community in need of protection and care. We see little chance that attempts to craft autonomous, generative selves will disappear even if economic opportunities expand, in part because of the salience of these values among the middle class (Bellah, Madsen, Sullivan, Swidler, and Tipton 1985). Evidence bolstering the claim that cultural forces may have relevance beyond economic conditions includes Kearney and Wilson's (2017) recent finding that the increases in economic opportunity due to localized fracking booms led to increased wages for less-educated men, but no corresponding increase in marriage.

This brings us back to the question of why labor force detachment is becoming more common among men with a high school diploma but no four-year college degree, especially when the official unemployment rate is so low. It is tempting to look for a single explanation for this increase. Although only a starting point, our findings suggest that these changes may be driven by the fact that the workplace, the family, and religion have all been transformed, along with men's sense of what constitutes fulfillment in all these domains. In addition, the salience of manual labor in identity formation seems to have weakened, compared to prior generations. If significant changes in any one of those arenas can be life-altering, the combined effect of all these changes will be quite unpredictable and will vary with the temperamental differences of the men who confront them.

Though our analysis should sound an alarm for the near term, we believe it is too soon to predict how these changes will play out over time as society adjusts to them. Society has faced shifts in the relationship of men to work, family, and religion before. Of the transition from mechanical solidarity (agrarian societies, with bonds based on likeness) to organic solidarity (industrial and postindustrial societies, where bonds are based on difference) in the late 19th century, Durkheim (1893 [1984], p. 339) wrote: "It has been rightly stated that morality … is in the throes of an appalling crisis. [T]he remedy for the ill is nevertheless not to seek to revive traditions and practices that no longer correspond to present-day social conditions." Rather, he argued, "We need to introduce greater justice into their relationships by diminishing those external inequalities that are the source of our ills." To ease the crisis of working-class men in labor force attachment, ill health, and mortality more than a century later, we may need to do the same.

The optimistic reading of the developments we have described is that working-class men are now sharing in the autonomy and generativity that was largely the province of middle- and upper-class men in previous generations. Moreover, the interest they show in being involved as fathers and in helping others could represent a widening of the boundaries of masculinity in ways that are more consistent with contemporary family and work life. The pessimistic reading is that these men are pursuing goals that they are unlikely to achieve due to their lack of social integration. They must find their way without ties to steady work, stable families, and organized religion. Without social support, their chances of success diminish. Those who fail to achieve the autonomous, generative selves they crave will have little to fall back on and few people to prevent them from sinking into despair.

■ *The authors thank Anne Case, Angus Deaton, Matthew Desmond, Gordon Hanson, Christopher Jencks, William Kornblum, Enrico Moretti, H. Luke Shaefer, Timothy Taylor, and members of Princeton's Workshop on Social Organization for their valuable suggestions. In particular, we thank Taylor and Jencks for superb editing and Jencks for the insights about the dangers of "one size fits all" explanations in understanding the decline in labor force participation. Elizabeth Powers conducted the majority of the interviews in Boston while Timothy Nelson interviewed most of the respondents in other sites.*

References

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the U.S. Labor Market." *American Economic Review* 96(2): 189–94.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. *Habits of the Heart: Individualism and Commitment in American Life*. Berkeley CA: University of California Press.

Case, Anne, and Angus Deaton. 2015. "Rising Morbidity and Mortality in Midlife among White non-Hispanic Americans in the 21st century." *PNAS* 112(49): 15078–83.

Case, Anne, and Angus Deaton. 2017. "Mortality and Morbidity in the 21st Century." *Brookings Papers on Economic Activity*, Spring: 397–476.

Cherlin, Andrew J. 2009. *The Marriage-Go-Round: The State of Marriage and the Family in America Today*. New York, NY: Alfred A. Knopf.

Cherlin, Andrew J. 2014. *Labor's Love Lost: The Rise and Fall of the Working-Class Family in America*. New York Russell Sage Foundation.

Claessens, Amy. 2007. "Gatekeeper Moms and (Un)Involved Dads: What Happens After a Breakup?" Chap. 9 in *Unmarried Couples with Children*, edited by Paula England and Kathryn Edin. New York: Russell Sage Foundation.

Coglianese, John. 2018. "The Rise of In-and-Outs: Declining Labor Force Participation of Prime Age Men." Working paper. Harvard University.

Durkheim, Emile. 1893 [1984]. *The Division of Labor in Society*. New York, NY: The Free Press.

Durkheim, Emile. 1897 [1997]. *Suicide: A Study in Sociology*. New York, NY: The Free Press.

Edin, Kathryn, and Timothy J. Nelson. 2013. *Doing the Best I Can: Fatherhood in the Inner-City*. Berkeley, CA: University of California Press.

Edin, Kathryn, Laura Tach, and Ronald Mincy. 2009. "Claiming Fatherhood: Race and the Dynamics of Father Involvement among Unmarried Men." *Annals of the American Academy of Political and Social Science* 621(1): 149–77.

Erikson, Erik H. 1993. *Childhood and Society*. New York: W. W. Norton & Company.

Friedland, Roger, and Robert R. Alford. 1991. "Bringing Society Back In: Symbols, Practices, and Institutional Contradictions." Chap. 10 in *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. J. DiMaggio, 232–263. University of Chicago Press.

Gans, Herbert J. 1962. *The Urban Villagers: Group and Class in the Life of Italian-Americans*. New York: Free Press.

Halle, David. 1984. *America's Working Man*. University of Chicago Press.

Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley, CA: University of California Press.

Inglehart, Ronald. 1977. *The Silent Revolution: Changing Values and Political Styles among Western Publics*. Princeton University Press.

Kearney, Melissa S., and Riley Wilson. 2017. "Male Earnings, Marrigable Men, and Nonmarital Fertility: Evidence from the Fracking Boom." NBER Working Paper 23408.

Kohn, Melvin. 1969. *Class and Confirmity: A Study in Values*. University of Chicago Press.

Komarovsky, Mirra. 1964. *Blue-Collar Marriage*. With the collaboration of Jane H. Philips. New York: Random House.

Kotre, John, and Kathy B. Kotre. 1998. "Intergenerational Buffers: 'The Damage Stops Here.'" Chap. 11 in *Generativity and Adult*

Development: How and Why We Care for the Next Generation, edited by Dan P. McAdams and Ed de St. Aubin. Washington, DC: American Psychological Association.

Lamont, Michèle. 2000. *The Dignity of Working Men: Morality and the Boundaries of Race, Class, and Immigration*. Cambridge, MA: Harvard University Press.

Maruna, Shadd. 2001. *Making Good: How Ex-Convicts Reform and Rebuild Their Lives.* Washington DC: American Psychological Association.

McAdams, Dan P., Holly M. Hart, and Shadd Maruna. 1998. "The Anatomy of Generativity." Chap. 1 in *Generativity and Adult Development,* edited by Dan P. McAdams and Ed de St. Aubin, Eds. Washington DC: American Psychological Association.

McLanahan, Sara. 2004. "Diverging Destinies: How Children are Faring Under the Second Demographic Transition." *Demography* 41(4): 607–627.

Putnam, Robert D. 2015. *Our Kids: The American Dream in Crisis.* New York: Simon and Schuster.

Rubin, Lillian B. 1976. *Worlds of Pain: Life in the Working-Class Family.* New York: Basic Books.

Silva, Jennifer M. 2013. *Coming Up Short: Working-Class Adulthood in an Age of Uncertainty.* New York: Oxford University Press.

Tach, Laura, Ronald Mincy, and Kathryn Edin. 2009. "Parenting as a 'Package Deal': Relationships, Fertility, and Nonresident Father Involvement among Unmarried Parents." *Demography* 47(1): 181–204.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.

Wilcox, W. Bradford, Andrew J. Cherlin, Jeremy E. Uecker, and Matthew Messel. 2012. "No Money, No Honey, No Church: The Deinstitutionalization of Religious Life among the White Working Class." *Research in the Sociology of Work* 23: 227–50.

Wilcox, W. Bradford, Nicholas H. Wolfinger, and Charles E. Stokes. 2015. "One Nation Divided: Culture, Civic Institutions, and the Marriage Divide." *The Future of Children* 25: 111–27.

Willis, Paul. 1977. *Learning to Labor: How Working Class Kids Get Working Class Jobs.* New York, NY: Columbia University Press.

Wuthnow, Robert. 2002. "United States: Bridging the Privileged and the Marginalized." Chap. 2 in *Democracies in Flux: The Evolution of Social Capital in Contemporary Society*, edited by R. D. Putnam. Oxford University Press.

Zelizer, Viviana A. 2002. "Enter Culture." Chap. 5 in *The New Economic Sociology: Developments in an Emerging Field*, edited by Mauro F. Guillén, Randall Collins, Paula England, and Marshall Meyer. New York: Russell Sage Foundation.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 229–242

Retrospectives
Ricardo on Machinery

Samuel Hollander

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please contact Joseph Persky, Professor of Economics, University of Illinois, Chicago, at jpersky@uic.edu.

Introduction

We are currently experiencing an outpouring of concern both popular and professional regarding technological unemployment. A report by the Institute for Public Policy Research, for example, estimates that 44 percent of UK jobs could be automated over the next decade or so (Lawrence, Roberts, and King 2017). In the United States, the Council of Economic Advisers (2016, p. 239) estimated that 83 percent of jobs with an hourly wage below $20 are threatened by automation. On the other side, a number of economists are quick to offer the rejoinder that while two centuries of technological progress since the Industrial Revolution has reshaped jobs and the labour market many times over, it has not been accompanied by a secular upward trend in unemployment (in this journal, Mokyr, Vickers, and Ziebarth 2015).

■ *Samuel Hollander is University Professor Emeritus, University of Toronto, Toronto, Canada. His email address is shollande@gmail.com.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.229 doi=10.1257/jep.33.2.229

I shall be discussing an apparent about-turn on the subject by David Ricardo (1772–1823), who at different times, even in different chapters of the same book, and, indeed, even at different places in the same chapter, seemed to be on both sides of the argument as to whether technological unemployment should be a matter for concern.

In a chapter entitled "On Machinery," added to the third edition of his *Principles of Political Economy* (1821), which comprises volume 1 of his *Collected Works* (1951–73), Ricardo announced that he had become concerned about the possibility, even likelihood, of technical change detrimental to labour's interests: "It is more incumbent on me to declare my opinion on this question," he explained, "because they have, on further reflection, undergone a considerable change" (Ricardo 1951–1973 1: 386). His original position, Ricardo explained, had assumed that all classes would benefit from the higher productivity and thus reduced prices allowed by use of machinery, including labourers who "would have the means of buying more commodities with the same money wages," while "no reduction in wages would take place, because the capitalist would have the same power of demanding and employing the same quantity of labour as before, although he might be under the necessity of employing it in the production of a new, or at any rate of a different commodity" (p. 387). In testimony before Parliament on December 16, 1819, Ricardo responded to a claim made regarding "the rapid inroad which machinery had made upon manual labour within only a few years. …[which] had thrown a great many hands out of employment" by stating unequivocally that "machinery did not lessen the demand for labour" (Ricardo 1951–1973, vol. 5: *Speeches and Evidence*, p. 30). (It should be noted that the "Hansard reports" of what was said in Parliament at this time were drawn from later newspaper reports rather than direct observation.) Or as Ricardo's comment was reported in *British Press*: "He never could think machinery could do mischief to any country, either in its immediate or its permanent effect" (p. 31).[1]

However, by 1821 Ricardo had become "convinced that the substitution of machinery for human labour, is often very injurious to the interests of the class of labourers," and that "the opinion entertained by the labouring class, that the employment of machinery is frequently detrimental to their interests, is not founded on prejudice and error, but is conformable to the correct principles of political economy" (Ricardo 1951–1973 1: 388, 392). He expressed his altered view in a number of places. The third edition of the *Principles* was the last published by Ricardo, but his new position was conveyed—again according to the Hansard report—in a House of Commons speech of May 30, 1823, shortly before his death in which he said: "[I]t was evident, that the extensive use of machinery, by throwing a large portion of labour into the market, while, on

[1] For additional confirmation that Ricardo had taken for granted a positive effect of machinery on labour's interests before the 1821 revision, his *Essay on Profits* (1815) asserts that "the effects of improved machinery, which it is now no longer questioned, has a decided tendency to raise the real wages of labour" (Ricardo 1951–1973 4: 35).

the other hand, there might not be a corresponding increase of demand for it, must, in some degree, operate prejudicially to the working classes" (Ricardo 1951–1973 5: 303). In a letter to J. R. McCulloch written in June 1821, Ricardo wrote: "If machinery could do all the work that labour now does, there would be no demand for labour. Nobody would be entitled to consume any thing who was not a capitalist, and who could not buy or hire a machine" (Ricardo 1951–1973 8: 399–400).

Ricardo's about-turn on this issue may seem clear-cut. However, in the very same "On Machinery" chapter, Ricardo also outlined qualifications to show that there was little need for concern. In fact, the index entry "A *qualified* use of machinery vindicated" (Ricardo 1951–1973 1: 436; Ricardo's emphasis) represents the primary message of the chapter to be a *defense* of machinery with qualification, rather than a justification of labour's fears subject to qualification.

Ricardo's opposing messages are reflected in contrasting reactions to the chapter "On Machinery." Some readers—including Thomas Robert Malthus and J. R. McCulloch—understood it as supporting working-class opposition to machinery. Others—including John Stuart Mill and Sir John Hicks—find therein the answer to such opposition. The story is rendered particularly interesting by the fact that Ricardo's prominent contemporary Thomas Robert Malthus may be shown to have priority over Ricardo regarding both the principle of labour displacement by machinery and its small practical import. But Malthus did not lay formal claim to a discovery and his contribution has gone unnoticed, while Ricardo is remembered as providing the drama—a financier opposing the unhampered operation of the market system should it threaten the welfare of the working-class majority.

Ricardo's Volte-Face on Machinery

To explain how machinery could be detrimental to the interests of labourers, Ricardo offered an arithmetical example entailing a capitalist employer who is both farmer and industrialist, and supposes that part of the workforce hitherto engaged in the production of food (and other necessaries) is diverted to machinery production, thereby reducing the so-called "wage fund."

In understanding the purpose of this example, it's important to bear in mind that Ricardo was not arguing that machinery must always be detrimental to labour. Instead, Ricardo was providing a counterexample to his earlier statements to the effect that machinery must always benefit labour. It's also important to bear in mind that Ricardo was himself dissatisfied with the example he chose to exposit his argument, representing it as "the most simple that I could select" (Ricardo 1951–1973 1: 390). His dissatisfaction is confirmed by the fact that (as we will discuss) he felt obliged to explain himself in correspondence immediately after publication, which suggests that the chapter was written in haste and scarcely revised properly.

Ricardo's (pp. 388–9) new case entails the application of newly discovered machinery. He sets out the initial state of affairs thus:

> A capitalist we will suppose employs a capital of the value of £20,000 and that he carries on the joint business of a farmer, and a manufacturer of necessaries. We will further suppose, that £7000 of this capital is invested in fixed capital, viz. in buildings, implements &c. &c. and that the remaining £13,000 is employed as circulating capital in the support of labour. Let us suppose, too, that profits are 10 per cent, and consequently that the capitalist's capital is every year put into its original state of efficiency, and yields a profit of £2000.
>
> Each year the capitalist begins his operations by having food and necessaries in his possession of the value of £13,000, all of which he sells in the course of the year to his own workmen for that sum of money, and, during the same period, he pays them the like amount of money for wages: at the end of the year they replace in his possession food and necessaries of the value of £15,000, £2000 of which he consumes himself, or disposes of as may best suit his pleasure and gratification. As far as these products are concerned, the gross produce for that year is £15,000, and the net produce £2000.

Ricardo next supposes that "the capitalist employs half his men in constructing a machine, and the other half in producing food and necessaries as usual. During that year he would pay the sum of £13,000 in wages as usual, and would sell food and necessaries to the same amount to his workmen; but what would be the case the following year?" The workmen as a group would again produce output of £15,000, but half of that would be in the form of food and necessaries valued at £7500, while the other half would be a machine valued at £7500, plus fixed capital as before of £7000. This sums to £22,000, of which £2000 would again be profit to the capitalist. Thus "[a]fter deducting this latter sum for his own [personal] expenses, he would have a no greater circulating capital than £5500 with which to carry on his subsequent operations; and, therefore, his means of employing labour, would be reduced in the proportion of £13,000 to £5500, and, consequently, all the labour which was before employed by £7500, would become redundant" (Ricardo 1951–1973 1: 389).

Before the machinery was in place, as Ricardo notes, "gross produce for that year is £15,000, and the net produce £2000." But after the machinery is in place, profit to the capitalist ("net produce") remains at £2000, while total output ("gross produce") for the year has fallen to £7500. Even if profit for the capitalist was somewhat higher than before, gross output could be lower. Thus, a capitalist might be led to reduce output in a search for higher profits. Or as Ricardo put it (p. 388): "My mistake arose from the supposition, that whenever the net income of a society increased, its gross income would also increase. I now, however, see reason to be satisfied that the one fund, from which landlords and capitalists derive their revenue, may increase, while the other, that upon which the labouring class mainly depend, may diminish, and therefore it follows, if I am right, that the same cause which may increase the net revenue of the country, may at the same time render the population redundant, and deteriorate the condition of the labourer."

The next section will describe the qualifications that Ricardo put on this idea, which suggest that his example is intended to apply only to a case in which machinery is suddenly discovered and then rapidly and extensively used, not to the more common case in which saving from profits gradually leads to additional investment in capital. Ricardo also points out that his chosen example should be distinguished from the substitution of machinery for labour in response to rising real wages during the growth process, which was an issue of long standing, when "[t]he demand for labour will continue to increase with an increase of capital, but not in proportion to its increase" (Ricardo 1951–1973 1: 395). But before considering such qualifications, let's enumerate some of the other issues with Ricardo's example.

First, Ricardo's basic model assumes the same profit rate of £2000 both before and after the new machinery, when of course it is the promise of a higher return that is the instigator of the conversion. It would seem straightforward for Ricardo to have presented an example with higher profits, but perhaps he felt that higher profits would overly complicate the example. In addition, the entrepreneur in the example does not seem to take into account that the purchasing power of even a constant net income would rise with cost reductions due to the advanced process in operation.

Second, while the concern throughout the discussion of machinery is the impact on *national* employment, Ricardo's basic model relates to an individual capitalist, and the connection from individual capitalist to overall economy is never clarified. However, one can imagine Ricardo envisioning that if many capitalists acted along the lines of the example, the effects on workers would be widespread.[2]

Third, Ricardo's example seems focused on the specific time period of the conversion to the new machinery undertaken *prior to operation of the new process*, without taking into account any effects from the later actual use of the machinery. In an explanatory letter to J. R. McCulloch on June 18, 1821, Ricardo confirms that the unemployment generated in his model relates specifically to the *conversion* itself: "I have said that when a manufacturer is in possession of a circulating capital he can employ with it a greater number of men, and if it should suit his purposes to substitute a fixed capital of an equal value for this circulating capital, it will be inevitably followed by a necessity for dismissing a part of his workmen. For a fixed capital cannot employ all the labour which it is calculated to supersede" (Ricardo

[2] Ricardo (1951–1973 1: 390–1) was prepared to make his case more realistically by introducing intersectoral exchange, supposing machinery to be applied in manufacturing. The demand for labour falls, say, in cloth-making. But because clothing-labour's consumption is reduced, the agricultural sector contracts in response, which "leads us to the same result; the demand for labour would diminish, and the commodities necessary to the support of labour would not be produced in the same abundance" (p. 391).

1951–1973 8: 390). A fall in gross produce implies a creation of labour redundancy at the conversion stage, but the redundancy does not increase after that stage.[3]

Fourth, Ricardo confirmed in later correspondence that his concern with machinery was not restricted to the creation of excess labour supply, but extended to wage reduction in correction of the excess: "Labour will fall because there will be a diminished demand for it" (Ricardo 1951–1973 8: 399). In addition, in a letter to McCulloch (June 30, 1821; p. 388), Ricardo pointed out that even if the new machinery allowed cheaper production, this wasn't going to be much help if the machinery had "tended to diminish the number of the class of buyers" because workers have suffered lost jobs or reduced wages.

Fifth, might the labour cost of repairing machinery alter Ricardo's example in a meaningful way? Responding to an objection by McCulloch that he had neglected *capital durability*, Ricardo admitted in a letter written June 18, 1821, that he should have been clearer regarding the "life of the machine" which he had taken for granted would be understood. Ricardo's (1951–1973 8: 388–9, letter of June 18, 1821) example relates to machinery in cloth-making:

> If I have not said whether the machine was to last one, ten, or a hundred years I have not been so explicit as I ought to have been. I admit too that it is as plain as any proposition in geometry that if it lasted only one year there could be no diminution in the demand for labour, but I do not admit that the same result would necessarily take place if the machine lasted for ten years. If the machine were to last one year only, the [output] produced must be as great a value at least as before, but if it were to last 10 years, a value much less than that, would afford the ordinary profits of stock, because although the same amount of capital would be employed, less of that capital would be employed in the maintenance of labour, and consequently a less deduction would be annually made from the gross value of the commodity produced. It is what remains after this deduction that invariably constitutes profits…. Give to the machine greater durability, and a less return … will be sufficient to compensate the manufacturer, because he must sacrifice fewer yards for the purpose of keeping his fixed capital [machinery] in its original efficient state.

[3] The term "calculated to supersede" confirms that dismissal of labourers occurs at the machine-construction period. The contrast with the period of actual operation is expressed in the *Principles* as "the discovery *and use* of machinery" and "the discovery, and *useful application*" of machinery (Ricardo 1951–1973 1: 390–1; emphasis added); and the matter of motivation at the conversion stage as: "the motives for employing machinery are always sufficient to ensure its employment, although it may, and frequently must, diminish both the quantity of the gross produce, and its value" (p. 392). The decision to convert turns therefore on the *calculation* that when the machinery is activated even a reduced gross output will suffice to generate a net income somewhat higher than initially and cover the costs corresponding to the reduced labour input. Notice that at the conversion stage the fall in gross produce is *in prospect only*; and that if the conversion is proven justified by events, the reduced gross output flow dictates continuance of activity at the lower level of employment due to the conversion.

Ricardo intimates here that the "conversion" problem arises from the fact that machinery need not be entirely replaced in as short a period as the circulating capital (wages) which it displaces. But this is a new gloss since Ricardo had in fact mentioned repair of machinery only fleetingly in his chapter (Ricardo 1951–1973 1: 389).

Ricardo's underlying presumption appears to be that output-reducing technologies are common. The underlying problem here is that Ricardo lacks a theory of production to justify when the output-reduction effects of the "discovery and use" of machinery are more or less likely to arise.[4] As we will see in the next section, Ricardo offers a number of qualifications to his pessimistic conclusions about machinery and the working class—in fact, the qualifications are so strong as to weaken greatly the concerns over machinery and jobs.

The "New View" Qualified

In approaching Ricardo's qualifications to his new view of machinery, I draw attention to the curious entry in the index of the third edition of his *Principles* noted at the outset of this paper: "A *qualified* use of machinery vindicated." That phrasing subtly diverts attention from the confession of error in having once supported machinery as necessarily favorable to all classes.

Ricardo seems to agree with "the opinion entertained by the labouring class that the employment of machinery is *frequently* detrimental to their interests" or "is *often* very injurious to the interests of the class of labourers." But Ricardo then withdraws much of the admission by insisting only on the *possibility* of the case adverse to labour.

For example, recall the proposition cited earlier that "the discovery and use of machinery *may* be attended with a diminution of gross produce" (emphasis added). To the same effect, Ricardo he wrote to Malthus immediately after publication "my sole complaint against [machinery] is that it *sometimes* actually diminishes the gross produce" (July 21, 1821; in Ricardo 1951–1973 9: 23; emphasis added).

Most significantly, Ricardo assures readers in the "On Machinery" chapter that cases of output-reducing investment in machinery were the exception: "The statements which I have made will not, I hope, lead to the inference that machinery should not be encouraged. To elucidate the principle, I have been supposing, that improved machinery is *suddenly* discovered, and extensively used; but the truth is, that these discoveries are gradual, and rather operate in determining the employment of the capital which is saved and accumulated, than in diverting capital from its actual employment" (Ricardo 1951–1973 1: 395). Even in the event of an abrupt capital conversion there would be "the stimulus to savings from revenue, which … an abundant net produce will afford," thereby reducing, even eliminating, the initial excess labour supply (p. 392). This indeed is Ricardo's closing theme: "I have

[4] Wicksell (1934, p. 137) objected to the Ricardian analysis on marginal-productivity grounds. Samuelson (1989) rejects Wicksell and "vindicates" Ricardo's allowances for an increase in net product and reduction in gross product (Samuelson 1988).

before observed, too, that the increase of net incomes, estimated in commodities, which is always the consequence of improved machinery, will lead to new savings and accumulations. These savings, it must be remembered, are annual, and must soon create a fund, much greater than the gross revenue, originally lost by the discovery of the machine, when the demand for labour will be as great as before, and the situation of the people will be still further improved by the increased savings which the increased net revenue will still enable them to make" (p. 396).

Reactions

Ricardo's effort to downplay the adverse effects for labour of machinery, immediately after raising the issue with so much fanfare, did not always succeed in calming his critics. Two in particular stand out.

Writing to Ricardo, Thomas Robert Malthus agreed "in the theory of your propositions, but practically I think that the cases are very rare in which for any length of time the gross produce is diminished by machinery" (letter of July 16, 1821; in Ricardo 1951–1973 9: 18). Malthus also expressed concern that workers would "take fast hold" on the strong affirmation that "the opinion entertained by the labouring class, that the employment of machinery is frequently detrimental to their interests, is not founded on prejudice and error, but is conformable to the principles of political economy."

J. R. McCulloch, hitherto a loyal Ricardo acolyte—he had shortly before changed his own view of machinery in the light of Ricardo's insistence that it could *not* diminish demand for labour—complained bitterly regarding "the extreme erroneousness of the principles to which you have incautiously lent the sanction of your name," which implied that "the laws against the Luddites are a disgrace to the statute book" (letter of June 5, 1821; in Ricardo 1951–1973 8: 382, 385). These allusions relate to illegal industrial action entailing machine-breaking, and to the harsh government reaction. McCulloch also interpreted Ricardo's new argument as an egregious surrender to those concerned with overproduction: "The fundamental differences that formerly existed (for I am sorry to think they have now nearly disappeared) between you and Messrs. Malthus and Sismondi induced many to believe that Political Economy was a thing of fudge, a fabric without a foundation – And I certainly think that those who were formerly of that opinion have a good deal better ground for entertaining it now" (p. 382).

But Ricardo firmly turned back the charge that he had countenanced a Malthusian theory of overproduction: "Mr. Malthus's objection to machinery is that it adds so much to the gross produce of the country that the commodities produced cannot be consumed—that there is no demand for them: mine, on the contrary is that the use of machinery often diminishes the quantity of the gross produce, and although the inclination to consume is unlimited, the demand will be diminished, by the want of means of purchasing. Can any two doctrines be more different?" (June 18, 1821; in Ricardo 1951–1973 8: p. 387).[5] (Again, one notes that Ricardo claims that machinery "often" diminishes gross output, which conflicts with the softening qualifications he also expresses.)

However, others reacted to Ricardo's "On Machinery" chapter by using his "qualifications" to defend the use of machinery as advantageous to labor. Thus, J. S. Mill rehearsed the Ricardian position in all editions of his *Principles of Political Economy*, the last appearing in 1871. (Mill does not refer explicitly to Ricardo, but it would be an extraordinary coincidence were "On Machinery" not open before him or at least firmly registered in his memory as he wrote!) Mill's account is important in that it confirms the interpretation offered above.

Mill (1848 [1965]) writes that "all increase of fixed capital, when taking place at the expense of circulating capital, must be, at least temporarily, prejudicial to the interests of the labourers. This is true, not of machinery alone, but of all improvements by which capital is sunk; that is, rendered permanently incapable of being applied to the maintenance and remuneration of labour" (pp. 93–4).Mill writes that the problem was to assure the employment of "as many labourers as before, *and pay them as highly*" (p. 95, emphasis added).

Mill's optimistic qualifications about the effect of machinery on workers exactly parallel those of Ricardo. Even temporary negative effects on the demand for labour reflecting the conversion of circulating into fixed capital do not come into play should innovation be financed from *net* accumulation which Mill (1848 [1965], p. 97) averred was the standard pattern:

> I do not believe that as things are actually transacted, improvements in production are often, if ever injurious, even temporarily, to the labouring classes in the aggregate. They would be so if they took place suddenly to a great amount, because much of the capital sunk must necessarily in that case be provided from funds already employed as circulating capital. But improvements are always introduced very gradually, and are seldom or never made by withdrawing circulating capital from actual production, but are made by the employment of the annual increase.

In addition, because the new technology itself tended to encourage both the "ability" and "motive" to save, it followed that "at the slow pace at which improvements are usually introduced, a great part of the capital which the improvement ultimately absorbs, is drawn from the increased profits and increased savings which

[5] Ricardo's representation here of Malthus's position regarding machinery is misleading. Malthus always denied that he was opposed to the use of machinery, insisting only that proper attention be accorded aggregate demand in the usual case where the gross output expands as a result. Elsewhere Ricardo conveys Malthus's position more accurately: "Mr. Malthus thinks that in many cases [inventions to save labour] would be disastrous presents to them, they must be accompanied, according to him by demand to make them beneficial. Now I think that demand depends only on supply, the means of obtaining abundance of commodities can never I think be otherwise than beneficial" (Note 243; in Ricardo 1951–1973 2: 365).

it has itself called forth" (Mill 1848 [1965] pp. 97–8). It was "[t]his tendency of improvements in production to cause increased accumulation, and thereby ultimately to increase the gross produce, even if temporarily diminishing it … which is the conclusive answer to the objections against machinery; and the proof thence arising of the ultimate benefit to labourers of mechanical inventions even in the existing state of society" (pp. 98–9). The adverse effect of new technology on employment involved "a case abstractedly possible [rather] than one which is frequently realized in fact" (p. 134).

Closer to our own day, Samuelson (1988, 1989) commended "On Machinery" for stating the possibility of an increase in net revenue combined with a fall in gross revenue as a result of the use of machinery. Sir John Hicks (1971, p. 925) too allowed that "[i]t is perfectly possible that the adoption of a (profitable) new invention may lead to a (temporary) reduction in final output." But most importantly for our focus on capital conversion as responsible for labour redundancy, Hicks (1973, p. 99) applauded Ricardo for conveying the message that "[t]o industrialize, without the savings to support your industrialization, is to ask for trouble." Hicks was no less impressed by Ricardo's allowance (quoted above) for the "new savings and accumulation" made from "the increase of net incomes, estimated in commodities, which is always the consequence of improved machinery." Hicks (1969, p. 153) wrote: "It was nevertheless to be expected (as Ricardo did in fact expect) that the time would come when the adverse effect of the swing to fixed capital would be exhausted, so that the favourable effect of the higher growth rate would alone survive."

On the Transition to the New View: A Malthus Influence?

Recall that upon reading "On Machinery," Malthus accepted Ricardo's theoretical argument regarding the effect of machinery in reducing gross output and demand for labour while insisting on the rarity of the phenomenon. That Malthus so readily accepted the theoretical case should come as no surprise, because in his own *Principles of Political Economy* (1820) he himself had provided a formulation that is almost the duplicate of Ricardo's. Malthus wrote:

> If … a capitalist who had employed £20,000 in productive labour, and had been in the habit of selling his goods for £22,000, making a profit of 10 per cent, were to employ the same quantity of labour in the construction of a machine worth £20,000, which would enable him to carry on his business without labour in future, except as his machine might require repair, it is obvious that, during the first year, the same value of the annual produce and the same demand for labour would exist; but in the next year it would only be necessary for the capitalist, in order to obtain the same rate of profits as before, to sell his goods for a little more that £2,000 instead of £22,000. The value of the annual produce would fall, the capital would not be increased, and the revenue would be decidedly diminished, and upon the principle that the demand for labour depends

upon the rate at which the value of the general produce, or of the capital and revenue taken together, increases, the slackness of the demand for labour under such circumstances would be adequately accounted for (pp. 261–2).

There is no mention here of the promise of an *increase* in net revenue (profits) to motivate the conversion—but there was none in Ricardo's initial formulation, either. Malthus (1820, p. 425) also argued that the case of output-decreasing machinery was unlikely to be encountered in reality:

> [I]t happens but seldom that we have to determine the amount of advantage or disadvantage occasioned by the increase of the neat [net], at the expense of the gross revenue. The interest of individual capitalists uniformly prompts them to the saving of labour, in whatever business they are engaged; and both theory and experience combine to shew that their successful efforts in this direction, by increasing the powers of production, afford the means of increasing, in the greatest practical degree, the amount and value of the gross produce…

Malthus thus preceded Ricardo with regard to the theory of conversion of circulating into fixed capital with its damaging consequences for labour's interests and to the qualification concerning the norm. That Malthus did not lay claim to the discovery may be explained by his perception of the theoretical possibility as a sort of the *curiosum,* but also, I suspect, because Malthus (1820 p. 401) maintained that "inventions" are usually *endogenous*: "Inventions to save labour are generally called forth by the wants of mankind in the progress of improvement; and therefore seldom much exceed those wants."

In keeping with the long friendship and detailed correspondence between Ricardo and Malthus (on which see Dorfman 1989), Ricardo composed "Notes" on Malthus's *Principles* upon its appearance in 1820. In Note 149, Ricardo wrote: "[T]o the capitalist it can be of no importance whether his capital consists of fixed or of circulating capital, but it is of the greatest importance to those who live by the wages of labour; they are greatly interested in increasing the gross revenue, as it is on the gross revenue that must depend the means of providing for the population. If capital is realized in machinery, there will be little demand for an increased quantity of labour" (Ricardo 1951–1973 2: 236).[6] Given the similarity in the numerical examples used by Malthus and Ricardo, it seems plausible that the seeds of Ricardo's case

[6] Piero Sraffa, Ricardo's editor, regards Note 149 as "a transition-stage in Ricardo's thinking on the subject" having in mind Ricardo's statement that it is "on the gross revenue that must depend the means of providing for the population," but lacking recognition in this note "that improved machinery might actually diminish the gross produce" (Sraffa in Ricardo 1951–1973 1: lviii–lix). However, Ricardo was in fact further away from his final position than Sraffa allows, considering the contrast between Malthus's representation in his passage of an *exogenous* redistribution of a given capital stock—corresponding precisely to the case about to appear in Ricardo's 1821 chapter—and Ricardo's comment which relates to the effect on labour demand of a growing total capital depending on its allocation between fixed and circulating capital.

may well have been sown (consciously or not) in the course of reading Malthus's *Principles*.

Malthus is relevant to our concern with the origins of Ricardo's new view for another and even more specific reason. Attached to one of the Malthus texts cited above is a footnote addressing Ricardo's chapter "On Gross and Net Revenue," which was published in the first two editions of the latter's *Principles*. That chapter Malthus *misunderstood* as maintaining that national advantage derived from a technological change permitting unchanged net revenue but *reduced* gross revenue and lower employment, to which Malthus (1820, pp. 425–6n) objected: "What is to become of the capital as well as the people in the case of such a change? It is obvious that a considerable portion of it must become redundant and useless." Malthus then confirmed his own approval of "all saving of labour and inventions in machinery," while distancing himself from what he took to be Ricardo's position:

> I quite agree with Mr. Ricardo, however, in approving all saving of labour and inventions in machinery; but it is because I think that their tendency is to increase the gross produce and to make room for a larger population and a larger capital. If the saving of labour were to be accompanied by the effects stated in Mr. Ricardo's instance, I should agree with M. de Sismondi and Mr. Owen in deprecating it as a great misfortune.

This reaction is precisely the same as the one Malthus expressed the following year upon reading "On Machinery," when he dissented from Ricardo's formulation that machinery is "frequently" detrimental to the working class.

Here, then, we find Malthus attributing to Ricardo the belief that machinery will frequently have a negative effect on the working class—and doing so a full year ahead of the publication of Ricardo's "On Machinery" chapter. In responding to Malthus, Ricardo in his Note 257 protested at the attribution to him of finding national advantage in deriving the same net income from a smaller workforce, implying approval of layoffs. His sole objective of his chapter "On Gross and Net Revenue" had been to argue that no social gain—with national "power" specifically in mind—flowed from generating a given net revenue by means of more rather than fewer men (Ricardo 1951–1973 2: 381–3). Yet to arrive at the case envisaged in "On Machinery," all that Ricardo had to do was alter his perspective slightly, from a *comparison* of alternative states of gross and net revenue to a *change* between states (as Malthus had in fact understood him to have intended). Whether this mental process was in fact the relevant one we cannot say, but it should be entertained as an open possibility.[7]

[7] A linkage has often been suggested between John Barton's *On the Condition of the Labouring Classes of Society* (1817) and Ricardo's "new" position. In an earlier exchange of correspondence with Barton, Ricardo had rejected a proposition stating—in Ricardo's description of the argument—that additions to fixed capital would lead to a smaller increase in output giving "less permanent employment to labour" than would a similar increase in circulating capital, since the same rate of profit can be earned with a smaller addition to output in the former than in the latter case (Ricardo to Barton, 10 May 1817; in

Recapitulation and Some Policy Implications

Ricardo's admission of error in 1821 relates to the recognition of labour redundancy generated by the conversion of circulating or wage-goods capital into fixed capital in the form of machinery. His example of how machinery can harm the working-class focused on the short-term time period when new machinery is first introduced. But he offered the qualification that unemployment due to machinery construction would, however, be counterbalanced by an increased demand for labour reflecting investment financed from savings enabled by higher productivity and profit margins relating to the new process. This in fact was the *primary* qualification to the pessimistic outlook. Furthermore, although Ricardo implicitly assumes constant real wages in his basic analysis, he did recognize the possibility of a fall in the real wage due to excess labour supply which in principle would encourage reabsorption of unemployed workers. But reduction in the real wage illustrated for him damage to labour's interests caused by machinery no less than did technological unemployment as such. In contrast with our own time, there was no concern with specificity of skill hindering labour transfers.

The Ricardo texts seem to offer an implicit concession that if innovations were introduced rapidly and on a large scale, *and not financed out of increased savings*, he might have been obliged to conclude that machinery should be in some manner discouraged. However, Ricardo warned that "employment of machinery could never be safely discouraged in a State, for if a capital is not allowed to get the greatest net revenue that the use of machinery will afford here, it will be carried abroad," whereas "while a capital is employed in this country, it must create a demand for some labour; machinery cannot be worked without the assistance of men, it cannot be made but with the contribution of their labour" (Ricardo 1951–73 1: 396–7). While Ricardo reaffirmed his concern with technological unemployment in his House of Commons speech of May 30, 1823, only a few months before his death (as noted above), he nonetheless insisted that "he would not tolerate any law to prevent the use of machinery," after taking into account the loss of export markets to foreign competitors not similarly constrained (Ricardo 1951–1973 5: 303).

Of course, there do exist possible interventions to shield the working-class from economic disruption that would be less intrusive than preventing use of machinery, and such ideas were not unknown in Ricardo's time. As one example, Joseph Lee (1656, pp. 22–3) had much earlier recommended regulation of land enclosure, such as the protection of labourers' cottages by contract, to assure that its benefits do not accrue solely to the landowning class. As another example closer to Ricardo's day, Bentham (1800 [1843], p. 39) maintained that "increase of wealth by saving labour [by introduction or improvement of machinery] is not so great as increase

---

Ricardo 1951–1973 7: 156–7). Barton is indeed referred to favourably in the "On Machinery" chapter, but with respect to substitution of machinery for labour in response to wage increases during the growth process resulting in a lag in the growth rate of labour demand behind that of capital, rather than to the main case of exogenous capital conversion (Ricardo 1951–1973 1: 395–6n).

of wealth by increase of quantity of labour; and that, consequently, opposition to machinery is well grounded, if no care is taken to provide immediate employment for the discharged hands."

Ricardo had traditionally argued against public works on the grounds (which came to be known as the "Treasury View") that such expenditure merely diverted funds "from other employments which would be equally if not more productive to the community" (letter to Malthus, January 3, 1817; in Ricardo 1951–1973 7: 116). But while conceding the potentially negative effects of machinery on employment, Ricardo continued to reject such proposals. [8] In the Hansard report of his House of Commons speech on May 30, 1823, Ricardo was still rejecting any "law to prevent the use of machinery." As for any policy proposal, all we find is a disappointing suggestion that "the people had the remedy in their own hands. A little foresight, a little prudence … a little of that caution which the better educated felt it necessary to use, would enable them to improve their situation" (Ricardo 1951–1973 5: 303). Such a sentiment tacitly acknowledges the inflammation of a social problem, but offers no immediate solution.

[8] It would be premature to draw definitive conclusions regarding Ricardo's overall social orientation from his position on public works. For example, a comprehensive study would take into account his courageous protestations, registered in his Parliamentary voting record and in correspondence, against the repression of public protests culminating in the "Manchester massacre" (Peterloo) of 1819 (Ricardo 1951–1973 5: xxii; 8: 80).

References

Barton, John. 1817. *On the Condition of the Labouring Classes of Society*. London: Arch.

Bentham, Jeremy. 1800 [1843]. "Manual of Political Economy." In *The Works of Jeremy Bentham* vol. 3, edited by John Bowring, pp. 31–84. Edinburgh: William Tait.

Council of Economic Advisers. 2016. *Economic Report of the President*. February. The White House.

Dorfman, Robert. 1989. "Thomas Robert Malthus and David Ricardo." *Journal of Economic Perspectives* 3(3): 153–64.

Hicks, John R. 1969. *A Theory of Economic History*. Oxford: Clarendon Press.

Hicks, John R. 1971. "A Reply to Professor Beach." *Economic Journal* 81(324): 922–25.

Hicks, John R. 1973. *Capital and Time: A Neo-Austrian Theory*. Oxford: Clarendon Press.

Lawrence, Mathew, Carys Roberts, and Loren King. 2017. "Managing Automation Employment, Inequality and Ethics in the Digital Age." Discussion Paper, IPPR Commission on Economic Justice, December. https://www.ippr.org/files/2018-01/cej-managing-automation-december2017.pdf.

Lee, Joseph. 1656. *A Vindication of a Regulated Inclosure*. London: Thomas Williams.

Malthus, Thomas Robert. 1820. *Principles of Political Economy*. London: John Murray.

Mill, John Stuart. 1848 [1965]. *Principles of Political Economy*. In *The Collected Works of John Stuart Mill*, vols. 2–3, edited John M. Robson. University of Toronto Press.

Mokyr, Joel, Chris Vickers, and Nicolas L. Ziebarth. 2015. "The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?" *Journal of Economic Perspectives* 29(3): 31–50.

Ricardo, David. 1951–1973. *The Works and Correspondence of David Ricardo*, 11 vols, edited by Piero Sraffa. Cambridge University Press.

Samuelson, Paul A. 1988. "Mathematical Vindication of Ricardo on Machinery." *Journal of Political Economy* 96(2): 274–81.

Samuelson, Paul A. 1989. "Ricardo was Right!" *Scandinavian Journal of Economics* 91(1): 47– 62.

Wicksell, Knut. 1934. *Lectures on Political Economy*, vol. 1. London: Routledge.

Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 243–250

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by email at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., St. Paul, MN 55105.

Smorgasbord

An OECD publication, *The Role and Design of Net Wealth Taxes in the OECD*, reports: "Net wealth taxes are far less widespread than they used to be in the OECD. … While 12 countries had net wealth taxes in 1990, there were only four OECD countries that still levied recurrent taxes on individuals' net wealth in 2017. Decisions to repeal net wealth taxes have often been justified by efficiency and administrative concerns and by the observation that net wealth taxes have frequently failed to meet their redistributive goals. The revenues collected from net wealth taxes have also, with a few exceptions, been very low. More recently, however, some countries have shown a renewed interest in net wealth taxes as a way

■ *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives, *based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.243 doi=10.1257/jep.33.2.243

to raise revenues and address wealth inequality. … While the tax system should help address wealth inequality, the question is whether a wealth tax is the most effective way to do so. The report assesses the case for and against net wealth taxes, looking at efficiency, equity and administrative arguments. … Overall, the report concludes that from both an efficiency and equity perspective, there are limited arguments for having a net wealth tax in addition to broad-based personal capital income taxes and well-designed inheritance and gift taxes." April 2018, http://www.oecd.org/ctp/the-role-and-design-of-net-wealth-taxes-in-the-oecd-9789264290303-en.htm.

Eric A. Posner, Glen Weyl, and Suresh Naidu open a discussion of "Antitrust Remedies for Labor Market Power." From the abstract: "Although the antitrust laws prohibit firms from restricting competition in labor markets as in product markets, the government does little to address the labor market problem, and private litigation has been rare and mostly unsuccessful. One reason is that the analytic methods for evaluating labor market power in antitrust contexts are far less sophisticated than the legal rules used to judge product market power. To remedy this asymmetry, we propose methods for judging the effects of mergers on labor markets. We also extend our approach to other forms of anticompetitive practices undertaken by employers against workers. We highlight some arguments and evidence indicating that market power may be even more important in labor markets than in product markets." *Harvard Law Review*, December 2018, 132, pp. 536–601, https://harvardlawreview.org/2018/12/antitrust-remedies-for-labor-market-power.

Santiago Levy discusses *Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico*. "From 1996 to 2015, the country's per capita GDP growth averaged only 1.2 percent per year. Moreover, this unimpressive figure arguably overestimates Mexico's performance, as it reflects the fact that because of the country's demographic transition, its labor force grew more rapidly than its population during these years (2.2 versus 1.4 percent). … Over the medium term, growth occurs because the labor force increases (in quantity and quality), because there is more investment in physical capital, and because the productivity of labor and capital (total factor productivity – TFP) increases. Decomposing Mexico's growth over this period into these three components, one finds that TFP growth averaged only 0.14 percent annually, without any corrections for the quality of the labor force. Considering increases in schooling (that is, taking into account that workers with more years of schooling can potentially contribute more to output than those with fewer years), yields a negative TFP growth rate of 0.53 percent. … The book points out that the main policies and institutions impeding growth are those related to taxation, labor and social insurance regulations, and enforcement of contracts. By documenting the central relevance of these issues to growth in Mexico, the book is an implicit criticism of the view that good macro, trade, and competition policies, accompanied by investments in education, are by themselves sufficient to bring prosperity to the country." Inter-American Development Bank, July 2018, https://flagships.iadb.org/en/Under-Rewarded-Efforts.

The Society of Actuaries and the Henry J. Kaiser Family Foundation have created Initiative 18|11 to consider ways of holding down US health care spending.

Their first report is "What Can We Do about the Cost of Health Care?" "[H]ealth care in the United States represents 18 percent of the gross domestic product compared with 11 percent in comparable countries, such as the United Kingdom. In dollar terms, the cost of health care here is roughly double that of similar countries. … At the conference, we focused on two key drivers: the price of goods and services and the chronic disease burden. … There was a consensus among the participants that one of the primary reasons for the 18|11 problem is the difference in prices. … [A]pproximately 50 percent of the increase in U.S. expenditures from 1996 to 2013 was due to increases in price and intensity. … Remarkably, 86 percent of health care spending is for patients with one or more chronic conditions. … The cost of chronic diseases goes far beyond the direct amounts spent on these diseases. In the United States, seven out of every 10 deaths are caused by chronic diseases each year. There are indirect costs through lost productivity and an unmeasurable loss in the quality of life and the loss of ability to perform activities of daily living, such as bathing and eating." January 2019, https://www.soa.org/research/topics/initiative-1811.

Collections

The Aspen Institute Economic Study Group has published a collection of 12 papers on the theme *Expanding Economic Opportunity for More Americans: Bipartisan Policies to Increase Work, Wages, and Skills*, edited by Melissa S. Kearney and Amy Ganz. From "A Policy Agenda to Develop Human Capital for the Modern Economy," by Austan Goolsbee, Glenn Hubbard, and Amy Ganz: "We propose a federal grant program to provide new funding to community colleges, contingent on institutional outcomes in degree completion rates and labor market outcomes. We believe a program of a similar scale to the 19th century Morrill Land Grant Program, which dramatically expanded access to higher education for working-class Americans, is needed to ensure our workforce meets the demands of the modern economy. … In 1910, fewer than 10% of Americans had a high school degree. By 1935, nearly 40% of the population had earned their degrees. This inflection point came from substantial new investments in the nation's education resources. We aim to achieve increases of a similar magnitude … by 2030. … We estimate an annual investment of $22 billion." From "Scaling Apprenticeship to Increase Human Capital," by Robert I. Lerman: "[T]he United States has lagged far behind other developed countries—countries like Germany and Switzerland, but also Australia, Canada, and England—in creating apprenticeships. In these countries, apprentices constitute about 2.5–3.0% of the labor force, or about 10 times the U.S. rate. Increasing the availability of apprenticeships would increase youth employment and wages, improve workers' transitions from school to careers, upgrade those skills that employers most value, broaden access to rewarding careers, increase economic productivity, and contribute to positive returns for employers and workers. … The experiences of Australia, Canada, and England demonstrate that scaling apprenticeship is quite possible, even outside countries with a strong tradition of apprenticeship. … Overall, the federal government has devoted less than $30 million (per year) to the Office of Apprenticeship (OA) to supervise, market, regulate, and publicize the system. … Were the United States to spend what Britain spends annually on apprenticeship, adjusting for differences in the size and composition of the labor force, it would provide at least $9 billion per year for apprenticeship. In fact, the British government spends as much on advertising its apprenticeship programs as the entire U.S. budget for apprenticeship." February 2019, https://assets.aspeninstitute.org/content/uploads/2019/01/ESG_Report_Expanding-Economic-Opportunity-for-More-Americans.pdf.

The *Regulatory Review*, from the Penn Program on Regulation, offers a series of nine short essays on "Bringing Expertise to the Gun Debate." From "Gun Regulation Is Costly—and Not the Only Option," by Jennifer Doleac: "But, in general, the effect of gun regulations on public safety is less clear than many advocates on either side think … It is difficult to disentangle the effects of gun laws from the effects of a community's feelings about guns, from a community's motivation to reduce gun violence, or from an increase in gun purchases that often comes before the laws take effect. … Are there other life-saving programs more deserving of these resources? … Summer jobs programs for teens reduce mortality by 18 to 20 percent among participants. This effect is driven by a reduction in young men killed by homicide or suicide. Cognitive behavioral therapy for at-risk young men lowers violent crime arrests by 45 to 50 percent for participants. Access to Medicaid in early childhood decreases suicide by 10 to 15 percent later in life. Mandating that health insurance cover mental health benefits at parity reduces the suicide rate by 5 percent. Access to antidepressants also reduces suicide rates: An increase in antidepressant sales equivalent to one pill per capita reduced suicide by 5 percent. In addition, repealing duty-to-warn laws for mental health providers—which require that they report a patient's violent threats, perhaps causing patients to be less honest—could reduce teen suicides by 8 percent and decrease homicides by 5 percent. … In the war over gun deaths, vast armies have gathered to contest gun regulations, a territory of uncertain value. Meanwhile, other zones of clear value are available and virtually unguarded." From "Reducing Information Asymmetry in the American Gun Market," by Amanda LeSavage: "Suicides constitute two-thirds of annual gun deaths in the United States. Individuals who live in homes with guns are approximately five times more likely to commit suicide by any means and approximately 17 times more likely to commit suicide with a gun than individuals who do not have guns in their homes. … Suicide usually results from an impulsive decision that can come as a surprise even to the victim. If an individual has access to a gun in a moment of such crisis and attempts suicide with that gun, there is an 85 percent chance of death. But less than 10 percent of people who attempt suicide by any other means actually die. That statistic is why the United States, which possesses almost half of the civilian-owned guns that exist worldwide, suffers from an alarmingly high suicide rate." November 5–15, 2018, https://www.theregreview.org/2018/11/05/bringing-expertise-gun-debate.

The Brookings Institution and the Kellogg School of Business hosted a three-paper conference on “Retirement, Pensions, and Social Security.” In their contribution, Robert L. Clark and John B. Shoven write: “The retirement crisis is in no small measure caused by trying to do the impossible. What we mean by this is that it is nearly impossible to finance 30-year retirements with 40-year careers. Yet with today’s average retirement ages (62 for women and 64 for men), we are trying to do just that. If a 64-/62-year-old couple retired today, the survivor of the couple would have about a 40 percent chance of living an additional 30 years. This division of adult life between work and retirement is at the heart of the financial problems of Social Security and state and local pension plans, and it threatens the adequacy of retirement resources for millions of Americans.” January 31, 2019, https://www.brookings.edu/topic/retirement-pensions-social-security.

Realizing Indonesia’s Economic Potential is a 13-chapter book edited by Luis E. Breuer, Jaime Guajardo, and Tidiane Kinda. From “Twenty Years after the Asian Financial Crisis,” by M. Chatib Basri: “Before the AFC [Asian financial crisis], Indonesia’s economy was lauded as a success story of structural transformation in East Asia. Its economy grew by an average of 7.6 percent per year from 1967 to 1996. … The World Bank (1993) cited Indonesia as a member of the newly industrialized economies, together with Malaysia and Thailand. However, the AFC reversed the picture completely, hitting the Indonesian economy hard … Hill (1999) referred to this as the strange and sudden death of a tiger economy.” From “Realizing Indonesia’s Economic Potential: An Overview,” by Luis E. Breuer and Tidiane Kinda: “Home to more than 260 million people, Indonesia is the fourth most populous country in the world and the largest economy in Southeast Asia. With GDP of about US$1 trillion, the country is the world’s sixteenth largest economy and the seventh largest in purchasing-power-parity terms. … Following two decades of socioeconomic progress, Indonesia is well positioned to continue its remarkable transformation. However, important reforms remain needed … These reforms, discussed at length in the book, include raising tax revenues to enhance infrastructure and human capital, streamlining complex regulations, opening up to FDI, and deepening the financial sector while preserving stability.” International Monetary Fund, August 2018, https://www.elibrary.imf.org/doc/IMF071/24870-9781484337141/24870-9781484337141/Other_formats/Source_PDF/24870-9781484355954.pdf.

Economists Speak

Tyler Cowen interviews “Daniel Kahneman on Cutting Through the Noise.” “[L]et me explain what I mean by noise. I mean, just randomness. … I’ll tell you where the experiment from which my current fascination with noise arose. I was working with an insurance company, and we did a very standard experiment. They constructed cases, very routine, standard cases. Expensive cases—we’re not talking of insuring cars. We’re talking of insuring financial firms for risk of fraud. So you have people who are specialists in this. This is what they do. Cases were constructed completely realistically, the kind of thing that people encounter every day. You have 50 people reading a case and putting a dollar value on it. … Suppose you take two people at random, two underwriters at random. You average the premium they set, you take the difference between them, and you divide the difference by the average. By what percentage do people differ? … And there is a common answer that you find, when I just talk to people and ask them, or the executives had the same answer. It's somewhere around 10 percent. That's what people expect to see in a well-run firm. Now, what we found was 50 percent, 5–0, which, by the way, means that those underwriters were absolutely wasting their time, in the sense of assessing risk. … And you find variability within individuals, depending morning, afternoon, hot, cold. A lot of things influence the way that people make judgments: whether they are full, or whether they've had lunch or haven't had lunch affects the judges, and things like that. Now, it's hard to say what there is more of, noise or bias. But one thing is very certain—that bias has been overestimated at the expense of noise. Virtually all the literature and a lot of public conversation is about biases. But in fact, noise is, I think, extremely important, very prevalent. There is an interesting fact—that noise and bias are independent sources of error, so that reducing either of them improves overall accuracy. There is room for … and the procedures by which you would reduce bias and reduce noise are not the same. So that's what I'm fascinated by these days." December 19, 2018; text and audio available at https://medium.com/conversations-with-tyler/tyler-cowen-daniel-kahneman-economics-bias-noise-167275de691f.

Eric Wallach offers "An Interview with Deidre McCloskey, Distinguished Professor Emerita of Economics and of History, UIC." "The central misconception is to think that one can claim the honorable title of 'liberal' if one approves of one form of liberty, such as mutual consent in sexual partners or the ability to drill for oil where you wish, but excludes the other form. Liberty is liberty, and is meaningless by parts. You are still a slave if only on odd days of the month. In Latin America, for example, the word 'liberal,' once meaningful there, has long been appropriated by conservatives who like to drill for oil where they wish, but hate gays. In the United States, it has been appropriated by sweet, or not so sweet, slow socialists, who celebrate diversity, but regard economic liberty as not worthy of much consideration. … I used to think freedom was freedom of speech, freedom of the press, freedom of conscience. Here is what it amounts to: you have to have the right to sow what you wish to, to make shoes or coats, to bake into bread the flour ground from the grain you have sown, and to sell it or not sell it as you wish; for the lathe-operator, the steelworker, and the artist it's a matter of being able to live as you wish and work as you wish and not as they order you." *The Politic*, February 10, 2019, http://thepolitic.org/an-interview-with-deirdre-mccloskey-distinguished-professor-emerita-of-economics-and-of-history-uic.

Hites Ahir interviews "Paul Cheshire on Urban Economics." On the gains from cities: "[M]y assessment is that cities are the most welfare enhancing human innovation in history: they empowered the division of labour, the invention of money, trade and technical inventions like the wheel—let alone government, the arts or culture."

On the study of land values: "Classical economists devoted far more effort to trying to understand the returns to land than they did to labour or capital: it was both the most important asset and the most important factor of production. When Adam Smith was writing only about 12 percent of Europe's population lived in cities and even in the most industrialised country, Britain, the value of agricultural land was about 3 times that of annual GDP. But as the value of other assets increased, interest in land diminished so that by about 1970 really only agricultural economists and a few urban economists were interested in it: and they did not talk to each other. But by 2010 residential property, mostly the land on which houses sat, was worth three times as much as British GDP. By the end of 2013 houses accounted for 61 percent of the UK's net worth: up from 49 percent 20 years ago. Land, now urban land, is valuable, so there is renewed interest." *Global Housing Watch Newsletter*, March 2019, http://unassumingeconomist.com/2019/03/paul-cheshire-on-urban-economics.

Discussion Starters

Scott Lincicome describes "The 'Protectionist Moment' That Wasn't: American Views on Trade and Globalization." "In fact, recent public opinion polling uniformly reveals that, first, foreign trade and globalization are generally popular, and in fact more popular today than at any point in recent history; second, a substantial portion of the American electorate has no strong views on U.S. trade policy or trade agreements; third, and likely due to the previous point, polls on trade fluctuate based on partisanship or the state of the U.S. economy; and, fourth, Americans' views on specific trade policies often shift depending on question wording, especially when the actual costs of protectionism are mentioned. These polling realities puncture the current conventional wisdom on trade and public opinion—in particular, that Americans have turned en masse against trade and globalization…" Cato Institute, *Free Trade Bulletin*, November 2, 2018, https://object.cato.org/sites/cato.org/files/pubs/pdf/ftb-72.pdf.

Michael Beckley considers "The Power of Nations: Measuring What Matters." "What makes some countries more powerful than others? This is the most important question for the study and practice of international relations. … [M]ost scholars measure power in terms of resources, specifically wealth and military assets. The logic of this approach is simple and sound: countries with more wealth and more military assets at their disposal tend to get their way more often than countries with fewer of these resources. Unfortunately, however, most scholars measure resources with gross indicators, such as gross domestic product (GDP); military spending; or the Composite Indicator of National Capability (CINC), which combines data on military spending, troops, population, urban population, iron and steel production, and energy consumption. … Standard gross indicators are not good enough; they are logically unsound and empirically unreliable, severely mischaracterizing the balance of power in numerous cases, including in some of the most consequential geopolitical events in modern history. … The hype about China's rise, however, has been based largely on gross indicators that ignore costs. When costs are accounted for, it becomes clear that the United States' economic and military lead over China is much larger than typically assumed—and the trends are mostly in America's favor. … [T]here is a large literature showing that GDP per capita serves as a reliable proxy for economic and military efficiency. … Military studies also show that the higher a country's GDP per capita, the more efficiently its military fights in battle. The reason is that a vibrant civilian economy helps a country produce advanced weapons, train skillful military personnel, and manage complex military systems. … GDP per capita thus provides a rough but reliable measure of economic and military efficiency. … Combining GDP with GDP per capita thus yields an indicator that accounts for size and efficiency, the two main dimensions of net resources. … Future studies can experiment with ways to improve this measure…" *International Security*, Fall 2018, 43:2, pp. 7–44, https://www.mitpressjournals.org/doi/pdf/10.1162/isec_a_00328.

Giana M. Eckhardt and Susan Dobscha consider "The Consumer Experience of Responsibilization: The Case of Panera Cares." "In this paper, we explore how consumers experience being tasked with solving social issues through their consumption choices. … We do this in the context of Panera Cares, a nonprofit division of Panera Bread Company … Panera Cares self identifies as a conscious capitalist organization … They enact conscious capitalism through a pricing approach which we label conscious pricing. This incorporates elements of pay what you want (PWYW), Pay It Forward (PIF), and traditional charitable donation behavior by asking consumers to pay what they feel is appropriate for their food and drinks based on their support of the social issue of food insecurity … This pricing strategy allows us to understand how consumers put a price on morality. … We demonstrate that consumers feel discomfort with the conscious pricing policy. This discomfort takes three forms: physical, psychological, and philosophical. Consumers have disdain for the embodied experience of dining near the food insecure in the physical space of the café, and they question Panera Cares' motives for engaging in conscious pricing. The food insecure experience discomfort as well. Rather than being empowered via a dignified dining experience, they feel ashamed or uncomfortable when trying to pay what they can for their food. Our findings suggest a pushback against tasked responsibilization." *Journal of Business Ethics*, January 2018, https://link.springer.com/article/10.1007/s10551-018-3795-4.


Begin making plans for the next
AEA ANNUAL MEETING
San Diego, CA
January 3–5, 2020
SHARE. COLLABORATE. LEARN.
AMERICAN ECONOMIC ASSOCIATION
ASSA
ALLIED SOCIAL SCIENCE ASSOCIATIONS
Webcasts of past meetings available at
www.aeaweb.org/webcasts

AT&T LTE
2:02 PM
aeaweb.org
38%
initial default event
ongoing default episode
1980
1985
1990


AIM HIGH.
ACHIEVE MORE.
MAKE A DIFFERENCE.
Whether you are a student, an established economist, or an emerging scholar in your field, our member resources can be an important part of your success
✓ Collaboration
✓ Career Services
✓ Peer Recognition
✓ Learning Resources
✓ Prestigious Research
✓ Member Communications
✓ Member Savings
ADVANCE YOUR CAREER
Starting at only $22, a membership is a smart and easy way to stay abreast of all the latest research and news in economics you need to know about.
AMERICAN ECONOMIC ASSOCIATION
Join or Renew
Your AEA Membership Today!
www.aeaweb.org/membership

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to *JEP* articles, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $22.00, $32.00, or $42.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of

Economic Perspectives

Spring 2019, Volume 33, Number 2

Symposia

Automation and Employment

Daron Acemoglu and Pascual Restrepo, "Automation and New Tasks: How Technology Displaces and Reinstates Labor"

Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb, "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction"

Jeremy Atack, Robert A. Margo, and Paul W. Rhode, "'Automation' of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study"

Hong Cheng, Ruixue Jia, Dandan Li, and Hongbin Li, "The Rise of Robots in China"

Fiscal Policy

Valerie A. Ramey, "Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?"

Pierre Yared, "Rising Government Debt: Causes and Solutions for a Decades-Old Trend"

Alberto Alesina, Carlo Favero, and Francesco Giavazzi, "Effects of Austerity: Expenditure- and Tax-based Approaches"

The Problems of Men

Ariel J. Binder and John Bound, "The Declining Labor Market Prospects of Less-Educated Men"

Courtney C. Coile and Mark G. Duggan, "When Labor's Lost: Health, Family Life, Incarceration, and Education in a Time of Declining Economic Opportunity for Low-Skilled Men"

Kathryn Edin, Timothy Nelson, Andrew Cherlin, and Robert Francis, "The Tenuous Attachments of Working-Class Men"

Features

Samuel Hollander, "Retrospectives: Ricardo on Machinery"

Timothy Taylor, "Recommendations for Further Reading"

AMERICAN ECONOMIC ASSOCIATION