The Journal of Economic Perspectives

A journal of the American Economic Association

Summer 2015

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 R. R. Donnelley Company, Jefferson City, Missouri, 65109, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Summer 2015, Vol. 29, No. 3. 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 $20.00, $30.00, or $40.00 depending on income; for an additional $15.00, you can receive this journal in print. E-reader versions are free. 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 29 • Number 3 • Summer 2015

Symposia

Articles

Features

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 29, Number 3—Summer 2015—Pages 3–30

Why Are There Still So Many Jobs? The History and Future of Workplace Automation[†]

David H. Autor

There have been periodic warnings in the last two centuries that automation and new technology were going to wipe out large numbers of middle class jobs. The best-known early example is the Luddite movement of the early 19th century, in which a group of English textile artisans protested the automation of textile production by seeking to destroy some of the machines. A lesser-known but more recent example is the concern over "The Automation Jobless," as they were called in the title of a *TIME* magazine story of February 24, 1961:

> The number of jobs lost to more efficient machines is only part of the problem. What worries many job experts more is that automation may prevent the economy from creating enough new jobs. . . . Throughout industry, the trend has been to bigger production with a smaller work force. . . . Many of the losses in factory jobs have been countered by an increase in the service industries or in office jobs. But automation is beginning to move in and eliminate office jobs too. . . . In the past, new industries hired far more people than those they put out of business. But this is not true of many of today's new industries. . . . Today's new industries have comparatively few jobs for the unskilled or semiskilled, just the class of workers whose jobs are being eliminated by automation.

Concerns over automation and joblessness during the 1950s and early 1960s were strong enough that in 1964, President Lyndon B. Johnson empaneled a

■ *David H. Autor is Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. From 2009 to 2014, he was Editor of the* Journal of Economic Perspectives.

[†] To access the Data Appendix and disclosure statement, visit
http://dx.doi.org/10.1257/jep.29.3.3

doi=10.1257/jep.29.3.3

"Blue-Ribbon National Commission on Technology, Automation, and Economic Progress" to confront the productivity problem of that period—specifically, the problem that productivity was rising so fast it might outstrip demand for labor. The commission ultimately concluded that automation did not threaten employment: "Thus technological change (along with other forms of economic change) is an important determinant of the precise places, industries, and people affected by unemployment. But the general level of demand for goods and services is by far the most important factor determining how many are affected, how long they stay unemployed, and how hard it is for new entrants to the labor market to find jobs. The basic fact is that technology eliminates jobs, not work" (Bowen 1966, p. 9). However, the Commission took the reality of technological disruption as severe enough that it recommended, as one newspaper (*The Herald Post* 1966) reported, "a guaranteed minimum income for each family; using the government as the employer of last resort for the hard core jobless; two years of free education in either community or vocational colleges; a fully administered federal employment service, and individual Federal Reserve Bank sponsorship in area economic development free from the Fed's national headquarters."

Such concerns have recently regained prominence. In their widely discussed book *The Second Machine Age*, MIT scholars Erik Brynjolfsson and Andrew McAfee (2014, p. 11) offer an unsettling picture of the likely effects of automation on employment:

> Rapid and accelerating digitization is likely to bring economic rather than environmental disruption, stemming from the fact that as computers get more powerful, companies have less need for some kinds of workers. Technological progress is going to leave behind some people, perhaps even a lot of people, as it races ahead. As we'll demonstrate, there's never been a better time to be a worker with special skills or the right education, because these people can use technology to create and capture value. However, there's never been a worse time to be a worker with only 'ordinary' skills and abilities to offer, because computers, robots, and other digital technologies are acquiring these skills and abilities at an extraordinary rate.

Clearly, the past two centuries of automation and technological progress have not made human labor obsolete: the employment-to-population ratio rose during the 20th century even as women moved from home to market; and although the unemployment rate fluctuates cyclically, there is no apparent long-run increase. But those concerned about automation and employment are quick to point out that past interactions between automation and employment cannot settle arguments about how these elements might interact in the future: in particular, the emergence of greatly improved computing power, artificial intelligence, and robotics raises the possibility of replacing labor on a scale not previously observed. There is no fundamental economic law that guarantees every adult will be able to earn a living solely on the basis of sound mind and good character. Whatever the future holds, the present clearly offers a resurgence of automation anxiety (Akst 2013).

In this essay, I begin by identifying the reasons that automation has not wiped out a majority of jobs over the decades and centuries. Automation does indeed substitute for labor—as it is typically intended to do. However, automation also complements labor, raises output in ways that lead to higher demand for labor, and interacts with adjustments in labor supply. Indeed, a key observation of the paper is that journalists and even expert commentators tend to overstate the extent of machine substitution for human labor and ignore the strong complementarities between automation and labor that increase productivity, raise earnings, and augment demand for labor.

Changes in technology do alter the types of jobs available and what those jobs pay. In the last few decades, one noticeable change has been "polarization" of the labor market, in which wage gains went disproportionately to those at the top and at the bottom of the income and skill distribution, not to those in the middle. I will offer some evidence on this phenomenon. However, I will also argue that this polarization is unlikely to continue very far into the foreseeable future.

The final section of this paper reflects on how recent and future advances in artificial intelligence and robotics should shape our thinking about the likely trajectory of occupational change and employment growth. I argue that the interplay between machine and human comparative advantage allows computers to substitute for workers in performing routine, codifiable tasks while amplifying the comparative advantage of workers in supplying problem-solving skills, adaptability, and creativity. The frontier of automation is rapidly advancing, and the challenges to substituting machines for workers in tasks requiring flexibility, judgment, and common sense remain immense. In many cases, machines both substitute for and complement human labor. Focusing only on what is lost misses a central economic mechanism by which automation affects the demand for labor: raising the value of the tasks that workers uniquely supply.

How Automation and Employment Interact

In 1900, 41 percent of the US workforce was employed in agriculture; by 2000, that share had fallen to 2 percent (Autor 2014), mostly due to a wide range of technologies including automated machinery. The mass-produced automobile drastically reduced demand for many equestrian occupations, including blacksmiths and stable hands. Successive waves of earth-moving equipment and powered tools displaced manual labor from construction. In more recent years, when a computer processes a company's payroll, alphabetizes a list of names, or tabulates the age distribution of residents in each Census enumeration district, it is replacing a task that a human would have done in a previous era. Broadly speaking, many—perhaps most—workplace technologies are designed to save labor. Whether the technology is tractors, assembly lines, or spreadsheets, the first-order goal is to substitute mechanical power for human musculature, machine-consistency for human handiwork, and digital calculation for slow and error-prone "wetware."

Given that these technologies demonstrably succeed in their labor saving objective and, moreover, that we invent many more labor-saving technologies all the time, should we not be somewhat surprised that technological change hasn't *already* wiped out employment for the vast majority of workers? Why doesn't automation *necessarily* reduce aggregate employment, even as it demonstrably reduces labor requirements per unit of output produced?

These questions underline an economic reality that is as fundamental as it is overlooked: tasks that cannot be substituted by automation are generally complemented by it. Most work processes draw upon a multifaceted set of inputs: labor and capital; brains and brawn; creativity and rote repetition; technical mastery and intuitive judgment; perspiration and inspiration; adherence to rules and judicious application of discretion. Typically, these inputs *each* play essential roles; that is, improvements in one do not obviate the need for the other. If so, productivity improvements in one set of tasks almost necessarily increase the economic value of the remaining tasks.

An iconic representation of this idea is found in the O-ring production function studied by Kremer (1993).[1] In the O-ring model, failure of any one step in the chain of production leads the entire production process to fail. Conversely, improvements in the reliability of any given link increase the value of improvements in all of the others. Intuitively, if $n - 1$ links in the chain are reasonably likely to fail, the fact that link n is somewhat unreliable is of little consequence. If the other $n - 1$ links are made reliable, then the value of making link n more reliable as well rises. Analogously, when automation or computerization makes some steps in a work process more reliable, cheaper, or faster, this increases the value of the remaining human links in the production chain.

As a contemporary example, consider the surprising complementarities between information technology and employment in banking, specifically the experience with automated teller machines (ATMs) and bank tellers documented by Bessen (2015). ATMs were introduced in the 1970s, and their numbers in the US economy quadrupled from approximately 100,000 to 400,000 between 1995 and 2010. One might naturally assume that these machines had all but eliminated bank tellers in that interval. But US bank teller employment actually rose modestly from 500,000 to approximately 550,000 over the 30-year period from 1980 to 2010 (although given the growth in the labor force in this time interval, these numbers do imply that bank tellers declined as a share of overall US employment). With the growth of ATMs, what are all of these tellers doing? Bessen observes that two forces worked in opposite directions. First, by reducing the cost of operating a bank branch, ATMs indirectly increased the demand for tellers: the number of tellers *per branch* fell by more than a third between 1988 and 2004, but the number of urban bank branches (also encouraged by a wave of

[1] The name of the O-ring production function refers to the 1986 accident of Space Shuttle Challenger, which exploded and crashed back to earth less than two minutes after takeoff, killing its seven crew members. The proximate cause of the Challenger crash was an inexpensive and seemingly inconsequential rubber O-ring seal in one of its booster rockets that failed after hardening and cracking during the icy Florida weather on the night before takeoff.

bank deregulation allowing more branches) rose by more than 40 percent. Second, as the routine cash-handling tasks of bank tellers receded, information technology also enabled a broader range of bank personnel to become involved in "relationship banking." Increasingly, banks recognized the value of tellers enabled by information technology, not primarily as checkout clerks, but as salespersons, forging relationships with customers and introducing them to additional bank services like credit cards, loans, and investment products.

This example should not be taken as paradigmatic; technological change is not necessarily employment-increasing or Pareto-improving. Three main factors can mitigate or augment its impacts. First, workers are more likely to benefit directly from automation if they supply tasks that are complemented by automation, but not if they primarily (or exclusively) supply tasks that are substituted. A construction worker who is expert with a shovel but cannot drive an excavator will generally experience falling wages as automation advances. Similarly, a bank teller who can tally currency but cannot provide "relationship banking" is unlikely to fare well at a modern bank.

Second, the elasticity of labor supply can mitigate wage gains. If the complementary tasks that construction workers or relationship bankers supply are abundantly available elsewhere in the economy, then it is plausible that a flood of new workers will temper any wage gains that would emanate from complementarities between automation and human labor input. While these kinds of supply effects will probably not offset productivity-driven wage gains fully, one can find extreme examples: Hsieh and Moretti (2003) document that new entry into the real estate broker occupation in response to rising house prices fully offsets average wage gains that would otherwise have occurred.

Third, the output elasticity of demand combined with income elasticity of demand can either dampen or amplify the gains from automation. In the case of agricultural products over the long run, spectacular productivity improvements have been accompanied by declines in the share of household income spent on food. In other cases, such as the health care sector, improvements in technology have led to ever-larger shares of income being spent on health. Even if the elasticity of final demand for a given sector is below unity—meaning that the sector shrinks as productivity rises—this does not imply that aggregate demand falls as technology advances; clearly, the surplus income can be spent elsewhere. As passenger cars displaced equestrian travel and the myriad occupations that supported it in the 1920s, the roadside motel and fast food industries rose up to serve the "motoring public" (Jackson 1993). Rising income may also spur demand for activities that have nothing to do with the technological vanguard. Production of restaurant meals, cleaning services, haircare, and personal fitness is neither strongly complemented nor substituted by current technologies; these sectors are "technologically lagging" in Baumol's (1967) phrase. But demand for these goods appears strongly income-elastic, so that rising productivity in technologically leading sectors may boost employment nevertheless in these activities. Ultimately, this outcome requires that the elasticity of substitution between leading and lagging sectors is less than or equal to unity (Autor and Dorn 2013).

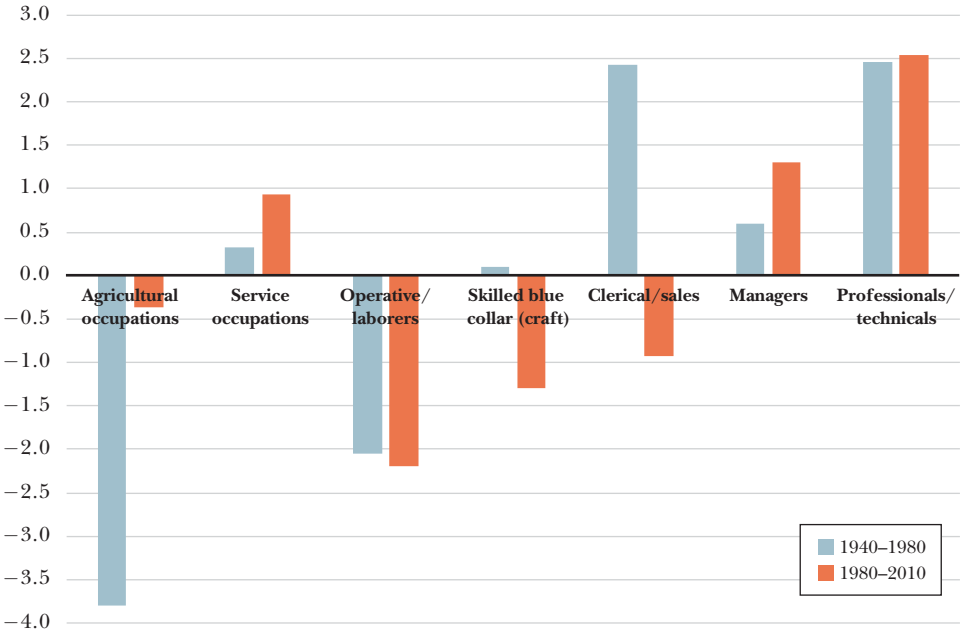
Over the very long run, gains in productivity have not led to a shortfall of demand for goods and services: instead, household consumption has largely kept pace with household incomes. We know this because the share of the population engaged in paid employment has generally risen over (at least) the past century despite vast improvements in material standards of living. An average US worker in 2015 wishing to live at the income level of an average worker in 1915 could roughly achieve this goal by working about 17 weeks per year.[2] Most citizens would not consider this tradeoff between hours and income desirable, however, suggesting that consumption demands have risen along with productivity. Of course, citizens in high-income countries work fewer annual hours, take more vacations, and retire earlier (relative to death) than a century ago—implying that they choose to spend part of their rising incomes on increased leisure. This is clearly good news on many fronts, but does it also imply that consumption demands are approaching satiation? I think not. In high-income countries, consumption and leisure appear to be complements; citizens spend much of their leisure time consuming—shopping, traveling, dining, and, less pleasantly, obtaining medical care.[3]

What about the Marxian concern that automation will immiserate workers by obviating the demand for labor? In simple economic models, this outcome cannot really occur because capital is owned by the economic agents who are presumably also the workers; but, alternatively, the returns could accrue to a narrow subset of agents. Sachs and Kotlikoff (2012) and Sachs, Benzell, and LaGarda (2015) explore multigenerational economic environments in which a burst of robotic productivity can enrich one generation of capital owners at the expense of future generations. These later generations suffer because the fruits of the productivity surge are consumed by the old, while the young face diminished demand for their labor and, in some cases, also experience credit constraints that inhibit their human capital investments. In these models, the fundamental threat is not technology per se but misgovernance; an appropriate capital tax will render the technological advance broadly welfare-improving, as these papers stress. Thus, a key takeaway is that rapid automation may create distributional challenges that invite a broad policy response, a point to which I will return.

[2] Douglas (1930; reproduced in US Bureau of the Census 1949) reports average annual earnings across all sectors in 1915 at $633. Inflating this to 2015 dollars using the US Bureau of Labor Statistics historical Consumer Price Index calculator yields a current dollar equivalent of $14,711. The BLS employment report from April 2015 reports mean weekly private nonfarm earnings of $858. Thus, it would take 17 weeks of work at the average US weekly wage to earn a full-time annual 1915 income.

[3] This outcome is a modern version of the "coal paradox" posed by William Stanley Jevons in his 1865 book *The Coal Question*. Jevons argued that as we became more efficient in mining coal, we would use more of it, not less. Modern environmental economists term this idea the "rebound effect." In this discussion, the broad parallel is that greater efficiency of production of all goods and services means that we consume more of them, not the same or less.

Figure 1
Average Change per Decade in US Occupational Employment Shares for Two Periods: 1940–1980 and 1980–2010



Source: Based on Katz and Margo (2014), table 1.6, panel A, which is based upon the 1920 through 2000 Census of population IPUMS and 2010 American Community Survey.
Notes: Observed long changes in US occupational employment shares over 1940–1980 and 1980–2010 are scaled by the number of intervening decades to yield average change per decade. Occupations are classified into occupational groups based on 1950 occupation codes using the consistent coding of occupations in all years into 1950 codes (the OCC1950 variable) in the IPUMS. Additional details are found in Katz and Margo (2014, p. 46).

Polarization in the US Labor Market

Even if automation does not reduce the quantity of jobs, it may greatly affect the qualities of jobs available. For the three decades or so from the end of World War II and up through the late 1970s, the US experienced rapid automation and technological change—inspiring, for example, the *TIME* magazine story in 1961 and Lyndon Johnson's 1964 National Commission mentioned earlier. While it's difficult to paint an accurate picture of occupational change over a large time interval, Figure 1, which draws from Katz and Margo (2014), provides a high-level overview by depicting the average change per decade in employment for seven broad occupational categories, ranked from lowest to highest paid, for two periods: 1940–1980 and 1980–2010. In the first four decades after World War II, the thrust of occupational change skewed strongly away from physically demanding, dangerous, and

menial work and towards skilled blue- and white-collar work. Agricultural employment declined by almost 4 percentage points per decade. Professional, technical, and managerial employment—the highest skill categories—grew by 3 percentage points per decade (2.5 for the professionals and technicians plus 0.5 for the managers). And among the vast middle group of workers between agriculture (at the bottom) and professional, technical, and managerial (the three groups at the top), service and skilled blue-collar occupations were stable, clerical/sales occupations rose, and operative and laborer occupations fell sharply.

Thus, physically demanding, repetitive, dangerous, and cognitively monotonous work was receding, ushered out by extraordinary productivity gains in agriculture. Rising consumer affluence spurred demand for manufactured goods and leisure complements. Growth of technologically intensive corporations, health care services, and higher education created employment for credentialed professionals and a cadre of supporting clerical, administrative, and sales workers. Though automation was clearly reducing labor demand across a large swath of occupations, it is easy to see why overall job prospects appeared broadly favorable during this period.

But after the late 1970s, these favorable winds slowed and in some cases reversed. While jobs at the top of the skill ladder—professional, technical, and managerial occupations—grew even more rapidly between 1980 and 2010 than in the four decades prior, positive occupational shifts outside of these categories mostly halted. Skilled blue-collar occupations shrank rapidly and clerical and sales occupations—the vulnerable "production jobs" of the information age—sharply reversed course. While physically demanding operative and laborer jobs continued to atrophy, low-paid personal services began absorbing an increasing share of noncollege labor. By this time, the vast movement away from agricultural work had already played out.

Many forces distinguish the labor markets of these two epochs of 1940–1980 and 1980–2010: a partial list would include changes in the relative supply of college and noncollege labor, rising trade penetration, offshoring, and globalization of production chains, declines in labor union penetration, the changing "bite" of the minimum wage, and certain shifts in tax policy. Of course, many of these factors combine and interact as well such that attributing changes to a single cause would be foolish. However, my focus here is on the effects of technological change, and especially information technology, on employment and occupations (and later wages). To understand the role that information technology has played (and may play), it is useful to start from first principles: What do computers do? And how does their widespread adoption change what workers do?

Fundamentally, computers follow procedures meticulously laid out by programmers. The typical pattern has been that for a computer to accomplish a task, a programmer must first fully understand the sequence of steps required to perform that task, and then must write a program that, in effect, causes the machine to simulate these steps precisely. (The field of machine learning, discussed below, provides an interesting exception to this process.) When a computer processes a company's

payroll, alphabetizes a list of names, or tabulates the age distribution of residents in each Census enumeration district, it is "simulating" a work process that would, in a previous era, have been done by humans using nearly identical procedures. The principle of computer simulation of workplace tasks has not fundamentally changed since the dawn of the computer era—but its cost has. An ingenious 2007 paper by William Nordhaus estimates that the cost of performing a standardized set of computations has fallen by at least 1.7 trillion-fold since the manual computing era, with most of that decline occurring since 1980. Thus, firms have strong economic incentives to substitute ever-cheaper computing power for relatively expensive human labor. What are the effects?

One first-order effect is, of course, substitution. As the price of computing power has fallen, computers and their robot cousins have increasingly displaced workers in accomplishing explicit, codifiable tasks. In Autor, Levy, and Murnane (2003), my coauthors and I label these activities as "routine tasks," not because they are mundane, but because they can be fully codified and hence automated (see Levy and Murnane 2004 for many examples). Routine tasks are characteristic of many middle-skilled cognitive and manual activities: for example, the mathematical calculations involved in simple bookkeeping; the retrieving, sorting, and storing of structured information typical of clerical work; and the precise executing of a repetitive physical operation in an unchanging environment as in repetitive production tasks. Because core tasks of these occupations follow precise, well-understood procedures, they are increasingly codified in computer software and performed by machines. This force has led to a substantial decline in employment in clerical, administrative support, and to a lesser degree, in production and operative employment.

But the scope for this kind of substitution is bounded because there are many tasks that people understand tacitly and accomplish effortlessly but for which neither computer programmers nor anyone else can enunciate the explicit "rules" or procedures. I have referred to this constraint as Polanyi's paradox, named after the economist, philosopher, and chemist who observed in 1966, "We know more than we can tell" (Polanyi 1966; Autor 2015). When we break an egg over the edge of a mixing bowl, identify a distinct species of birds based on a fleeting glimpse, write a persuasive paragraph, or develop a hypothesis to explain a poorly understood phenomenon, we are engaging in tasks that we only tacitly understand how to perform. Following Polanyi's observation, the tasks that have proved most vexing to automate are those demanding flexibility, judgment, and common sense—skills that we understand only tacitly.[4]

Polanyi's paradox also suggests *why* high-level reasoning is straightforward to computerize and certain sensorimotor skills are not. High-level reasoning uses a set

[4] Computer scientists often refer to this phenomenon as Moravec's paradox, after Moravec (1988) who wrote, "[I]t is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."

of formal logical tools that were developed specifically to address formal problems: for example, counting, mathematics, logical deduction, and encoding quantitative relationships. In contrast, sensorimotor skills, physical flexibility, common sense, judgment, intuition, creativity, and spoken language are capabilities that the human species evolved, rather than developed. Formalizing these skills requires reverse-engineering a set of activities that we normally accomplish using only tacit understanding. Hoffman and Furcht (2014) discuss the challenge that Polanyi's paradox poses for scientific innovation more broadly.

If computers largely substitute for routine tasks, how do we characterize the nonroutine tasks for which they do not substitute? In Autor, Levy, and Murnane (2003), we distinguish two broad sets of tasks that have proven stubbornly challenging to computerize. One category includes tasks that require problem-solving capabilities, intuition, creativity, and persuasion. These tasks, which we term "abstract," are characteristic of professional, technical, and managerial occupations. They employ workers with high levels of education and analytical capability, and they place a premium on inductive reasoning, communications ability, and expert mastery. The second broad category includes tasks requiring situational adaptability, visual and language recognition, and in-person interactions—which we call "manual" tasks. Manual tasks are characteristic of food preparation and serving jobs, cleaning and janitorial work, grounds cleaning and maintenance, in-person health assistance by home health aides, and numerous jobs in security and protective services. These jobs tend to employ workers who are physically adept and, in some cases, able to communicate fluently in spoken language. While these activities are not highly skilled by the standards of the US labor market, they present daunting challenges for automation. Equally noteworthy, many outputs of these manual task jobs (haircuts, fresh meals, housecleaning) must be produced and performed largely on-site or in person (at least for now), and hence these tasks are not subject to outsourcing. The potential supply of workers who can perform these jobs is very large.

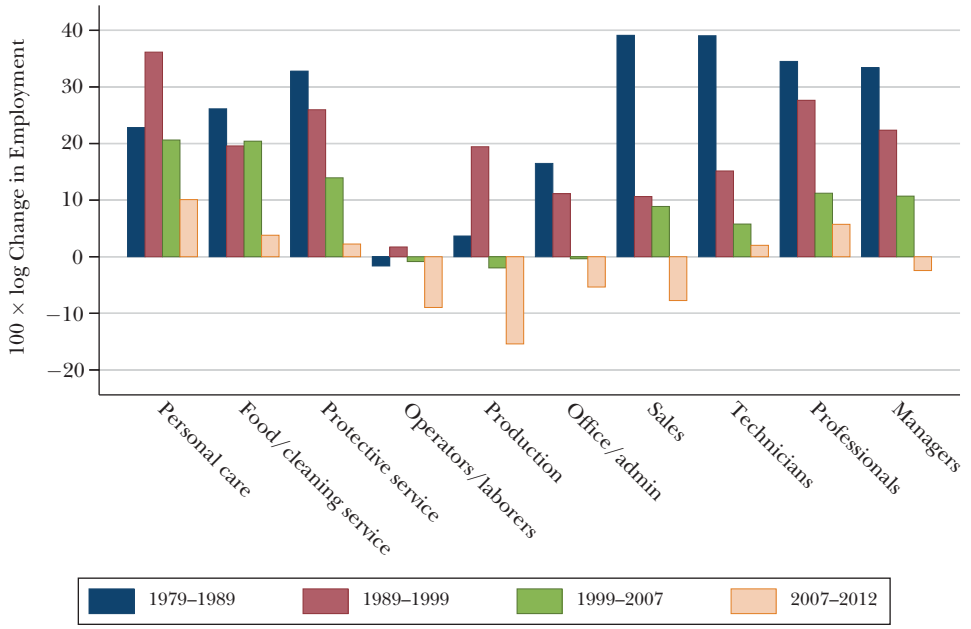
Because jobs that are intensive in either abstract or manual tasks are generally found at opposite ends of the occupational skill spectrum—in professional, managerial, and technical occupations on the one hand, and in service and laborer occupations on the other—this reasoning implies that computerization of "routine" job tasks may lead to the simultaneous growth of high-education, high-wage jobs at one end and low-education, low-wage jobs at the other end, both at the expense of middle-wage, middle education jobs—a phenomenon that Goos and Manning (2003) called "job polarization." A large body of US and international evidence confirms the presence of employment polarization at the level of industries, localities, and national labor markets (Autor, Katz, and Kearney 2006, 2008; Goos and Manning 2007; Autor and Dorn 2013; Michaels, Natraj, and Van Reenen 2014; Goos, Manning, and Salomons 2014; Graetz and Michaels 2015; Autor, Dorn, and Hanson 2015).[5]

[5] Mishel, Shierholz, and Schmitt (2013) offer an extended, and for the most part extremely careful, critique of the literature on technological change, employment, and wage inequality. Their paper argues

Figure 2

Change in Employment by Major Occupational Category, 1979–2012

(the y-axis plots 100 times log changes in employment, which is nearly equivalent to percentage points for small changes)



Sources: Author using data from the 1980, 1990, and 2000 Census IPUMS files, American Community Survey combined file 2006–2008, and American Community Survey 2012. The sample includes the working-age (16–64) civilian noninstitutionalized population. Employment is measured as full-time equivalent workers.
Notes: Figure 2 plots percentage point changes in employment (more precisely, the figure plots 100 times log changes in employment, which is close to equivalent to percentage points for small changes) by decade for the years 1979–2012 for ten major occupational groups encompassing all of US nonagricultural employment. Agricultural occupations comprise no more than 2.2 percent of employment in this time interval, so this omission has a negligible effect.

Figure 2 illustrates this pattern for the United States by plotting percentage point changes in employment by decade for the years 1979–2012 for ten major occupational groups encompassing all of US nonagricultural employment. (More

that the growth of low-wage service employment does not commence in the United States until the 2000s, a finding that is at odds with all other work using contemporary occupation codes of which I am aware (including the Bureau of Labor Statistic's own tabulations of Occupational Employment Statistics data for this time period provided in Alpert and Auyer 2003, table 1). At a methodological level, work in this area always requires adjustments and judgment calls in comparing occupational data across Census years, but the adjustments that Mishel et al. apply to the data generate occupational patterns that appear anomalous. Substantively, I believe the main issue is not whether *employment polarization* has occurred—on this, the evidence appears unambiguous—but the extent to which these occupational employment shifts are helpful for understanding *wage polarization* or wage inequality more broadly.

precisely, the figure plots 100 times log changes in employment, which are close to equivalent to percentage points for small changes. Agricultural occupations comprise no more than 2.2 percent of employment in this time interval, so this omission has a negligible effect.) These ten occupations can be divided into three groups. On the right-hand side of the figure are managerial, professional, and technical occupations, which are highly educated and highly paid. Moving leftward, the next four columns display employment growth in middle-skill occupations, comprising sales; office and administrative support; production, craft and repair; and operator, fabricator, and laborer. The leftmost three columns of Figure 2 depict employment trends in service occupations, defined by the Census Bureau as jobs that involve helping, caring for, or assisting others. The majority of workers in service occupations have no post-secondary education, and average hourly wages in service occupations are in most cases below the other seven occupational categories.

As Figure 2 illustrates, the rapid employment growth in both high- and low-education jobs has substantially reduced the share of employment accounted for by "middle-skill" jobs. In 1979, the four middle-skill occupations (sales; office and administrative workers; production workers; and operatives) accounted for 60 percent of employment. In 2007, this number was 49 percent, and in 2012, it was 46 percent. The employment share of service occupations was essentially flat between 1959 and 1979, and so their rapid growth since 1980 marks a sharp trend reversal (Autor and Dorn 2013).

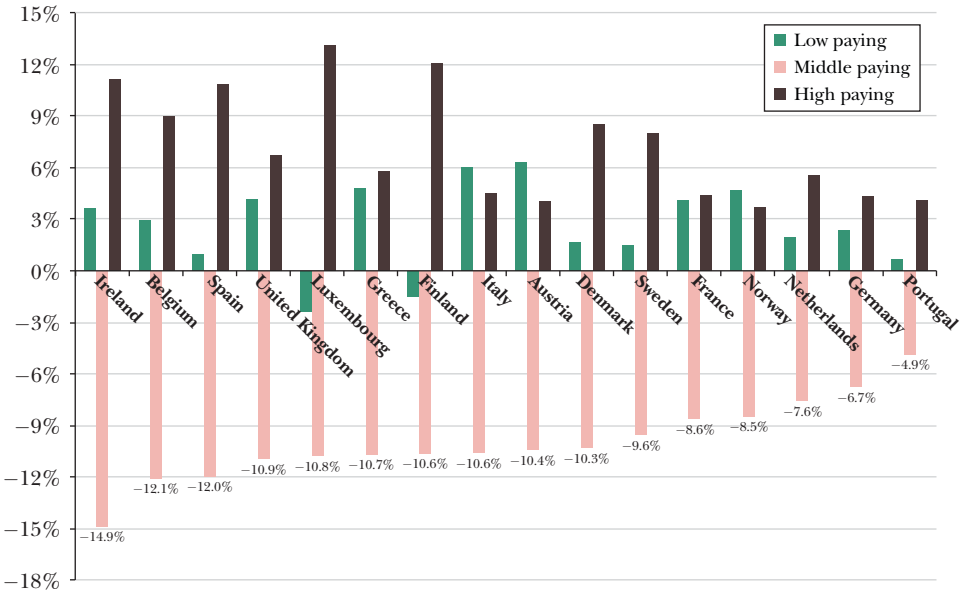
The polarization of employment across occupations is not unique to the United States. Figure 3 plots changes in the share of employment between 1993 and 2010 within three broad sets of occupations—low-, middle-, and high-wage—covering all nonagricultural employment in 16 European Union economies. In all countries, middle-wage occupations declined as a share of employment while both high-wage and low-wage occupations increased their shares of employment over this 17-year period. While the US and EU data are not precisely comparable, the US economy would fall roughly in the middle of the pack of this set of countries in terms of its employment polarization. The comparability of these occupational shifts across a large set of developed countries makes it likely that a common set of forces contributes to these shared labor-market developments. Simultaneously, the substantial differences among countries underscores that no single factor or common cause explains the diversity of experiences across the United States and the European Union.

Does Employment Polarization Lead to Wage Polarization?

From the barbell shape of occupational employment growth depicted in Figures 2 and 3, one might surmise that occupational polarization would also catalyze wage polarization—that is, rising relative wages in both high-education, abstract task-intensive jobs and in low-education, manual task-intensive jobs. However, this

Figure 3

Change in Occupational Employment Shares in Low, Middle, and High-Wage Occupations in 16 EU Countries, 1993–2010



Source: Goos, Manning, and Salomons (2014, table 2).

Notes: High-paying occupations are corporate managers; physical, mathematical, and engineering professionals; life science and health professionals; other professionals; managers of small enterprises; physical, mathematical, and engineering associate professionals; other associate professionals; life science and health associate professionals. Middle-paying occupations are stationary plant and related operators; metal, machinery, and related trade work; drivers and mobile plant operators; office clerks; precision, handicraft, craft printing, and related trade workers; extraction and building trades workers; customer service clerks; machine operators and assemblers; and other craft and related trade workers. Low-paying occupations are laborers in mining, construction, manufacturing, and transport; personal and protective service workers; models, salespersons, and demonstrators; and sales and service elementary occupations.

reasoning does not take into account the role played by the three mitigating forces discussed above: complementarity, demand elasticity, and labor supply.

Let's first consider the effect of computerization on wages in abstract task-intensive occupations such as managerial, professional, and technical occupations. These occupations all draw upon large bodies of constantly evolving expertise: for example, medical knowledge, legal precedents, sales data, financial analysis, programming languages, and economic statistics. Information technology and computerization should strongly complement workers performing abstract task-intensive jobs. By dramatically lowering the cost and increasing the scope of information and analysis available to them, computerization enables workers performing abstract tasks to further specialize in their area of comparative advantage, with less time spent on acquiring and crunching information, and more time spent on interpreting and applying it. By the same token, information technology substitutes for many of the

support occupations that these professions employ, including medical secretaries, paralegals, and research assistants. Similarly, computerization and information technology appears to allow "delayering" of management structures (Caroli and Van Reenen 2001). Arguably, many of the middle managers displaced by delayering performed routine information-processing tasks.

If demand for the output of abstract task-intensive activities is inelastic, these productivity gains might work to lower expenditure on these outputs, which could mitigate wage gains. However, all outward evidence suggests that as technology has boosted the output of the professions, demand for their services has more than kept pace. Health care is an obvious example, but one can readily make similar arguments about finance, law, engineering, research, and design.

What about reactions from labor supply? If workers could quickly move into the highly educated professions, such a shift would mute earnings gains. But of course, many professions require both college and graduate degrees, so the production pipeline for new entrants is at least five to ten years in length. Indeed, young US adults, particularly US males, have responded with remarkable sluggishness to the rising educational premium over the last 30 years (Autor 2014). For example, in 1975, approximately 40 percent of hours worked by males with fewer than ten years of experience (a group that has made the more recent choices about college) were supplied by those with a college education. Forty years later in 2005, this share was almost unchanged. For women workers with less than ten years of experience, the share of total hours worked by those with a college education was 42 percent in 1982 but had risen to 53 percent by 2005. In the last decade, the share of hours worked by those with less than ten years of experience and a college degree has increased for both men and women: in 2012, it was 52 percent of hours for men in this group and 62 percent of the hours for women. Thus, while the stock of workers with college and graduate degrees has certainly grown, the supply response has not been nearly large enough to swamp the contemporaneous movements in labor demand.

Workers in abstract task-intensive occupations therefore benefit from information technology via a virtuous combination of strong complementarities between routine and abstract tasks, elastic demand for services provided by abstract task-intensive occupations, and inelastic labor supply to these occupations over the short and medium term. In combination, these forces mean that information technology should raise earnings in occupations that make intensive use of abstract tasks and among workers who intensively supply them.

These same synergies do not apply to jobs that are intensive in manual tasks, such as janitors and cleaners, vehicle drivers, security guards, flight attendants, food service workers, and home health aides. Most manual task-intensive occupations are only minimally reliant on information or data processing for their core tasks, and involve only limited opportunities for either direct complementarity or substitution.[6]

[6] There are partial exceptions to this generalization: global positioning system satellites and scheduling software allows truckers and delivery services to minimize wasted mileage; calendar, contact, and billing

Aggregate evidence suggests that final demand for manual task-intensive work—services in particular—is relatively *price* inelastic (Baumol 1967; Autor and Dorn 2013). If so, productivity gains in manual task-intensive occupations that tend to reduce their price per unit of service provided will not necessarily raise expenditure on their outputs. On the other hand, demand for manual task-intensive work appears to be relatively *income* elastic (Clark 1951; Mazzorali and Ragusa 2013), so that rising aggregate incomes will tend to increase demand for these activities. New technology and productivity growth in other areas may therefore *indirectly* raise demand for manual task-intensive occupations by increasing societal income.

Labor supply to manual task-intensive occupations is intrinsically elastic, due to their generally low education and training requirements. This insight does not preclude the possibility that wages in manual tasks will rise, at least to some extent. As Baumol (1967) observed, even absent productivity growth in technologically lagging occupations, wages in these occupations *must* rise over time with societal income to compensate workers for *not* entering other sectors (again, assuming that demand for these activities is relatively inelastic). But it does suggest that wage increases in these jobs will be restrained to some extent by the labor supply response, including from workers displaced in other sectors of the economy.

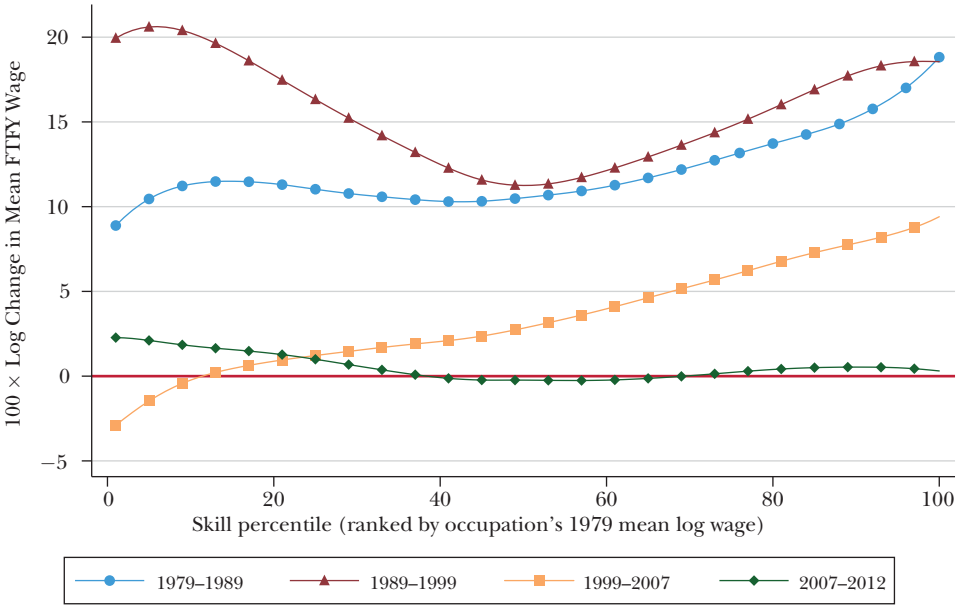
Overall, manual task-intensive activities are at best weakly complemented by computerization, do not benefit from elastic final demand, and face elastic labor supply that tempers demand-induced wage increases. Thus, while information technology has strongly contributed to *employment* polarization measured in quantity of jobs, we would not generally expect these employment changes to culminate in a corresponding *wage* polarization except perhaps at certain times or in certain labor markets. Indeed, in Autor and Dorn (2013), we present evidence that wages for manual-task occupations rose during the 1990s when labor markets were extremely tight, but after 2000, the expansion of manual task-intensive service occupations accelerated while wages in these occupations fell.

For insight about the evolution of wage patterns, consider Figure 4. The horizontal axis of this figure is based on a ranking of all 318 detailed occupations from lowest to highest by their initial skill level, as measured by its 1979 mean hourly occupational wage. These categories are weighted by their initial size, and then grouped into 100 bins of equal size. The vertical axis of the figure then shows the percentage change in wages over each of four periods across the skill distribution—with the line smoothed for clarity. (Again, more precisely, the figure plots 100 times log changes in employment, which is nearly equivalent to percentage points for small changes.)

The right-hand two-thirds of Figure 4 look like the plots of employment polarization. From 1979 through 2007, wages rose consistently across the high-skill portion

software assists home health workers to manage data more effectively; and computerized ordering systems enable food service workers to rapidly tally customer tabs. In a few years time, many retailers may employ RFID "chip" technology that will scan purchases without needing a human checkout cashier at all.

Figure 4
Changes in Mean Wages by Occupational Skill Percentile among Full-Time, Full-Year (FTFY) Workers, 1979–2012
(the y-axis plots 100 times log changes in employment, which is nearly equivalent to percentage points for small changes)



Sources: Author, calculated using 1980, 1990, and 2000 Census IPUMS files; American Community Survey combined file 2006–2008, American Community Survey 2012.
Notes: The figure plots changes in mean log wages over each period, by 1979 occupational skill percentile rank using a locally weighted smoothing regression (bandwidth 0.8 with 100 observations), where skill percentiles are measured as the employment-weighted percentile rank of an occupation's mean log wage in the Census IPUMS 1980 5 percent extract. The sample includes the working-age (1–64) civilian non-institutionalized population with 48+ annual weeks worked and 35+ usual weekly hours. Weekly wages are calculated as annual earnings divided by weeks worked.

of the figure, which is disproportionately made up of the abstract task-intensive categories of professional, technical, and managerial occupations. By contrast, wage growth in the middle-skill, typically routine task-intensive occupations was less rapid and generally decelerated over time. For the low-education, manual task-intensive occupations heavily represented on the left-hand side of Figure 4, in the 1980s, wage growth was a little *more* rapid than in the middle-skill occupations—and in the 1990s, it was much more rapid. However, that changed in the 2000s: while Figure 2 showed that employment growth in these occupations exceeded that in all other categories between 1999 and 2007, Figure 4 shows wage growth was generally negative in the low-skill percentiles, lower than in all other categories (Mishel, Shierholz, and Schmitt 2013). During this time period, my strong hunch is that the explanation is that declining employment in middle-skill routine task-intensive

jobs led middle-skill workers—including new entrants, those displaced from routine task-intensive jobs, and those who lost jobs during recession—to enter manual task-intensive occupations instead (Smith 2013; Cortes, Jaimovich, Nekarda, and Siu 2014; Foote and Ryan 2014).

A final set of facts illustrated by Figure 4 is that overall wage growth was anemic throughout the 2000s, even prior to the Great Recession. Between 1999 and 2007, real wage changes were negative below approximately the 15th percentile, and were below 5 percentage points up to the 70th percentile of the distribution. Indeed, wage growth was greater at all percentiles during both the 1980s and 1990s than in the *pre-recession* 2000s.[7] Of course, wage growth was essentially zero at all percentiles from 2007 to 2012.

Why are the rapidly rising earnings of the top 1 percent (as discussed in Atkinson, Piketty, and Saez 2011, for example) not strongly evident in Figure 4? One reason reflects substance; another is an artifact of the data. Substantively, the plot depicts changes in earnings by *occupational* percentile rather than *wage* percentile. Wage growth by occupational percentile is less concentrated than wage growth across wage percentiles because the highest earners are found across a variety of occupations. In addition, the very highest percentiles of earnings are censored in public use Census and American Community Survey data files, which further masks earnings gains at extreme quantiles.

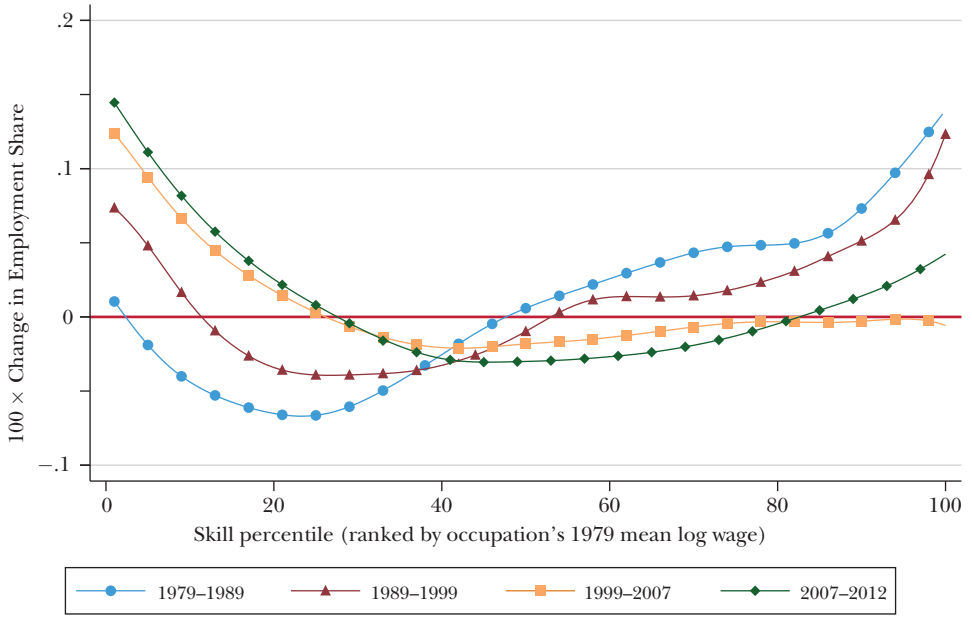
The Recent Slowdown in the Growth of High-Skill Occupations

The hypothesis that automation and information technology has led to occupational and, to a lesser degree, wage polarization in the US labor force can explain some key features of the US and the cross-national data. But reality invariably proves more complicated than any single theory anticipates.

For my thesis linking technological change to occupational change, one concern is the unexplained deceleration of employment growth in abstract task-intensive occupations after 2000 (Beaudry, Green, and Sand 2014, forthcoming; Mishel, Shierholz, and Schmitt 2013). Figure 5 follows the format of Figure 4 but instead of showing (approximate) percentage changes in wages on the vertical axis, it shows percentage changes in the employment share of the jobs ranked by their skill level in 1979. Since the sum of shares must equal one at any time period, the changes in these shares across the decades must total zero, and thus, the height at each skill percentile measures the growth in each occupation's employment relative to the whole.

[7] Because the 2000–2007 interval is two years shorter than the 1979–1989 period, one should multiply the later changes by 1.25 to put them on the same temporal footing. But even after making such an adjustment, wage growth was still considerably weaker at all percentiles from 2000–2007 than in the earlier two decades.

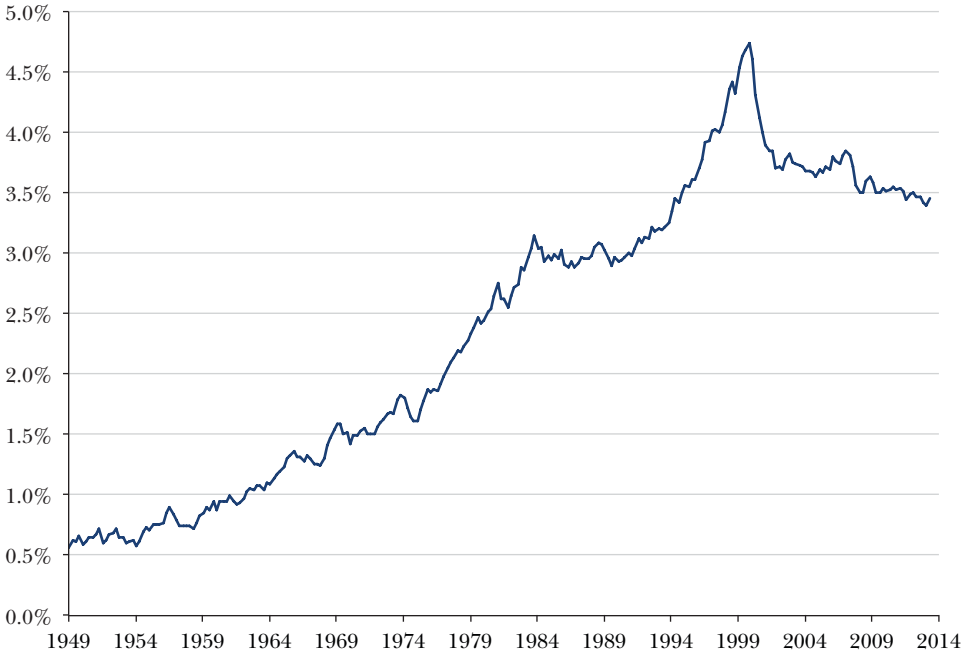
Figure 5
Smoothed Employment Changes by Occupational Skill Percentile, 1979–2012



Sources: Author, calculated using 1980, 1990, and 2000 Census Integrated Public Use Microdata Series (IPUMS) files; American Community Survey combined file 2006–2008, American Community Survey 2012.
Notes: The figure plots changes in employment shares by 1980 occupational skill percentile rank using a locally weighted smoothing regression (bandwidth 0.8 with 100 observations), where skill percentiles are measured as the employment-weighted percentile rank of an occupation's mean log wage in the Census IPUMS 1980 5 percent extract. Employment in each occupation is calculated using workers' hours of annual labor supply times the Census sampling weights. Consistent occupation codes for Census years 1980, 1990, and 2000, and 2008 are from Autor and Dorn (2013).

Figure 5 contributes three nuances to the occupational polarization story above. First, the pace of employment gains in low-wage, manual task-intensive jobs has risen successively across periods, as shown at the left-hand side of the figure. Second, the occupations that are losing employment share appear to be increasingly drawn from higher ranks of the occupational distribution. For example, the highest ranked occupation to lose employment share during the 1980s lay at approximately the 45th percentile of the skill distribution. In the final two subperiods, this rank rose still further to above the 75th percentile—suggesting that the locus of displaced middle-skill employment is moving into higher-skilled territories. Third, growth of high-skill, high-wage occupations (those associated with abstract work) decelerated markedly in the 2000s, with no relative growth in the top two deciles of the occupational skill distribution during 1999 through 2007, and only a modest recovery between 2007 and 2012. Stated plainly, the growth of occupational employment across skill levels looks U-shaped earlier in the period, with gains at low-skill and high-skill levels. By the 2000s, the pattern of occupational employment across

Figure 6
Private Fixed Investment in Information Processing Equipment and Software as a Percentage of Gross Domestic Product, 1949–2014



Source: FRED, Federal Bank of St. Louis. http://research.stlouisfed.org/fred2/graph/?g=GXc (accessed 8/3/2014).

skill levels began to resemble a downward ramp. In Autor (2015), I present a more detailed breakdown of these patterns, and in particular suggest that the set of abstract task-intensive jobs is not growing as rapidly as the potential supply of highly educated workers.

What explains the slowing growth of abstract task-intensive employment? One interpretation is that automation, information technology, and technological progress in general are encroaching upward in the task domain and beginning to substitute strongly for the work done by professional, technical, and managerial occupations. While one should not dismiss this possibility out of hand, it doesn't fit well with the pattern of computer and software investment. If information technology is increasingly replacing workers high in the skill distribution, one would expect a surge of corporate investment in computer hardware and software. Instead, Figure 6 shows that in early 2014, information processing equipment and software investment was only 3.5 percent of GDP, a level last seen in 1995 at the outset of the "dot-com" era. To me, the evidence in Figure 6 suggests a temporary dislocation of demand for information technology capital during the latter half of the 1990s, followed by a sharp correction after 2000. I suspect that the huge falloff in

information investment may have dampened innovative activity and demand for high-skilled workers more broadly.

As noted earlier, technological change is far from the only factor affecting US labor markets in the last 15 years. For example, the deceleration of wage growth and changes in occupational patterns in the US labor market after 2000, and further after 2007, is surely associated to some extent with two types of macroeconomic events. First, there are the business cycle effects—the bursting of the "dot-com" bubble in 2000, and the collapse of the housing market and the ensuing financial crisis in 2007–2008—both of which curtailed investment and innovative activity. Second, there are the employment dislocations in the US labor market brought about by rapid globalization, particularly the sharp rise of import penetration from China following its accession to the World Trade Organization in 2001 (Autor, Dorn, and Hanson 2013; Pierce and Schott 2012; Acemoglu, Autor, Dorn, Hanson, and Price forthcoming). China's rapid rise to a premier manufacturing exporter had far-reaching impacts on US workers, reducing employment in directly import-competing US manufacturing industries and depressing labor demand in both manufacturing and nonmanufacturing sectors that served as upstream suppliers to these industries.

Of course, these forces are in various ways linked with the spread of automation and technology. Advances in information and communications technologies have changed job demands in US workplaces directly and also indirectly, by making it increasingly feasible and cost-effective for firms to source, monitor, and coordinate complex production processes at disparate locations worldwide and altering competitive conditions for US manufacturers and workers. This multidimensional complementarity among causal factors makes it both conceptually and empirically difficult to isolate the "pure" effect of any one factor.

Polanyi's Paradox: Will It Be Overcome?

Automation, complemented in recent decades by the exponentially increasing power of information technology, has driven changes in productivity that have disrupted labor markets. This essay has emphasized that jobs are made up of many tasks and that while automation and computerization can substitute for some of them, understanding the interaction between technology and employment requires thinking about more than just substitution. It requires thinking about the range of tasks involved in jobs, and how human labor can often complement new technology. It also requires thinking about price and income elasticities for different kinds of output, and about labor supply responses.

The tasks that have proved most vexing to automate are those demanding flexibility, judgment, and common sense—skills that we understand only tacitly. I referred to this constraint above as Polanyi's paradox. In the past decade, computerization and robotics have progressed into spheres of human activity that were considered off limits only a few years earlier—driving vehicles, parsing legal documents, even

performing agricultural field labor. Is Polanyi's paradox soon to be at least mostly overcome, in the sense that the vast majority of tasks will soon be automated?[8]

My reading of the evidence suggests otherwise. Indeed, Polanyi's paradox helps to explain what has *not* yet been accomplished, and further illuminates the paths by which more *will* ultimately be accomplished. Specifically, I see two distinct paths that engineering and computer science can seek to traverse to automate tasks for which we "do not know the rules": environmental control and machine learning. The first path circumvents Polanyi's paradox by regularizing the environment, so that comparatively inflexible machines can function semi-autonomously. The second approach inverts Polanyi's paradox: rather than teach machines rules that we do not understand, engineers develop machines that attempt to infer tacit rules from context, abundant data, and applied statistics.

Environmental Control

Most automated systems lack flexibility—they are brittle. Modern automobile plants, for example, employ industrial robots to install windshields on new vehicles as they move through the assembly line. But aftermarket windshield replacement companies employ technicians, not robots, to install replacement windshields. Evidently, the tasks of removing a broken windshield, preparing the windshield frame to accept a replacement, and fitting a replacement into that frame demand more real-time adaptability than any contemporary robot can cost-effectively approach.

The distinction between assembly line production and the in-situ repair highlights the role of environmental control in enabling automation. Engineers can in some cases radically simplify the environment in which machines work to enable autonomous operation, as in the familiar example of a factory assembly line. Numerous examples of this approach to environmental regularization are so ingrained in daily technology that they escape notice, however. To enable the operation of present-day automobiles, for example, humanity has adapted the naturally occurring environment by leveling, re-grading, and covering with asphalt a nontrivial percentage of the earth's land surface.[9]

The ongoing automation of warehouses provides another example. Large online retailers, such as Amazon.com, Zappos.com, and Staples, operate systems of warehouses that have traditionally employed legions of dexterous, athletic "pickers," who run and climb through shelves of typically non-air-conditioned warehouses to locate, collect, box, label, and ship goods. There is at present no cost-effective robotic

[8] For a glimpse of the view that just about anything can now be computerized, see the widely cited (albeit unpublished) article by the economists Carl Frey and Michael Osborne, who write (2013, p. 24) that, "recent developments in ML [machine learning] and MR [mobile robotics], building upon big data, allow for pattern recognition, and thus enable computer capital to rapidly substitute for labour across a wide range of non-routine tasks. Yet some inhibiting engineering bottlenecks to computerization persist. Beyond these bottlenecks, however, we argue that it is largely already technologically possible to automate almost any task, provided that sufficient amounts of data are gathered for pattern recognition."

[9] According to Wikipedia, so-called impervious surfaces (mostly roads and parking lots) cover 43,000 square miles of land in the lower 48 United States—roughly equal to the land area of the state of Ohio (http://en.wikipedia.org/wiki/Impervious_surface, accessed 8/4/2014).

facsimile for these human pickers. The job's steep requirements for flexibility, object recognition, physical dexterity, and fine motor coordination are too formidable.

But large components of warehousing can be automated, as demonstrated by Kiva Systems, a robotic warehousing startup that was purchased by Amazon in 2012. The core of the Kiva system is a dispatch program that oversees the flow of all goods through the warehouse, coordinating the work of robots, which carry shelves, with the work of humans. As objects arrive at the facility for stocking, the dispatch software directs robots to transport and line up empty shelves to a loading area, where *human* stockers place merchandise on shelves. Robots then carry the loaded shelves back to a storage warehouse, where the dispatch software directs their placement to optimize product availability for expected product demand. As new orders arrive, the dispatch software sends robots to retrieve shelves and lines them up in a packing area. Then a human picker, directed by a laser pointer controlled by the dispatch software, takes objects from the assembled shelves, packs them in shipping boxes, applies a shipping label, and drops the package in a chute for delivery. As items are picked, the robots take the shelves away until needed again for packing or restocking. Thus, in a Kiva-operated warehouse, robots handle only the routine task of moving shelves across a level surface; workers handle merchandise; and the dispatch software coordinates the activity.

While Kiva Systems provides a particularly clear example of exploiting environmental control to extend the reach of automation, the same principle is often lurking behind more sophisticated packaging. Perhaps the least recognized—and most mythologized—is the self-driving Google Car. Computer scientists sometimes remark that the Google car does not drive on roads, but rather on maps. A Google car navigates through the road network primarily by comparing its real-time audio-visual sensor data against painstakingly hand-curated maps that specify the exact locations of all roads, signals, signage, and obstacles. The Google car adapts in real time to obstacles, such as cars, pedestrians, and road hazards, by braking, turning, and stopping. But if the car's software determines that the environment in which it is operating differs from the environment that has been preprocessed by its human engineers—when it encounters an unexpected detour or a crossing guard instead of a traffic signal—the car requires its human operator to take control. Thus, while the Google car appears outwardly to be adaptive and flexible, it is somewhat akin to a train running on invisible tracks.

These examples highlight both the limitations of current technology to accomplish nonroutine tasks, and the capacity of human ingenuity to surmount some of these obstacles by re-engineering the environment in which work tasks are performed.

Machine Learning

Polanyi's paradox—"we know more than we can tell"—presents a challenge for computerization because, if people understand how to perform a task only tacitly and cannot "tell" a computer how to perform the task, then seemingly programmers cannot automate the task—or so the thinking has gone. But this understanding

is shifting rapidly due to advances in machine learning. Machine learning applies statistics and inductive reasoning to supply best-guess answers where formal procedural rules are unknown. Where engineers are unable to program a machine to "simulate" a nonroutine task by following a scripted procedure, they may nevertheless be able to program a machine to master the task autonomously by studying successful examples of the task being carried out by others. Through a process of exposure, training, and reinforcement, machine learning algorithms may potentially infer how to accomplish tasks that have proved dauntingly challenging to codify with explicit procedures.

As a concrete example, consider the task of visually identifying a chair (discussed in Autor, forthcoming). An engineer applying a conventional rules-based programming paradigm might attempt to specify what features of an object qualify an object as a chair—it possesses legs, arms, a seat, and a back, for example. But one would soon discover that many chairs do not possess all of these features (for example, some chairs have no back, or no arms). If the engineer then relaxed the required feature set accordingly (chair back optional), the included set would grow to encompass many objects that are not chairs, such as small tables. The canonical approach to recognizing objects by pre-specifying requisite features—and more sophisticated variants of this approach—would likely have very high misclassification rates. Yet, any grade-school child could perform this task with high accuracy. What does the child know that the rules-based procedure does not? Unfortunately, we cannot enunciate precisely what the child knows—and this is precisely Polanyi's paradox.

Machine learning potentially circumvents this problem. Relying on large databases of so-called "ground truth"—a vast set of curated examples of labeled objects—a machine learning algorithm attempts to infer what attributes of an object make it more or less likely to be designated a chair. This process is called "training." When training is complete, the machine can apply this statistical model to attempt to identify chairs that are distinct from those in the original dataset. If the statistical model is sufficiently good, it may be able to recognize chairs that are somewhat distinct from those in the original training data, like chairs of different shapes, materials, or dimensions. Machine learning does not require an explicit physical model of "chairness." At its core, machine learning is an atheoretical brute force technique—what psychologists call "dustbowl empiricism"—requiring only large training databases, substantial processing power, and, of course, sophisticated software.[10]

How well does machine learning work in practice? If you use a search engine or Google Translate, operate a smartphone with voice commands, or follow movie suggestions from Netflix, you can assess for yourself how successfully these technologies function. For example, if the majority of users who recently searched for the terms "degrees bacon" clicked on links for Kevin Bacon rather than links for best bacon cooking temperatures, the search engine would tend to place the Kevin Bacon links higher in the list of results. My general observation is that

[10] Varian (2014) provides an introduction to machine learning techniques for economists.

the tools are inconsistent: uncannily accurate at times; typically only so-so; and occasionally unfathomable. Moreover, an irony of machine learning algorithms is that they also cannot "tell" programmers why they do what they do. IBM's Watson computer famously triumphed in the trivia game of *Jeopardy* against champion human opponents. Yet Watson also produced a spectacularly incorrect answer during its winning match. Under the category of *US Cities*, the question was, "Its largest airport was named for a World War II hero; its second largest, for a World War II battle." Watson's proposed answer was Toronto, a city in Canada. Even leading-edge accomplishments in this domain can appear somewhat underwhelming. A 2012 *New York Times* article (Markoff 2012) described Google's X Lab's recent project (Le et al. 2012) to apply a neural network of 16,000 processors to identify images of cats on YouTube. The article's headline ruefully poses the question, "How Many Computers to Identify a Cat? 16,000."

Since the underlying technologies—the software, hardware, and training data—are all improving rapidly (Andreopouos and Tsotsos 2013), one should view these examples as prototypes rather than as mature products. Some researchers expect that as computing power rises and training databases grow, the brute force machine learning approach will approach or exceed human capabilities. Others suspect that machine learning will only ever "get it right" on average, while missing many of the most important and informative exceptions. Ultimately, what makes an object a chair is that it is purpose-built for a human being to sit upon. Machine-learning algorithms may have fundamental problems with reasoning about "purposiveness" and intended uses, even given an arbitrarily large training database of images (Grabner, Gall, and Van Gool 2011). One is reminded of Carl Sagan's (1980, p. 218) remark, "If you wish to make an apple pie from scratch, you must first invent the universe."

Conclusions

Major newspaper stories offer fresh examples daily of technologies that substitute for human labor in an expanding—although still circumscribed—set of tasks. The offsetting effects of complementarities and rising demand in other areas are, however, far harder to identify as they occur. My own prediction is that employment polarization will *not* continue indefinitely (as argued in Autor 2013). While some of the *tasks* in many current middle-skill jobs are susceptible to automation, many middle-skill *jobs* will continue to demand a mixture of tasks from across the skill spectrum. For example, medical support occupations—radiology technicians, phlebotomists, nurse technicians, and others—are a significant and rapidly growing category of relatively well-remunerated, middle-skill employment. Most of these occupations require mastery of "middle-skill" mathematics, life sciences, and analytical reasoning. They typically require at least two years of postsecondary vocational training, and in some cases a four-year college degree or more. This broad description also fits numerous skilled trade and repair occupations, including plumbers, builders, electricians, heating/ventilating/air-conditioning installers, and

automotive technicians. It also fits a number of modern clerical occupations that provide coordination and decision-making functions, rather than simply typing and filing, like a number of jobs in marketing. There are also cases where technology is enabling workers with less esoteric technical mastery to perform additional tasks: for example, the nurse practitioner occupation that increasingly performs diagnosing and prescribing tasks in lieu of physicians.

I expect that a significant stratum of middle-skill jobs combining specific vocational skills with foundational middle-skills levels of literacy, numeracy, adaptability, problem solving, and common sense will persist in coming decades. My conjecture is that many of the tasks currently bundled into these jobs cannot readily be unbundled—with machines performing the middle-skill tasks and workers performing only a low-skill residual—without a substantial drop in quality. This argument suggests that many of the middle-skill jobs that persist in the future will combine routine technical tasks with the set of nonroutine tasks in which workers hold comparative advantage: interpersonal interaction, flexibility, adaptability, and problem solving. In general, these same demands for interaction frequently privilege face-to-face interactions over remote performance, meaning that these same middle-skill occupations may have relatively low susceptibility to offshoring. Lawrence Katz memorably titles workers who virtuously combine technical and interpersonal tasks as "the new artisans" (see Friedman 2010), and Holzer (2015) documents that "new middle skill jobs" are in fact growing rapidly, even as traditional production and clerical occupations contract.[11]

This prediction has one obvious catch: the ability of the US education and job training system (both public and private) to produce the kinds of workers who will thrive in these middle-skill jobs of the future can be called into question. In this and other ways, the issue is not that middle-class workers are doomed by automation and technology, but instead that human capital investment must be at the heart of any long-term strategy for producing skills that are complemented by rather than substituted for by technological change. In 1900, the typical young, native-born American had only a common school education, about the equivalent of sixth to eighth grades. By the late 19th century, many Americans recognized that this level of schooling was inadequate: farm employment was declining, industry was rising, and their children would need additional skills to earn a living. The United States responded to this challenge over the first four decades of the 20th century by becoming the first nation in the world to deliver universal high school education to its citizens (Goldin and Katz 2008). Tellingly, the high school movement was led by the farm states. Societal adjustments to earlier waves of technological advancement were neither rapid, automatic, nor cheap. But they did pay off handsomely.

[11] A creative paper by Lin (2011) studies the growth of "new work" by documenting the differential growth of US employment in newly introduced Census occupation codes during the 1980s and 1990s in high-education and high-technology cities. New occupational titles are generally clustered across two categories: those associated with using new technologies such as web developer or database administrator; and novel personal services, such as personal chefs and stylists.

A final point, typically neglected in recent dismal prophesies of machine-human substitution, is that if human labor is indeed rendered superfluous by automation, then our chief economic problem will be one of distribution, not of scarcity. The primary system of income distribution in market economies is rooted in labor scarcity; citizens possess (or acquire) a bundle of valuable "human capital" that, due to its scarcity, generates a flow of income over the career path. If machines were in fact to make human labor superfluous, we would have vast aggregate wealth but a serious challenge in determining who owns it and how to share it. One might presume that with so much wealth at hand, distribution would be relatively straightforward to resolve. But history suggests that this prediction never holds true. There is always perceived scarcity and ongoing conflict over distribution, and I do not expect that this problem will become any less severe as automation advances. Are we actually on the verge of throwing off the yoke of scarcity so that our primary economic challenge soon becomes one of distribution? Here, I recall the observations of economist, computer scientist, and Nobel laureate Herbert Simon (1966), who wrote at the time of the automation anxiety of the 1960s: "Insofar as they are economic problems at all, the world's problems in this generation and the next are problems of scarcity, not of intolerable abundance. The bogeyman of automation consumes worrying capacity that should be saved for real problems . . ." A half century on, I believe the evidence favors Simon's view.

■ *This paper draws from an essay prepared for the Federal Reserve Bank of Kansas City's economic policy symposium on "Re-Evaluating Labor Market Dynamics," August 21–23, 2014, in Jackson Hole, Wyoming (Autor 2015) as well as the essay "The Paradox of Abundance: Automation Anxiety Returns" (Autor forthcoming). I thank Erik Brynjolfsson, Chris Foote, Frank Levy, Lisa Lynch, Andrew McAfee, Brendan Price, Seth Teller, Dave Wessel, participants in the MIT CSAIL/Economists Lunch Seminar, and the editors of this journal for insights that helped to shape my thinking on this subject. I thank Sookyo Jeong and Brendan Price for superb research assistance.*

References

Acemoglu, Daron, David H. Autor, David Dorn, Gordon Hanson, and Brendan Price. Forthcoming. "Import Competition and the Great U.S. Employment Sag of the 2000s." *Journal of Labor Economics.*

Akst, Daniel. 2013. "What Can We Learn from Past Anxiety over Automation?" *The Wilson Quarterly*, Summer.

Alpert, Andrew, and Jill Auyer. 2003. "Evaluating the BLS 1988–2000 Employment Projections." *Monthly Labor Review*, October, pp. 13–37.

Andreopoulos, Alexander, and John K. Tsotsos. 2013. "50 Years of Object Recognition: Directions Forward." *Computer Vision and Image Understanding* 117(8): 827–91.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49(1): 3–71.

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

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. 2015. "Polanyi's Paradox and the Shape of Employment Growth." In *Re-Evaluating Labor Market Dynamics*, pp. 129–79. Federal Reserve Bank of Kansas City.

Autor, David H. Forthcoming. "The Paradox of Abundance: Automation Anxiety Returns." In *Performance and Progress: Essays on Capitalism, Business and Society*, edited by Subramanian Rangan. London: Oxford University Press.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103(5): 1553–97.

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. 2015. "Untangling Trade and Technology: Evidence from Local Labor Markets." *Economic Journal* 125(584): 621–46.

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., 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.

Baumol, William J. 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *American Economic Review* 57(3): 415–26.

Beaudry, Paul, David A. Green, and Benjamin M. Sand. 2014. "The Declining Fortunes of the Young since 2000." *American Economic Review* 104(5): 381–86.

Beaudry, Paul, David A. Green, and Benjamin M. Sand. Forthcoming. "The Great Reversal in the Demand for Skill and Cognitive Tasks." *Journal of Labor Economics.*

Bessen, James. 2015. "Toil and Technology." *Finance and Development* 52(1).

Bowen, Harold R. (Chairman). 1966. "Report of the National Commission on Technology, Automation, and Economic Progress: Volume I." Washington: U.S. Government Printing Office.

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

Caroli, Eve, and John Van Reenen. 2001. "Skill-Biased Organizational Change? Evidence from a Panel of British and French Establishments." *Quarterly Journal of Economics* 116(4): 1449–92.

Clark, Colin. 1951. *The Conditions of Economic Progress*, 2nd edition. London: Macmillan.

Cortes, Guido Matias, Nir Jaimovich, Christopher J. Nekarda, and Henry E. Siu. 2014. "The Micro and Macro of Disappearing Routine Jobs: A Flows Approach." NBER Working Paper 20307, July.

Douglas, Paul H. 1930. *Real Wages in the United States, 1890–1926.* New York: Houghton Mifflin Company.

Foote, Christopher L., and Richard W. Ryan. 2014. "Labor-Market Polarization over the Business Cycle." Public Policy Discussion Paper 12-8, Federal Reserve Bank of Boston, April.

Frey, Carl Benedikt, and Michael A. Osborne. 2013. "The Future of Employment: How Susceptible are Jobs to Computerization?" Oxford Martin School, September.

Friedman, Thomas L. 2010. "The Election That Wasn't." *New York Times*, October 23.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology.* Cambridge: Harvard University Press.

Goos, Maarten, and Alan Manning. 2003. "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain." Center for Economic Performance Discussion Papers DP0604, December.

Goos, Maarten, and Alan Manning. 2007. "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain." *Review of Economics and Statistics* 89(1): 118–33.

Goos, Maarten, Alan Manning, and Anna Salomons. 2014. "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring." *American Economic Review* 104(8): 2509–26.

Grabner, Helmut, Juergen Gall, and Luc Van Gool. 2011. "What Makes a Chair a Chair?" In *Computer Vision and Pattern Recognition 2011*, IEEE Conference, June 20–25, pp. 1529–36.

Graetz, Georg, and Guy Michaels. 2015. "Robots at Work." IZA Discussion Papers 8938, Institute for the Study of Labor (IZA).

Herald Press, The. 1966. "Skirting the Automation Question." February 7, p. 2. (*The Herald Press* of St. Joseph, Michigan.)

Hoffman, William, and Leo Furcht. 2014. *The Biologist's Imagination: Innovation in the Biosciences*. New York, NY: Oxford University Press.

Holzer, Harry J. 2015. "Job Market Polarization and U.S. Worker Skills: A Tale of Two Middles." Brookings Institution Economic Studies Working Paper, April 6.

Hsieh, Chang-Tai, and Enrico Moretti. 2003. "Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry." *Journal of Political Economy* 111(5): 1076–1122.

Jackson, Kristin. 1993. "The World's First Motel Rests upon Its Memories." *Seattle Times*, April 25.

Jevons, William Stanley. 1865. *The Coal Question; An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal Mines*. London: Macmillan and Co.

Katz, Lawrence F., and Robert A. Margo. 2014. "Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective." In *Human Capital in History*, edited by Leah Platt Bouston, Carola Frydman, and Robert A. Margo, 15–57. University of Chicago Press.

Kremer, Michael. 1993. "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics* 108(3): 551–75.

Le, Quoc V., Marc' Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean, and Andrew Y. Ng. 2012. "Building High-Level Features Using Large Scale Unsupervised Learning." In *Proceedings of the 29th International Conference on Machine Learning*, June 26–July 1, Edinburgh, Scotland, UK.

Levy, Frank and Richard J. Murnane. 2004. *The New Division of Labor: How Computers Are Creating the Next Job Market*. Princeton University Press.

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

Markoff, John. 2012. "How Many Computers to Identify a Cat? 16,000." *New York Times*, June 25.

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.

Michaels, Guy, Ashwini Natraj, and John Van Reenen. 2014. "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years." *Review of Economics and Statistics* 96(1): 60–77.

Mishel, Lawrence, Heidi Shierholz, and John Schmitt. 2013. "Don't Blame the Robots: Assessing the Job Polarization Explanation of Growing Wage Inequality." EPI-CEPR Working Paper, November 19.

Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.

Nordhaus, William D. 2007. "Two Centuries of Productivity Growth in Computing." *Journal of Economic History* 67(1): 17–22.

Pierce, Justin R., and Peter K. Schott. 2012. "The Surprisingly Swift Decline of U.S. Manufacturing Employment." NBER Working Paper 18655.

Polanyi, Michael. 1966. *The Tacit Dimension*. New York: Doubleday.

Ruggles, Steven, Trent J. Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. University of Minnesota.

Sachs, Jeffrey D., Seth G. Benzell, Guillermo LaGarda. 2015. "Robots: Curse or Blessing? A Basic Framework." NBER Working Paper 21091, April.

Sachs, Jeffrey D., and Laurence J. Kotlikoff. 2012. "Smart Machines and Long-Term Misery." NBER Working Paper 18629, December.

Sagan, Carl. 1980. *Cosmos*. New York: Random House.

Simon, Herbert. A. 1966. "Automation" (a letter in response to "Where Do We Go From Here?" March 17, 1966 issue) *New York Review of Books*, May 26.

Smith, Christopher L. 2013. "The Dynamics of Labor Market Polarization." Federal Reserve Board, Finance and Economics Discussion Series No. 2013-57, August.

TIME. 1961. "The Automation Jobless." February 24.

US Bureau of the Census. 1949. *Historical Statistics of the United States, 1789–1945*. U.S. Government Printing Office, Series D 134-144.

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

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 31–50

The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?

Joel Mokyr, Chris Vickers, and Nicolas L. Ziebarth

Technology is widely considered the main source of economic progress, but it has also generated cultural anxiety throughout history. From generation to generation, literature has often portrayed technology as alien, incomprehensible, increasingly powerful and threatening, and possibly uncontrollable (Ellul 1967; Winner 1977). The myth of Prometheus is nothing if not a cautionary tale of these uncontrollable effects of technology. In *Civilization and its Discontents*, Sigmund Freud (1930 [1961], pp. 38–39) assessed what technology has done to homo sapiens, making him into a kind of God with artificial limbs, "a prosthetic God. When he puts on all his auxiliary organs he is truly magnificent; but those organs have not grown onto him and they still give him much trouble at times."

So it is surely not without precedent that the developed world is now suffering from another bout of such angst. In fact, these worries about technological change have often appeared at times of flagging economic growth. For example, the Great Depression brought the first models of secular stagnation in Alvin Hansen's 1938 book *Full Recovery or Stagnation?* Hansen drew on the macroeconomic ideas of John Maynard Keynes in fearing that economic growth was over, with population growth and technological innovation exhausted. Keynes was also drawn into the debate and offered a meditation on the future of technology and unemployment in his well-known essay, "Economic Possibilities for our Grandchildren."

■ *Joel Mokyr is Robert H. Strotz Professor of Arts and Sciences and Professor of Economics and History, Northwestern University, Evanston, Illinois. Chris Vickers is Assistant Professor of Economics, Auburn University, Auburn, Alabama. Nicolas L. Ziebarth is Assistant Professor of Economics, University of Iowa, Iowa City, Iowa. Their email addresses are j-mokyr@northwestern.edu, czvickers@auburn.edu, and nicolas-ziebarth@uiowa.edu.*

http://dx.doi.org/10.1257/jep.29.3.31 doi=10.1257/jep.29.3.31

This was originally written as a set of lectures in 1928 after a decade of dismal economic performance in the United Kingdom and then revised in 1930 to incorporate remarks about the Great Depression (Pecchi and Piga 2008, p. 2). Keynes (1930) remained optimistic about the future in the face of staggering unemployment, writing: "We are suffering, not from the rheumatics of old age, but from the growing-pains of over-rapid changes, from the painfulness of readjustment between one economic period and another. The increase of technical efficiency has been taking place faster than we can deal with the problem of labour absorption; the improvement in the standard of life has been a little too quick." More recently, Winner's (1977) "Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought" was published during the economic doldrums of the mid and late 1970s. Today, distinguished economists such as Lawrence Summers (2014), in a speech to the National Association of Business Economists, can be heard publicly musing about the possibility of secular stagnation. In his Martin Feldstein lecture, Summers (2013b) discussed a downright "neo-Luddite" (that famous protest movement against technological innovation in nineteenth century England) position on the effects of technology for long-term trends in employment.

Anxieties over technology can take on several forms, and we focus on what we view as three of the most prominent concerns. The first two worries are based on an "optimistic" view that technology will continue to grow and perhaps accelerate. First, one of the most common concerns is that technological progress will cause widespread substitution of machines for labor, which in turn could lead to technological unemployment and a further increase in inequality in the short run, even if the long-run effects are beneficial. Second, there has been anxiety over the moral implications of technological process for human welfare, broadly defined. In the case of the Industrial Revolution, the worry was about the dehumanizing effects of work, particularly the routinized nature of factory labor. In modern times, perhaps the greater fear is a world like that in Kurt Vonnegut's 1952 novel *Player Piano,* where the elimination of work itself is the source of dehumanization (for example, Rifkin 1995). As Summers said (as quoted "not perfectly verbatim" in Kaminska 2014), while "[t]he premise of essentially all economics . . . is that leisure is good and work is bad. . . . economics is going to have to find a way to recognize the fundamental human satisfactions that come from making a contribution . . ." A third concern cuts in the opposite direction, suggesting that the epoch of major technological progress is behind us. In recent years, even in the face of seemingly dizzying changes in information technology, pessimists such as Gordon (2012), Vijg (2011), and Cowen (2010) have argued that our greatest worry should be economic and productivity growth that will be too slow because of, for example, insufficient technological progress in the face of "headwinds" facing western economies. Some of these so-called "headwinds," including slow productivity and population growth, formed the basis of Hansen's (1939) secular stagnation hypothesis. The argument of this paper is that these worries are not new to the modern era and that understanding the history provides perspective on whether this time is truly different. The next section of the paper considers the role of these three anxieties among economists, primarily

focusing on the historical period from the late 18th to the early 20th century, while the final section offers some comparisons between the historical and current manifestations of these three concerns.

Anxieties over Technology from the Industrial Revolution to the Great Depression

Short-Term Disruption, Long-Term Benefits

We begin with the first technological transition that was extensively written about and debated in real-time by economists and many others, the Industrial Revolution. From the late 18th century onward, the debate centered on how technological progress affected workers and how these effects might differ between the short run and the long run. In the short run, could technological innovation lead to lower employment or lower wages? If there were long-run negative effects, were these innovations still worthwhile?

Prominent economists of that time had divided opinions. For example, Thomas Mortimer (1772, p. 104) wrote that he wished never to see machines such as saw mills and stamps as they would "exclude the labour of thousands of the human race, who are usefully employed . . ." In a much-debated chapter inserted into the third edition of his *Principles of Political Economy,* David Ricardo presented a candid admission of a change of opinion. In the past, Ricardo (1821 [1971], p. 380) noted, he had been convinced that an application of machinery to any branch of production was a general good, but he had more recently concluded that the "substitution of machinery for human labour is often very injurious to the interests of the class of labourers . . . [It] may render the population redundant and deteriorate the condition of the labourer." Berg (1980, pp. 67–68) underlines the eccentric economics of Ricardo on the long- versus short-run effects of technological change. Ricardo did not think that technological unemployment was the necessary result of technological progress in a specific industry. However, because of his "wage-fund" theory in which capital spent on machinery was taken out of the funds available to pay for workers, employment might be reduced as a result of investment in machinery. Ricardo felt that this case was a rather restrictive one, and that in the long run higher productivity would lead higher saving and eventually rising demand for labor.

While many writers conceded possibly negative effects of machinery on employment in the short run, they typically distinguished short-run dislocations from possible long-run effects. For example, Sir James Steuart (1767, vol. I, p. 122), widely regarded as one of the most insightful writers on economics before Adam Smith, wrote that he would disapprove of machinery only in cases in which it "might force a man to be idle" who would have no other way of earning his bread than his current employment. But normally, Steuart argued, technological unemployment would occur only if the innovation was introduced very suddenly. Even then, the dislocation to employment would be temporary, while the advantages of higher productivity would be permanent. Post-Ricardians such as John Stuart Mill conceded that improvements could

temporarily be injurious to workers, but Mill (1848 [1929], p. 97) quickly added: "I do not believe that . . . improvements in production are often, if ever, injurious, even temporarily, to the labouring classes in the aggregate." Karl Marx, from a rather different perspective, also argued that technological unemployment was a serious problem in the short run, in the broader context of the immiseration of workers under a capitalist system. But for Marx as well, technological improvement was part of a social and political process that would lead eventually to widespread prosperity. (Of course, the Marxist vision of progress also eventually required a wholesale overthrow of the existing capitalist economic system.)

Some of the most interesting thinking on the long-run effects of technology came from, for a lack of a better term, "reactionaries." These writers, while conceding the power of technology, were deeply doubtful on whether it benefited society as a whole. Yet many resigned themselves to the change and even encouraged adoption for noneconomic reasons. For example, William Mildmay (1765, p. 42) conceded that machinery might "destroy the necessity of labour" but still recommended its introduction, because other nations would otherwise outcompete Britain. This fatalism about technology was perhaps best reflected in writing of intellectuals in the antebellum US South, as in some examples cited by Genovese (1992, p. 21): "The fate of Russia in the Crimean War, declared Thomas L. Clingman, the powerful politician from North Carolina, teaches the need for railroads as a matter of military survival. Even the most 'reactionary' of southerners—even George Fitzhugh—had to agree." Other Southerners such as Thomas Roderick Dew of Virginia saw the tension between slavery and the fact that "[m]ilitary might depended upon that vaunted economic progress which the free-labor system excelled in generating" (Genovese 1992, p. 19). For these writers, technology adoption had an aspect of the prisoner's dilemma, in which each party would be foolish to pass on utilizing the newest advances even if the end result was to make everyone worse off.

Given all of this handwringing over short-run costs of technology, one may have thought that technological unemployment during the Industrial Revolution was a serious problem. However, it is one thing to argue that in certain configurations some temporary unemployment can be caused by the introduction of "machinery." It is another to demonstrate that such technological unemployment actually occurred on a large scale, and indeed the evidence is that it did not (Mokyr 2002, p. 256). In fact, on closer examination of the better-known British protests of the day that were supposedly focused technological innovations in textiles, like the Luddite (1811–16) and Captain Swing (1830–32) riots, the role actually played by the concerns of laborers about being replaced by machinery has been greatly exaggerated. These upheavals were comparable to the "Occupy Wall Street" movement (though substantially more frightening to those nearby) with a multitude of causes and a somewhat unclear set of goals. The Luddite riots started in Nottingham where workers were more concerned with low wages and work practices, in general, rather than mechanization per se. In Lancashire, machine-breaking seems to have been the result of machines being a convenient target in a dispute between

industrialists and their employees. In Yorkshire, on the other hand, wool-croppers were well-organized and clearly determined to slow down mechanization (Thomis 1970). Similarly, the Swing riots were as much aimed against cheap Irish migrant labor as they were directed against the new steam-threshers (Stevenson 1979, p. 243), yet they were one of the few instances in which mechanization was actually slowed down by political action.

Indeed, the broad claim that Britain's working-class leaders in the early decades of the 19th century resisted the machine because of the technological unemployment it caused is difficult to square with complaints about "long hours of alienated factory labour, and the smoking blight of rapidly expanding industrial towns" (Berg 1980, p. 17). The problem with the factories was not in the low quantity of work they offered, but the low quality of work in the mills.

Still, there is no doubt that by disrupting the demand for certain types of labor, the Industrial Revolution caused considerable distress, even if on balance it may not have reduced the overall demand for labor. In the British economy in the early 19th century, the workers most affected by an influx of capital investment were those employed in domestic cottage industries, which traditionally had very low capital intensity and low productivity. The handloom weavers and frame knitters with their little workshops were quite rapidly wiped out by factories after 1815 (Bythell 1969). While factory wages were rising, the real incomes of most domestic workers and independent artisans were falling (Allen 1992, pp. 255–56; 296–97; Lyons 1989). Modern work by Goldin and Katz (1998) has documented the skill-biased changes in technology from the early 20th century United States, while that of Katz and Margo (2013) has found a "hollowing out" in the skill distribution of 19th-century American manufacturing. In the early 19th century, the gaps in wages constituted the market "signal" that the death knell was sounding for a rural-industry lifestyle and culture based on cottage industries. This process was surely a painful one. Vickers and Ziebarth (2012) have suggested that some of this pain may have spilled over into higher rates of criminal behavior for those with skills depreciated by mechanization. Still, this transition was by and large a one-generation affair (Lyons 1989). Furthermore, after 1840, emigration to North America became increasingly an option for those whose livelihood disappeared.

The 19th-century workers who had jobs in the newly industrializing part of the economy had legitimate concerns with regard to wages, standards of living, and inequality. Precisely when the innovations of the Industrial Revolution translated into higher median living standards in England is a matter of dispute. Feinstein (1998) argues for almost no increase in wages prior to 1815, and even by the mid-1850s he estimates an increase of less than 30 percent compared to the 1780s. The record in the United States at that time is similarly controversial and uneven; Margo (2000, p. 51) finds generally positive rates of real wage growth from 1820–60, but with considerable variation across regions, time, and occupational class. For artisans in the Midwest, for example, wages actually fell slightly over this period. In the Northeast, real wages rose slowly in the 1820s and 1830s, with a relatively rapid rise in the 1840s.

Wages, of course, are an incomplete measure of the standard of living, and the nonpecuniary effects of industrialization on well-being have been much debated. Using commodity data, Mokyr (1988) finds a rate of consumption growth between 1815–1819 and 1845–1849 of well under a half a percent a year, with even that mostly concentrated in the later years. On the other hand, scholars such as Williamson (1981) characterize the urban disamenities associated with industrialization as "trivial": for example, coal smoke may have been unhealthy, but cheap fuel also offered warm houses and better-cooked food. However, the evidence for the pessimistic case has grown stronger. In a review, Voth (2004) argues that the modest real wages gained by 1850 were "bought by longer hours of more intensive work, performed in more dangerous and unhealthy workplaces." Estimates of inequality are still somewhat conjectural for this period and particularly sensitive to measurement issues and the construction of price indices for different classes. Still, Lindert (2000) suggests that inequality in Britain rose most quickly between 1740 and 1810, and at a somewhat slower rate after that. Allen (2005) argues for a sharpening of income inequality in Britain between 1780 and 1850. In the United States, the timing of rises in inequality is more disputed. Nevertheless, with regard to wages, standard of living, and inequality, the position of workers was at best uneven, even if technological unemployment per se was largely an exaggerated issue.

In the end, the fears of the Luddites that machinery would impoverish workers were not realized, and the main reason is well understood. The mechanization of the early 19th century could only replace a limited number of human activities. At the same time, technological change increased the demand for other types of labor that were complementary to the capital goods embodied in the new technologies. This increased demand for labor included such obvious jobs as mechanics to fix the new machines, but it extended to jobs for supervisors to oversee the new factory system and accountants to manage enterprises operating on an unprecedented scale. More importantly, technological progress also took the form of *product* innovation, and thus created entirely new sectors for the economy, a development that was essentially missed in the discussions of economists of this time. The children of the displaced handloom weavers not only had the option to work in machine-intensive cotton mills; they could also become trained engineers and telegraph operators. Nineteenth-century political economists lacked an ability to predict new job categories like the personal fashion consultants, cybersecurity experts, and online-reputation managers of the twenty-first century.

One sign that the magnitude of technological unemployment and labor displacement turned out to be relatively small during this time period—and that questions of the treatment and standard of living of workers using the new technologies loomed larger—was that the "machinery question" largely disappeared from the discourse of classical, non-Marxist political economy (Berg 1980, p. 130) in the late 19th century. While popular attention from time to time returned to the question of technological unemployment, the economics profession was dismissive. Upward trends in living standards could not be denied, and so as Woirol (1996, p. 20) notes, "the employment effects of technological change ceased to be seen as a relevant

problem." Among his examples, Woirol mentions Wells (1889, p. 374), who notes "undoubtedly a feeling of apprehension among the masses that the opportunities for employment . . . are less favorable than formerly," but concludes there is "little evidence thus far that labor has been disturbed or depressed" by machinery. Similarly, Woirol mentions the economist (and later president of Yale) Arthur Hadley (1896), who responded to the claim that mechanization "displaces a large amount of human labor, thus taking income away from employees" by noting there had been "a conspicuous increase of employment in those lines where improvements in machinery have been greatest."

The question did return in the early 20th century in the writings of Knut Wicksell (1901 [1934]) who argued, in a pure neoclassical model, that technological progress could either lower or raise the marginal product of labor, and thus wages, depending on whether the technology was labor-saving or labor-augmenting. Wicksell (p. 164) concluded, "The capitalist saver is thus, fundamentally, the friend of labour, though the technical inventor is not infrequently its enemy." However, Wicksell was careful to distinguish possible short-run deleterious effects from long-run outcomes and so continued, "The great inventions by which industry has from time to time been revolutionized at first reduced a number of workers to beggary. . . . [but then] as accumulation continues, these evils must disappear . . . and wages will rise" (p. 164). At about the same time, J. B. Clark (1907, p. 452) noted, in a widely cited remark: "The well-being of workers requires that progress should go on, and it cannot do so without causing temporary displacement of laborers." These comments from Wicksell and Clark roughly summarize the consensus of the economics profession in the early twentieth century.

During the Great Depression, a seemingly never-ending period of high unemployment, the attraction of the technological unemployment hypothesis could not be fully resisted. Ewan Clague (1935), a labor economist who was to serve as Commissioner of Labor Statistics for the Department of Labor from 1946 to 1965, stated in the pages of the *Journal of the American Statistical Association* that "the present outlook is for the rate of displacement of labor to exceed the rate of reabsorption so that technological employment will continue to be large." He concluded by noting that "in time . . . the surplus of older workers will be eliminated by age and death." Part of this worry was based not on rapid innovations in manufacturing, but on labor-saving changes in agriculture such as the tractor, which were one factor driving massive flows of people from rural areas to the cities. Others such as Edna Lonigan (a US Department of Labor economist), while cognizant of the debates over the effects of labor saving technology, rejected this argument. Lonigan (1939, p. 255) stated flatly that "our present unemployment has little to do with machines" and argued that there is no necessary connection between innovation and unemployment. Instead "it is in the failure of the price system . . . to permit the creation of new employment, that the source of the worker's growing insecurity is to be found." Of course, grasping for a secular supply-side explanation of high unemployment is not unique to the Depression of the 1930s. Many people have been quick to jump to such an explanation in the

Great Recession of 2007–2009 and its aftermath, even when a deficiency in aggregate demand may offer a more plausible explanation.

In the end, one should be careful in dismissing the performance of earlier generations of economists in predicting the effects of technological development on employment. While the predictions of widespread technological unemployment were, by and large, wrong, we should not trivialize the costs borne by the many who were actually displaced. It is true that, in the long run, wages for laborers increased to reflect dramatically increased productivity. It is also true that, for the Industrial Revolution, by many estimates it took longer than an average working lifetime to do so, and in the long run, we are all dead.

Technology and the Alienation of Labor

Besides questions of employment and wages, technological innovation brings worries about the nature of work and the so-called alienation of labor. Even before the Industrial Revolution, and in between extolling the value of specialization, Adam Smith (1776, p. 385) cautioned against the moral effects of this process, as when he wrote: "The man whose whole life is spent in performing a few simple operations . . . generally becomes as stupid and ignorant as it is possible for a human creature to become." Karl Marx, more well-known than Smith as a critic of industrialization, argued that the capitalist system alienates individuals from others and themselves. In the *Economic and Philosophic Manuscripts of 1844*, Marx wrote, "The height of this servitude is that it is only as a *worker* that he can maintain himself as a *physical subject,* and that it is only as a *physical subject* that he is a worker" (as cited in Elster 1986, p. 38).

This view of industrial capitalism as treating people like cogs in the machine became a central preoccupation of Marx and his followers, but it was no means unique to leftist revolutionaries. One can see it reflected in Thomas Jefferson's (1787, chap. 19) rosy views about the "yeoman farmer" as the basis for democracy: "Those who labour in the earth are the chosen people of God, if ever he had a chosen people, whose breasts he has made his peculiar deposit for substantial and genuine virtue." Even many reactionary supporters of slavery in the United States, such as John C. Calhoun (1837) came to a similar conclusion, viewing what some called the Northern wage-slavery system as one way in which one portion of the community lived on the labor of the other, outright slavery being the other. For Calhoun, all economic systems entailed coercion and limited freedom. Chattel slavery at least had the feature that a certain class of individuals—the slave-owners—could live a life elevated above the dirty, nasty nature of work. The factories created by industrialization offered no such option. Some such as Thomas Roderick Dew (quoted in Genovese 1992, pg. 18) held up slavery as a model because "only slavery . . . could guarantee republican liberties for the propertied, security for the propertyless, and stability for the state and society."

It is not as if the horrors of factory work were invented out of whole cloth. In the early days of the factory system, work conditions were often difficult, harsh, and unforgiving. The British Parliament in 1837 published a report titled "Evils of

the Factory System: Demonstrated by Parliamentary Evidence," partially focused on child labor. It contained graphic descriptions of accidents involving "the integuments, and the muscles, and the skin stripped off down to the bone, and in some instances a finger or two might be lost" (Wing 1837 [1967], p. clxxii). Such examples could be multiplied ad nauseam. Near the end of the 19th century, the Factory Inspectors in the State of Illinois (1895, p. 79) described factory work as involving "a degree of toil disproportionate to the condition and capacity of those engaged, while the effects of the unremitting and monotonous character of most of the work, can but stand in direct causative relation to the disturbances and depressions . . . the unremitting and monotonous character of factory work [is] productive of lessened vigor and vitality."

For Marx and others, it was not just that new factory jobs were dirty and dangerous. Jeffersonian encomiums aside, the pastoral life of small shop owners or yeoman farmers had not entailed particularly clean and safe work either. Instead the point was that this new work was in a deeper way unfit for humans and the process of covert coercion that forced people into these jobs and disciplined them while on the job was debasing. Thompson (1963, p. 305) argues that the factories were resisted even before the use of power because of "the discipline; the factory bell or hooter; the time-keeping which over-rode ill-health, domestic arrangements, or the choice of more varied occupations." One need not accept the reactionary view that such constraints on paid workers made 19th-century wage labor not very different from slavery to recognize, as many social reformers of the time did, that a lack of personal control over work raises meaningful issues.

There is little disagreement that work in the preindustrial period was much more variable for both predictable reasons (the seasonal pattern of agriculture) and unpredictable ones (driven by the small-scale nature of home production). Still these small-scale enterprises gave at least the worker-owners the appearance of autonomy in when and how they worked. The rise of the factory system ended that freedom, bringing jobs linked to a system of a set hours that most workers labor under still today. We should be careful not to take too seriously the idealization of preindustrial work by historians such as Thompson (1963), who wrote on the "work-life balance" present in preindustrial revolution economies. Freudenberger and Cummins (1976) argue that rather than reflecting some desire to consume leisure, what appeared to be a relatively short average pre–Industrial Revolution work week may have been primarily for necessary recuperation for work in the presence of high rates of disease and undernourishment.

Part of the loss of control in moving to factory work involved the physical separation of home from place of work. While today people worry about the exact opposite phenomenon with the lines between spheres of home and work blurring, this disjunction was originally a cause of great anxiety, along with the separation of place-of-work from place-of-leisure. Preindustrial societies had "no clearly defined periods of leisure as such, but economic activities, like hunting or market-going, obviously have their recreational aspects, as do singing or telling stories at work" (Thomas 1964, p. 52). Workers who were "considerably dissatisfied, because they

could not go in and out as they pleased" (Pollard 1963) had to be habituated into the factory system, by means of fines, locked gates, and other penalties. The preindustrial domestic system, by contrast, allowed a much greater degree of flexibility.

The process of industrialization also reduced the large share of transactions before the Industrial Revolution that took place within a context of personal relationships. Premodern commercial institutions were based on personal relationships that allowed for trading in the absence of well-developed formal and legal institutions (as discussed in Greif 1993). Factory work was part of a process in which personal relationships became less important in labor markets. Some such as Zucker (1986) have suggested that the "social overhead capital sector," consisting of intermediaries such as banking and insurance, increased dramatically in the late 19th century in response to the breakdown in traditional reputation mechanisms that were driven by personal contact.

Finally, there are claims that as work life and personal life separated, what was perceived as usual or virtuous behavior may have shifted, too. As industrialization in the 19th century eroded the transparent link between effort and success as understood by artisans, the moral understanding of work was transformed with the disappearance of what has been called the "moral economy," making room for a market economy. The changing nature of work provided purchase to those who viewed the rising standard of living associated with industrialization as something of a poisoned chalice. Again a number of antebellum southerners such as Henry William Ravenel held this view. As Ravenal wrote (cited in Genovese 1992, p. 30), "It is too sad proof that with all the progress made in the Arts and Sciences . . . with all the great improvements in manufactures and material prosperity, mankind are no better now than at any previous time—the evil passions of our fallen state are just as prominent and as easily brought into exercise . . ."

Historical Perspectives on a Horizon for Technological Progress

The question of whether sustained progress faced an inevitable horizon, technological or otherwise, has roots stretching back to classical antiquity. Robert Nisbet (1980) argued in his book *History of the Idea of Progress* that the ancients already ascribed to the "Idea of Progress," the claim that improvement in the moral and economic lot of man was possible. As one example, while the story of Prometheus suggests a mixed view of progress, Nisbet notes that Prometheus defends himself by pointing out the terrible condition in which he found mankind and what people were able to achieve with the gift of fire. This optimism continues through the Romans, particularly in the Lucretius' *De Rerum Natura* (*On the Nature of Things*), where he sketches out perhaps the earliest evolutionary account of the universe starting from atoms in the void. But while the idea that progress is possible and preferable was not new in the classical economists of the 18th and 19th centuries—or more broadly, to the Enlightenment period—the Enlightenment was a period in which the belief in progress became a central organizing concept of the discourse on the dynamics of society. Progress included *material* progress, or what we would think of today as economic growth. The conscious belief in the possibility of continuous betterment

of society and a detailed set of prescriptions for how to bring it about were innovations associated with the Enlightenment (Mokyr 2010, p. 33).

Perhaps most striking is the "faith," for lack of a better word, that these Enlightenment thinkers had in progress. They believed that progress was possible, that it was desirable, and that they knew how to bring it about. Others weren't so sure. Of course, the age of industrialization had its skeptics and pessimists. But our focus here is on those who were "pessimistic" about technology because they did not think much was left to be done. Consider the comment from Nobel prize-winning Albert Michelson (1903, pp. 23–25): "The more important fundamental laws and facts of physical science have all been discovered, and these are so firmly established that the possibility of their ever being supplanted in consequence of new discoveries is exceedingly remote." He goes on to quote a quip sometimes attributed to Lord Kelvin that "our future discoveries must be looked for in the sixth place of decimals."

With regards to economic progress, a number of 19th century economists came close to Lord Kelvin's view. John Stuart Mill wrote about the "stationary state" in his *Principles of Political Economy* (1848, book IV, chap. VI): "It is only in the backward countries of the world that increased production is still an important object: in those most advanced, what is economically needed is a better distribution . . ." For Mill, this stationary state of development did not imply no improvement whatsoever. "[A] stationary condition of capital and population implies no stationary state of human improvement. . . . Even the industrial arts might be as earnestly and as successfully cultivated, with this sole difference, that instead of serving no purpose but the increase of wealth, industrial improvements would produce their legitimate effect, that of abridging labor." This perspective is quite similar to the standard view of mid-19th century socialists descending from Marx's thinking about a communist society where alienation of workers from labor would end.

John Maynard Keynes (1930) set out a related view of technological progress in his essay, "Economic Possibilities for our Grandchildren." Keynes glimpsed a far-off technological horizon, where "for the first time since his creation man will be faced with his real, his permanent problem—how to use his freedom from pressing economic cares, how to occupy the leisure, which science and compound interest will have won for him, to live wisely and agreeably and well." For Keynes, the old Adamite adage that mankind would earn a living "in the sweat of thy brow" (as written in Genesis, 3:19) would change gradually over time: "For many ages to come the old Adam will be so strong in us that everybody will need to do some work if he is to be contented. We shall do more things for ourselves than is usual with the rich to-day, only too glad to have small duties and tasks and routines. Three-hour shifts or a fifteen-hour week may put off the problem for a great while. For three hours a day is quite enough to satisfy the old Adam in most of us!" Keynes was truly hopeful for such an outcome, but in a functional sense, this view of the future seems not much different from dystopias such as in the 2008 movie *Wall-E*, where humans free from economic cares spend their time floating on a futuristic version of a "lazy river."

At least in the time of Keynes (as compared to the earlier generations of economists), this future of radically reduced work hours would have followed naturally from simple extrapolation of ongoing trends. Whaples (2001) noted that the work week in US manufacturing declined from 59.6 hours in 1900 to 50.6 in 1930. At that rate of decline, the work week would have fallen to 25.4 hours by 2015. To the extent there was a decline in the number of people employed, and abstracting from the business cycle, the change in work in the late 19th and early 20th century concentrated in the young and old. From 1880 to 1920, the male labor force participation rate fell from 87.3 to 78.8 percent among those aged 65–69, but actually increased a tiny bit from 97.5 to 98.0 percent among those aged 40–44 (Sobek 2001).

The extrapolations of the end of work and "stationary states" were built on a model in which eventually people would have their fill, and no more economic progress would be necessary. It was taken as given that the level of technological progress necessary to reach this stage was possible. Whether there was additional technology to discover afterwards was beside the point. Some technological pessimists today come close to the "stationary state" theorists by drawing on a favorite analogy that the low-hanging fruit of technology progress have been picked; others view the limits on technological progress as epistemological—there are limits on what we can know.

Looking Ahead

Technology and the End of Work?

While we should not fault the lack of imagination of 19th century political economists in predicting the jobs of tomorrow, the limits they placed on the ways in which human labor could be used do seem peculiar from a modern viewpoint. The mechanical innovations of the Industrial Revolution acted as a substitute for human (and animal) strength as well as dexterity, but the machines of that time could not reason, compare, compute, read, smell, sense, hear, or make snap decisions. However, if artificial intelligence and robotics continue on their present trend, future machines will be able to carry out these human capabilities, at least in certain contexts and to a certain extent. Thus, it seems frighteningly plausible that this time will be different, and large sections of the labor market will be dislocated or "hollowed out," in the Katz and Margo (2013) terminology. Some scholars such as Beaudry, Green, and Sand (2013) have suggested that a peak in demand for high-skilled workers and cognitive tasks already occurred around the year 2000. In some theoretical models, as in Sachs, Benzell, and LaGarda (2015), a rise in "robotic productivity," which substitutes completely for labor, can result in declines in consumption, at least in the short term. But it is worth recalling that such predictions have been made repeatedly in the recent past. For example, 20 years ago in the mid-1990s, Rifkin (1995) described the spread of technology as "[l]ike a deadly epidemic inexorably working its way through the marketplace, the strange, seemingly inexplicable new economic disease spreads, destroying lives and destabilizing

whole communities in its wake" (p. 3) and cited approvingly a union leader who predicted "that within thirty years, as little as 2 percent of the world's current labor force 'will be needed to produce all the goods necessary for total demand'" (p. 8).

As we peer into the hazy future, we find it useful to distinguish two possible effects of the substitution of capital for labor: 1) how much people will work on average; and 2) how that work will be distributed. Leisure has increased over the medium term and the long term. Maddison's (2001, p. 347) computations show that between 1870 and 1998 the number of annual hours worked per employee in the highly industrialized western economies fell almost precisely by half, from roughly 2,950 hours per worker in 1870 to 1,500 hours per worker in 1998. Since 2000, OECD figures show another decline: the OECD average fell by 75 hours worked per year (although less in the United States). For economists, it would seem peculiar to fret too much about a long-term decline in work hours: indeed, the earlier discussion pointed out that there is a tradition of economists either forecasting or hoping that technology would reduce the need for work hours.

On the other hand, some economists and other social theorists have suggested that a reduced workweek is not an unalloyed good, because of underlying preferences for accomplishment and labor for its own sake. Freeman (2008, p. 141), for example, suggests that "evolution presumably imbued us with a work ethic for our survival and not a Garden of Eden existence." Phelps (2008, p. 101) writes that "if a challenging career is not the main hope for self-realization, what else could be?" Also recall Summers's call (as quoted in Kaminska 2014) for economists to "recognize the fundamental human satisfactions that comes from making a contribution." It seems plausible that attitudes toward work and the work ethic itself are not a hard-wired human universal, but rather a culturally conditioned set of beliefs and may not persist in the same form in the face of changes in the structure of the economy induced by technological change. After all, through much of history there has been a leisure class of (mostly) landowners who rarely felt the need to get dirt under their fingernails. Keynes (1930) viewed the old Adamite adage of "in the sweat of thy brow" as quite dispensable.

And as mentioned above, Keynes (1930) noted that with the decline of work, man must face the problem of how to occupy his leisure. Here technological progress has clearly changed the rules of the game. One of the underappreciated aspects of twentieth-century technological progress has been the increased marginal utility of leisure through increases in the variety of leisure and declines in the cost of leisure-directed techniques. Of course, the ultimate value of leisure activities is a matter of judgment. As Jeremy Bentham (1825, p. 206) famously wrote: "pushpin [a childish game often associated with a useless waste of time] is of equal value with the arts and sciences of music and poetry." However, it is no stretch to submit that it may be a net gain to human welfare to have fewer hours spent on a job, driving along US interstate highways, or selling tokens in a London Underground station and more hours using modern technology: for example, to watch dramas or sports of a mind-boggling variety on a high-definition flat screen; to attend virtual rock concerts or operas with high-quality sound;

to defeat the Trojans or win the tank battle of Kursk from a living room sofa using a joystick; or to "network" with friends through social media. This modern difference between leisure and work is particularly striking when compared to "leisure" in the preindustrial past that involved a fair amount of sitting in the dark. As noted, there has historically been a leisure class of people whose lives seemed quite pleasant, although their leisure activities were labor- or resource-intensive activities like golf, hunting, and formal dances. The United States was historically unusual in lacking this class, and European "visitors to the Northern States commented on the drawn faces and frantic busyness of Jacksonian Americans" and the absence of a leisure class (Rodgers 1978, p. 5). What makes today different is the fact that so much high-quality leisure activity can be accessed by all at low average cost and near-zero marginal cost.

If this predicted decline in labor hours worked was spread evenly across the working population, that decline would be a minor concern—particularly with the rise of "quality" leisure. Instead, much like the distribution of income and wealth, work hours appear to be diverging across segments of the population. Using US data, Aguiar and Hurst (2007) show that people with less than a high school education increased their leisure by almost ten hours per week from 1965 to 2003 (dominated by an increase in television watching) while college graduates increased by less than one hour per week (with an increase in television watching offset by a large decline in socializing). In a similar vein, Aguiar, Hurst, and Karabarbounis (2013) find that about half of the work hours lost by US workers in the recent recession were reallocated to leisure activity, with most of this accounted for by sleeping and television watching. As an article in *The Economist* (2014) noted, "the workers who are now working the longest hours . . . also happen to be among the most educated and best paid. The so-called leisure class has never been more harried." The welfare implications of this dual phenomenon in economic inequality of labor and leisure hours need to be further explored.

At least part of this widening inequality in hours worked is driven by the highest-skilled workers increasing their work effort, but it is also driven by outright declines in work for lower-skill workers. This change is reflected in relatively high unemployment rates for those with just a high school degree—in March 2015, for example, the US unemployment rate was 6.0 percent for those whose education ended with a high school degree versus 3.5 percent for those with a bachelor's degree—as well as in the 17-percentage-point difference in labor participation rates between these groups, based on US Bureau of Labor Statistics data. A common pattern in recent years is that routine tasks with little unpredictable variability are more likely to be mechanized, while jobs that require continuous adjustment to new information and new physical settings along with fine sensory motor-coordination are more difficult to automate. Many middle-skill jobs, both in manufacturing plants and in offices, have tended to be more susceptible to automation (as Autor discusses in this symposium). However, those middle-skill workers can then end up competing for lower-skill jobs. In this way, we are already seeing some of this labor-saving technology affecting the supply side of the lower-skilled labor force

(Jaimovich and Siu 2014; Charles, Hurst, and Notowidigdo 2014). Perhaps if these kinds of technological developments lead to an economy where an ever-larger share of the population works for relatively low wages but can still enjoy a high standard of living through a variety of low-cost leisure opportunities, political disruption may be minimal. But we do not discount the possibility that these shifts will lead to an era that redefines what goods government is responsible for providing, on a par with the political turmoil that led to the Depression-era New Deal or the creation of the German welfare state in the 19th century.

In the end, it is important to acknowledge the limits of our imaginations. Technophobic predictions about the future of the labor market sometimes suggest that computers and robots will have an absolute *and* comparative advantage over humans in all activities, which is nonsensical. The future will surely bring new products that are currently barely imagined, but will be viewed as necessities by the citizens of 2050 or 2080. These product innovations will combine with new occupations and services that are currently not even imagined. Discussions of how technology may affect labor demand are often focused on existing jobs, which can offer insights about which occupations may suffer the greatest dislocation, but offer much less insight about the emergence of as-yet-nonexistent occupations of the future. If someone as brilliant as David Ricardo could be so terribly wrong in how machinery would reduce the overall demand for labor, modern economists should be cautious in making pronouncements about the end of work.

Technology and the Characteristics of Work

Even if ongoing technological developments do not spell the end of work, they will surely push certain characteristics of future jobs back toward pre-factory patterns. These changes involve greater flexibility in when and where work takes place. Part and parcel of this increase in flexibility is the breakdown of the separation between work and home life. The main way in which flexibility seems to be manifesting itself is not through additional self-employment, but instead through the rise of contract firms who serve as matchmakers, in a phenomenon often driven by technology. For example, Autor (2001) notes that there was a decline in independent contractors, independent consultants, and freelancers as a portion of the labor force from 1995 to 1999—peak years for expansion of information technology industries—though there was a large increase in the fraction of workers employed by contract firms. The Census Bureau's counts "nonemployer businesses," which includes, for example, people with full-time employment reported in the Current Population Survey but who also received outside consulting income. The number of nonemployer businesses has grown from 17.6 million in 2002 to 22.7 million in 2012. In what is sometimes called the "sharing economy," firms like Uber and AirBnB have altered industries like cab driving and hotel management by inserting the possibility of flexible employment that is coordinated and managed through centralized online mechanisms. Firms such as oDesk or Amazon's Mechanical Turk allow for the outsourcing of tasks over the Internet that can be divided into finely sliced components.

It is not surprising that greater flexibility can be a mixed blessing. On one side, it can help in balancing work and family. For example, Goldin (2014) argues that industries in which jobs have more temporal flexibility have greater gender equality in earnings. In a survey of employers, Matos and Galinsky (2014) find that certain kinds of flexibility have become more prevalent since 2008, particularly flexibility with regard to time and place during the day, making it possible for workers to attend to personal or family needs. On the other side, flexibility can be a backdoor for employers to extract more effort from employees with an expectation that they always be accessible. In their report on workplace flexibility, the Council of Economic Advisors (2010) suggest that there has been little change since 1998 in the prevalence of these kinds of job-sharing characterized by a reduction in hours. Moreover, a penalty is strongly present for part-time work, across a variety of countries (Bardasi and Gornick 2008). Also, flexibility can often mean variable pay. The use of temp and contract workers in the "on-demand" economy (also known as contingent labor or "precarious workers") has also meant that these workers may experience a great deal of uncertainty as to how many hours they will work and when they will be called by the employers. Almost 50 percent of part-time workers receive only one week of advance notice on their schedule (as reported in Greenhouse 2014). As a recent cover story in the *Economist* (2015) has noted, workers in the on-demand economy can end up both flexible and rootless, and this creates a host of new incentive problems and monitoring costs.

The rise in flexible scheduling and other technology-driven changes, including telecommuting, have weakened the separation of work and home life. The proportion of workers working primarily from home nearly doubled from 1980 to 2010, rising from 2.3 to 4.3 percent (according to data from GlobalWorkplaceAnalytics.com). At the same time, the wage penalty for doing so has nearly disappeared (Bloom, Liang, Roberts, and Ying 2013). This change may have made time at work more pleasant, while occasionally also making time at home less pleasant. A number of white-collar professionals can sympathize with the feeling of being "always on," including doctors on call, lawyers on a case, and academics during the term. But those most affected are lower-income people like Fatimah Muhammad, quoted in Greenhouse (2014) as having "to call the [Joe Fresh clothing] store each morning, to see whether it needed her to work that day. 'I felt kind of stuck. I couldn't make plans,' said Ms. Muhammad."

The Technological Horizon

Making specific predictions about the future of technology or the economy is almost always imprudent. That said, we are skeptical for a number reasons that a horizon is relatively near—say, within a few decades—either for technological progress or for the widespread satiation of consumer demand. First, we do not foresee humanity running out of pressing technological problems anytime soon. In many cases, these problems are an outgrowth of previous technological advances. For example, the need for clean energy generation is due to industrialization and its resulting greenhouse gases in the first place. Another striking example is the

need for new antibiotics to the treat the bacteria that have become resistant to the first-generation of such wonder-drugs as penicillin and sulfa. We also expect that competition between firms, nations, and major trading blocs will stimulate continued efforts at technological gains. Even 18th-century British writers (such as Mildmay quoted earlier) who were suspicious about the effects of technological change for workers felt compelled to accept the innovations if only to ensure that Britain did not fall behind.

Finally, there is an underappreciated growth in the tools available for science and technology researchers. Across the sciences, extraordinary large amounts of data can now be stored and searched. New findings can rapidly be transmitted across the global networks of science and research. As Ridley (2010) pointed out, "The cross-fertilization of ideas between, say, Asia and Europe that once took years, decades, or centuries can now happen in minutes while Australia, the Americas, and Africa eavesdrop." One field that has been particularly affected by the development of new tools is genetics, particularly the polymerase chain reaction, which has seen the cost of sequencing a single human genome decline from $3 billion spent by the Human Genome Project to close to $5,000 in 2013 (Hayden 2014).

From our perspective, the more extreme of modern anxieties about long-term, ineradicable technological unemployment, or a widespread lack of meaning because of changes in work patterns seem highly unlikely to come to pass. As has been true now for more than two centuries, technological advance will continue to improve the standard of living in many dramatic and unforeseeable ways. However, fundamental economic principles will continue to operate. Scarcities will still be with us, most notably of time itself. The law of comparative advantage strongly suggests that most workers will still have useful tasks to perform even in an economy where the capacities of robots and automation have increased considerably.

The path of transition to this economy of the future may be disruptively painful for some workers and industries, as transitions tend to be. However, while the earliest transitions such as the Industrial Revolution were done with little governmental support for those displaced, this one will require public policy to ameliorate the harshest effects of dislocation. In particular, we believe that there is a distinct possibility that wages for some classes of workers may need to be supplemented through some income redistribution. In addition, it may be necessary to expand the set of publicly provided goods to include certain "primary goods" (Rawls 1971) such as food, housing, education, and health care that are necessary for a modern life to go well. For many others, cheaply produced goods and increasingly automated and freely available services should allow access to increasing levels of material well-being and health.

We suspect that in this new world, as material goods like food, clothing, and housing become relatively less expensive, the connection between standard measurements of output and human well-being—a long-standing source of contention—will become even more tenuous. This world would truly be the fulfillment of Simon Kuznets's (1934, p. 7) dictum that "the welfare of a country can scarcely be inferred from a measure of national income." In a world of cheap goods, while

inequality in terms of wealth or income may rise, inequality in the form of access to "primary" resources (in the Rawlsian sense) would be greatly diminished. The long-term trend toward greater leisure will continue, and one can even imagine an economy that reaches the stage in which only those who want to work actually will do so. The story of work in a world of continuing innovation is a good illustration of what is known as Amara's Law, named after systems engineer Roy Amara, long-time president of the for-profit think-tank the Institute for the Future: "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run." As we reflect on the economics of this new economy, we let Keynes (1930) offer a word of advice: "Meanwhile there will be no harm in making mild preparations for our destiny, in encouraging, and experimenting in, the arts of life as well as the activities of purpose."

References

Allen, Robert C. 1992. *Enclosure and the Yeoman: The Agricultural Development of the South Midlands, 1450–1850*. Oxford: Clarendon Press.

Allen, Robert C. 2005. "Capital Accumulation, Technological Change, and the Distribution of Income during the British Industrial Revolution." Department of Economics Discussion Paper 239, Oxford University.

Aguiar, Mark, and Erik Hurst. 2007. "Measuring Trends in Leisure: The Allocation of Time over Five Decades." *Quarterly Journal of Economics* 122(3): 969–1006.

Aguiar, Mark, Erik Hurst, and Loukas Karabarbounis. 2013. "Time Use during the Great Recession." *American Economic Review* 103(5): 1664–96.

Autor, David H. 2001. "Wiring the Labor Market." *Journal of Economic Perspectives* 15(1): 25–40.

Bardasi, Elena, and Janet C. Gornick. 2008. "Working for Less: Women's Part-time Wage Penalties across Countries." *Feminist Economics* 14(1): 37–72.

Beaudry, Paul, David A. Green, and Benjamin M. Sand. 2013. "The Great Reversal in the Demand for Skill and Cognitive Tasks." NBER Working Paper 18901.

Bentham, Jeremy. 1825. *The Rationale of Reward*. London, England: John and H. L. Hunt.

Berg, Maxine. 1980. *The Machinery Question and the Making of Political Economy, 1815–1848*. Cambridge University Press.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2013 "Does Working from Home Work? Evidence from a Chinese Experiment." NBER Working Paper 18871.

Bythell, Duncan. 1969. *The Handloom Weavers: A Study in the English Cotton Industry during the Industrial Revolution*. Cambridge University Press.

Calhoun, John C. 1837. "The 'Positive Good' of Slavery in 1837." Speech in the U.S. Senate, February 6th.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2014. "Housing Booms, Labor Market Outcomes, and Educational Attainment." http://faculty.wcas.northwestern.edu/noto/research/charles_hurst_noto_housing_booms.pdf.

Clague, Ewan. 1935. "The Problem of Unemployment and the Changing Structure of Industry." *Journal of the American Statistical Association* 30(189): 209–214.

Council of Economic Advisors. 2010. "Work-Life Balance and the Economics of Workplace Flexibility." March. https://www.whitehouse.gov/files/documents/100331-cea-economics-workplace-flexibility.pdf.

Cowen, Tyler. 2010 "The Great Stagnation." New York: Dutton.

Clark, Andrew E. 2003. "Unemployment as a Social Norm: Psychological Evidence from Panel Data." *Journal of Labor Economics* 21(2): 323–51.

Clark, John Bates. 1907. *Essentials of Economic Theory*. London: MacMillan.

Economist, The. 2014. "Why is Everyone So Busy?" December 20.

Economist, The. 2015. "The Future of Work: There's an App for That." January 3.

Ellul, Jacques. 1967. *The Technological Society.* Vintage Books.

Elster, Jon (ed.) 1986. *Karl Marx: A Reader.* Cambridge University Press.

Feinstein, Charles H. 1998. "Pessimism Perpetuated: Real Wages and the Standard of Living in Britain during and after the Industrial Revolution." *Journal of Economic History* 58(3): 625–58.

Fogel, Robert W. 2009. "Forecasting the Cost of U.S. Health Care in 2040." *Journal of Policy Modeling* 31(4): 482–88.

Freeman, Richard B. 2008. "Why Do We Work More than Keynes Expected?" Chap. 9 in *Revisiting Keynes: Economic Possibilities for our Grandchildren*, edited by Lorenzo Pecchi and Gustavo Piga. MIT Press.

Freud, Sigmund. [1930] 1961. *Civilization and Its Discontents.* New York: W.W. Norton.

Freudenberger, Herman, and Gaylord Cummins. 1976. "Health, Work, and Leisure before the Industrial Revolution." *Explorations in Economic History* 13(1): 1–12.

Genovese, Eugene D. 1992. *The Slaveholders' Dilemma: Freedom and Progress in Southern Conservative Thought, 1820–1860.* Columbia, SC: University of South Carolina Press.

Goldin, Claudia. 2014 "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4): 1091–1119.

Goldin, Claudia, and Lawrence F. Katz. 1998. "The Origins of Technology-Skill Complementarity." *Quarterly Journal of Economics* 113(3): 693–732.

Gordon, Robert J. 2012. "Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds." NBER Working Paper 18315.

Greenhouse, Steven. 2014 "A Push to Give Steadier Shifts to Part-Timers." *New York Times,* July 15.

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83(3): 525–48.

Hadley, Arthur T. 1896. *Economics.* New York: Putnam's Sons.

Hamermesh, Daniel S, Daiji Kawaguchi, and Jungmin Lee. 2014. "Does Labor Legislation Benefit Workers? Well-Being after an Hours Reduction." IZA Discussion Paper 8077.

Hansen, Alvin Harvey. 1938. *Full Recovery or Stagnation?* W.W. Norton.

Hansen, Alvin H. 1939. "Economic Progress and Declining Population Growth." *American Economic Review* 29(1): 1–15.

Hayden, Erika Check. 2014. "Technology: The $1000 Genome." *Nature,* March 19, 507(7492): 294–95.

Illinois, State of. 1895. *Report of the Factory Inspectors of Illinois for the Year Ending Dec. 15, 1894.* Springfield: Ed. F. Hartman, State Printer.

Jaimovich, Nir, and Henry E. Siu. 2014. "The Trend is the Cycle: Job Polarization and Jobless Recoveries." NBER Working Paper 18334, version revised March 20, 2014.

Jefferson, Thomas. 1787. "Notes on the State of Virginia." University of Virginia American Studies. http://xroads.virginia.edu/~hyper/JEFFERSON/intro.html.

Kaminska, Izabella. 2014. "Larry Summers on Forwarding the Doozer Economy." *Financial Times.* FT Alphaville blog, April 17.

Katz, Lawrence F., and Margo, Robert A. 2013. "Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective." NBER Working Paper 18752.

Keynes, John Maynard. 1930. "Economic Possibilities for Our Grandchildren." In *Essays in Persuasion* [1963], pp. 358–73. New York: W.W. Norton.

Kuznets, Simon. 1934. "National Income, 1929–1932." Senate Document no. 124, 73rd US Congress, 2nd session.

Lindert, Peter H. 2000. "When Did Inequality Rise in Britain and America?" *Journal of Income Distribution* 9(1): 11–25.

Lonigan, Edna. 1939. "The Effect of Modern Technological Conditions upon the Employment of Labor." *American Economic Review* 29(2): 246–59.

Lyons, John S. 1989. "Family Response to Economic Decline: Handloom Weavers in Early Nineteenth-Century Lancashire." *Research in Economic History* vol.12, pp. 45–91.

Maddison, Angus. 2001. *The World Economy: A Millennial Perspective.* Paris: OECD.

Margo, Robert A. 2000. *Wages and Labor Markets in the United States, 1820–1860.* University of Chicago Press.

Marx, Karl. 1844 [1988]. *The Economic and Philosophic Manuscripts of 1844.* Translated by Martin Milligan. Amherst, NY: Prometheus Books.

Matos, Kenneth, and Ellen Galinsky. 2014. "2014 National Study of Employers." Families and Work Institute.

Michelson, Albert A. 1903. *Light Waves and Their Uses.* University of Chicago Press.

Mildmay, William. 1765. *The Laws and Policy of England, Relating to Trade, Examined by the Maxims and Principles of Trade in General.* London: T. Harrison.

Mill, John Stuart. 1848 [1929]. *Principles of Political Economy,* edited by W. J. Ashley. London: Longmans, Green and Co.

Mokyr, Joel. 1988. "Is There Still Life in the Pessimistic Case? Consumption during the

Industrial Revolution, 1790–1850." *Journal of Economic History* 48(1): 69–92.

Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.

Mokyr, Joel. 2010. *The Enlightened Economy: An Economic History of Britain, 1700–1850*. Yale University Press.

Mortimer, Thomas. 1772. *The Elements of Commerce, Politics and Finances*. London: Hooper.

Nisbet, Robert. 1980. *History of the Idea of Progress*. Basic Books.

Pecchi, Lorenzo, and Gustavo Piga (eds). 2008. *Revisiting Keynes: Economic Possibilities for our Grandchildren*. MIT Press.

Phelps, Edmund S. 2008. "Corporatism and Keynes: His Philosophy of Growth." In *Revisiting Keynes: Economic Possibilities for our Grandchildren*, edited by Lorenzo Pecchi and Gustavo Piga. MIT Press.

Pollard, Sidney. 1963. "Factory Discipline in the Industrial Revolution." *Economic History Review* 16(2): 254–71.

Rawls, John. 1971. *A Theory of Justice*. Belknap/Harvard University Press.

Ricardo, David. 1821 [1971]. *Principles of Political Economy*, 3rd edition, edited by R.M. Hartwell. Harmondsworth: Pelican Classics.

Ridley, Matt. 2010. "When Ideas Have Sex." http://www.fastcodesign.com/1662632/frog-design-when-ideas-have-sex.

Rifkin, Jeremy. 1995. *The End of Work*. New York: J.B. Putnam.

Rifkin, Jeremy. 2014. *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*. Palgrave Macmillan.

Rodgers, Daniel T. 1978. *The Work Ethic in Industrial America, 1850–1920*. Second edition. University of Chicago Press.

Sachs, Jeffrey D, Seth G. Benzell, and Guillermo LaGarda. 2015. "Robots: Curse or Blessing? A Basic Framework." NBER Working Paper 21091.

Smith, Adam. 1776 [1965]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. New Rochelle: Arlington House.

Sobek, Matthew. 2001. "New Statistics on the U.S. Labor Force, 1850–1990." *Historical Methods* 34(2): 71–81.

Steuart, James. 1767. *An Inquiry into the Principles of Political Economy*. London: Printed for A. Millar, and T. Cadell.

Stevenson, John. 1979. *Popular Disturbances in England, 1700–1870*. New York: Longman.

Summers, Lawrence. 2013a. "Why Stagnation Might Prove to Be the New Normal." *Financial Times*, December 15.

Summers, Lawrence H. 2013b. "Economic Possibilities for Our Children." The 2013 Martin Feldstein Lecture. *NBER Reporter* no. 4, pp. 1–6.

Summers, Lawrence H. 2014. "U.S. Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound." *Business Economics* 49(2): 65–73.

Thomas, Keith. 1964. "Work and Leisure in Pre-Industrial Society." *Past & Present* no. 29, pp. 50–66.

Thomis, Malcolm. 1970. *The Luddites*. New York: Schocken.

Thompson, E. P. 1963. *The Making of the English Working Class*. Victor Gollancz Ltd.

Vickers, Chris, and Nicolas L. Ziebarth. 2012 "Economic Development and the Demographics of Criminals in Victorian England." Unpublished paper, Auburn University and University of Iowa.

Vijg, Jan. 2011. *The American Technological Challenge: Stagnation and Decline in the 21st Century*. New York: Algora Publishing.

Voth, Hans-Joachim. 2004. "Living Standards and the Urban Environment." Chap. 10 in *The Cambridge Economic History of Modern Britain: Volume I*. Edited by Roderick Floud and Paul Johnson. Cambridge University Press.

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

Whaples, Robert. 2001. "Hours of Work in U.S. History." EH.Net Encyclopedia, edited by Robert Whaples.

Wicksell, Knut. 1901 [1934]. *Lectures on Political Economy*, Vol. I: *General Theory*. Translated by E. Classen. London: George Routledge.

Williamson, Jeffrey G. 1981. "Urban Disamenities, Dark Satanic Mills, and the British Standard of Living Debate." *Journal of Economic History* 41(1): 75–83.

Wing, Charles. 1837 [1967]. *Evils of the Factory System Demonstrated by Parliamentary Evidence*. London: Frank Cass reprints.

Winner, Langdon. 1977. *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. MIT Press.

Woirol, Gregory R. 1996. *The Technological Unemployment and Structural Unemployment Debates*. Westport, CT: Greenwood.

Zucker, Lynne G. 1986. "Production of Trust: Institutional Sources of Economic Structure, 1840–1920." *Research in Organizational Behavior* 8: 53–111.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 51–60

Is a Cambrian Explosion Coming for Robotics?

Gill A. Pratt

About half a billion years ago, life on earth experienced a short period of very rapid diversification called the "Cambrian Explosion." Many theories have been proposed for the cause of the Cambrian Explosion, with one of the most provocative being the evolution of vision, which allowed animals to dramatically increase their ability to hunt and find mates (for discussion, see Parker 2003). Today, technological developments on several fronts are fomenting a similar explosion in the diversification and applicability of robotics. Many of the base hardware technologies on which robots depend—particularly computing, data storage, and communications—have been improving at exponential growth rates. Two newly blossoming technologies—"Cloud Robotics" and "Deep Learning"—could leverage these base technologies in a virtuous cycle of explosive growth. In Cloud Robotics—a term coined by James Kuffner (2010)—every robot learns from the experiences of all robots, which leads to rapid growth of robot competence, particularly as the number of robots grows. Deep Learning algorithms are a method for robots to learn and generalize their associations based on very large (and often cloud-based) "training sets" that typically include millions of examples. Interestingly, Li (2014) noted that one of the robotic capabilities recently enabled by these combined technologies is vision—the same capability that may have played a leading role in the Cambrian Explosion.

■ *Gill A. Pratt will be stepping down next month from his position as a Program Manager of the Defense Advanced Research Projects Agency (DARPA), Arlington, Virginia, where he oversaw the DARPA Robotics Challenge and several other programs in robotics while on leave from the faculty of Franklin W. Olin College, Needham, Massachusetts. The views expressed in this paper are those of the author, not Olin College, DARPA, the US Department of Defense, or the US Government.*

http://dx.doi.org/10.1257/jep.29.3.51 doi=10.1257/jep.29.3.51

How soon might a Cambrian Explosion of robotics occur? It is hard to tell. Some say we should consider the history of computer chess, where brute force search and heuristic algorithms can now beat the best human player yet no chess-playing program inherently knows how to handle even a simple adjacent problem, like how to win at a straightforward game like tic-tac-toe (Brooks 2015). In this view, specialized robots will improve at performing well-defined tasks, but in the real world, there are far more problems yet to be solved than ways presently known to solve them.

But unlike computer chess programs, where the rules of chess are built in, today's Deep Learning algorithms use general learning techniques with little domain-specific structure. They have been applied to a range of perception problems, like speech recognition and now vision. It is reasonable to assume that robots will in the not-too-distant future be able perform any associative memory problem at human levels, even those with high-dimensional inputs, with the use of Deep Learning algorithms. Furthermore, unlike computer chess, where improvements have occurred at a gradual and expected rate, the very fast improvement of Deep Learning has been surprising, even to experts in the field. The recent availability of large amounts of training data and computing resources on the cloud has made this possible; the algorithms being used have existed for some time and the learning process has actually become simpler as performance has improved.

While the so-called "neural networks" on which Deep Learning is often implemented differ from what is known about the architecture of the brain in several ways, their distributed "connectionist" approach is more similar to the nervous system than previous artificial intelligence techniques (like the search methods used for computer chess). Several characteristics of real brains are yet to be accomplished, such as episodic memory and "unsupervised learning" (the clustering of similar experiences without instruction), but it seems likely that Deep Learning will soon be able to replicate the performance of many of the perceptual parts of the brain. While questions remain as to whether similar methods can also replicate cognitive functions, the architectures of the perceptual and cognitive parts of the brain appear to be anatomically similar. There is thus reason to believe that artificial cognition may someday be put into effect through Deep Learning techniques augmented with short-term memory systems and new methods of doing unsupervised learning. To date, there are no huge datasets about cognition that are similar to the picture and speech datasets that have been so effective for computer learning about perception. But some methodologies for collecting such datasets, described below, may be possible.

The timing of tipping points is hard to predict, and exactly when an explosion in robotics capabilities will occur is not clear. Commercial investment in autonomy and robotics—including and especially in autonomous cars—has significantly accelerated, with high-profile firms like Amazon, Apple, Google, and Uber, as well as all the automotive companies, announcing significant projects in this area. In the next two sections of this paper, I examine some key technologies contributing to the present excitement in the robotics field. As with other technological developments, there has been a significant uptick in concerns about the societal implication of robotics and

artificial intelligence. Thus, I offer some thoughts about how robotics may affect the economy and some ways to address potential difficulties.

Eight Technical Drivers

A number of technologies relevant to the development of robotics are improving at exponential rates. Here, I discuss eight of the most important. The first three technological developments relate to individual robots; the next two relate to connectivity; and the final three relate to the capacities of the Internet that will shape the future of cloud robotics.

1) Exponential growth in computing performance. Robots are made up of computers that allow sensors and actuators to collaborate, and the processing power of computers keeps rising. Moore's law, as originally proposed by Gordon Moore (1965, 1975), one of the founders of Intel, refers to the doubling of transistor count on integrated circuits roughly every 18–24 months with similar improvements in processing speed. This relationship has held for many decades, although it is now approaching some fundamental limits. Semiconductor companies are now etching transistors on to chips at a scale of 14 nanometers (for example, see http://www.intel.com/content/www/us/en/silicon-innovations/intel-14nm-technology.html), where a nanometer is an almost unimaginably small one-billionth of a meter: for scale, a typical sheet of paper is about 100,000 nanometers thick. This small scale is approaching physical limits because it involves working at close to the atomic level. However, there do appear to be technologies that can continue Moore's law for at least another few years (Bauer, Veira, and Weig 2013), such as those that move beyond two-dimensional integrated circuits and work at a system level with three-dimensional and multi-chip systems.

2) Improvements in electromechanical design tools and numerically controlled manufacturing tools. Modern computer-aided design tools have significantly improved the productivity of electromechanical designers, including the quality of what is designed and the sophistication of what can be designed. Numerically controlled manufacturing tools—including the new "additive processes" such as 3D printing—can build such designs with great precision and little cost to additional complexity. Embedded processors have allowed for tremendous control complexity and performance and reliability improvements, all of which increase the competence and reliability of robots.

3) Improvements in electrical energy storage. If robots are to be mobile, they need to find ways to store or generate sufficient power to operate for reasonable periods—at least between episodes of recharging. Over the last few decades, advances in electric batteries and fuel cells have had a poor history of living up to promises. Standard lithium-ion batteries still remain nearly an order of magnitude less energy-dense than hydrocarbon fuels (including gasoline, as well as sugars and fats), but the gap is slowing closing. The high demand and fierce competition in the portable electronic markets of laptops, tablets, and cell phones, not to mention hybrid and electric vehicles, has continued to spark innovation and steady improvement in

energy storage. Supercapacitors, a new technology, charge and discharge much faster that standard batteries, and can be recharged hundreds or thousands of times, but significant improvements still need to be made in how much energy they can store. (For an accessible discussion of the tradeoffs between lithium-ion batteries and supercapacitors, see Miret 2013.) For many robot applications with ready access to recharging infrastructure, battery energy storage is already sufficient, and energy storage should continue to improve over time.

4) Improvements in electronics power efficiency. Robots running on electric batteries use electronics for power management of motors, and robots with many motors are particularly sensitive to the cost and performance of these electronics. Power-related semiconductors have taken advantage of general technology improvements in the integrated circuit industry and have also become much less expensive due to the continual improvements in portable devices, which all include batteries that are sensitive to the efficiency of power electronics. LED lighting is another rapidly growing market in which power-electronics semiconductors enable the production of light more efficiently with less power. New types of compound semiconductors (gallium nitride and silicon carbide) promise to usher in even higher performance and lower prices. The computational needs for the combination of cloud robotics and Deep Learning are currently being provided by graphics processing units—extremely high-performance computer chips originally developed for video games. In the future, computation may be provided by neuromorphic (brain-inspired) hardware, which often consumes less power.

5) Exponential expansion of the availability and performance of local wireless digital communications. Early robots were essentially stand-alone machines. Their capacities to remember and to solve problems were limited by the programming that they could carry around with them. Updating their information or reprogramming them was a costly and time-consuming process. However, flexible web-connected robots offer different possibilities for programming, problem-solving, learning, and updating. High-performance wireless digital communications are becoming ubiquitous, as are products that leverage that infrastructure, including a wide variety of tablets along with more specialized products. For example, the "learning thermostat" produced by Nest not only can be adjusted from your phone, but it also remembers and learns when you turn it higher or lower, and starts to make future adjustments automatically. The Google Chromecast device lets you take any content from your computer or mobile device and show it on your television using a wireless connection. Average worldwide WiFi Speeds, which were 10 megabits per second in 2014, will nearly double by 2018. There were 48 million public WiFi hotspots globally in 2014, a number that is expected to increase by a factor of seven by 2018 (Cisco 2015). The latest WiFi standard (802.11ac) surpasses one gigabit per second, as does the latest cellular data standard (5G). With these changes, robots will be able to communicate wirelessly within their facilities more quickly than ever.

6) Exponential growth in the scale and performance of the Internet. As wireless communication within facilities evolves, so will Internet communication outside the facility. The global Internet is presently estimated to carry about 88 exabytes (that

is, 10^{18} bytes) of traffic per month, which is predicted to double in three years, with no saturation in sight. There are currently about 13 billion devices connected to the Internet, already two for each human being on Earth; this ratio is projected to reach three for each human being on Earth by 2019 (Cisco 2015).

7) Exponential growth of worldwide data storage. On a global basis, total information stored is on the order of a 10^{21} bytes,[1] with volume ever increasing due to explosive demand for entertainment and social media. By comparison, the human brain has on the order of 10^{14} synapses. If we crudely equate each synapse to one byte of storage, the current level of worldwide data storage is on the order of 10 million human brains. Of course, bytes on a spinning disk drive accessed serially by a computer are not densely interconnected the way synapses are inside of a brain, so this remarkable amount of information storage has not been combined to emulate even one brain, much less 10 million. But it is a lot of data.

8) Exponential growth in global computation power. Worldwide total computation performance has reached on the order of a 10^{21} instructions per second.[2] More importantly, many billions of disk drives have been produced (although perhaps only a billion or so are running now), and several large Internet companies run millions of high-performance servers in parallel, each with high-performance multiple core processors. As a result, any computations that can be broken into parallel operations—because no communication of intermediate results are needed to solve the separate problems—can be parceled out and solved quickly. (Problems that can be broken into a number of pieces that can be solved in parallel are sometimes called "embarrassingly parallel" problems.) Many problems in robot autonomy can be solved this way.

Cloud Robotics

These technological developments suggest that the capabilities of robots themselves are increasing rapidly, along with the ability to interconnect robots. Several big ideas, collectively known as "cloud robotics," are poised to leverage many of these technologies to enable the revolution in robot capabilities. The potential gains from cloud robotics can be summarized with four big ideas.

Big Idea #1: Memory-Based Autonomy

The exponential growth in computing and storage performance has led researchers to explore memory-based methods of solving the perception, planning, and control problems relevant to the development of additional degrees of robot autonomy. Instead of decomposing these tasks into a set of hand-coded algorithms customized

[1] Hilbert and Lopez (2011) estimate total global information stored at 2.9 x 10^{20} bytes in 2007, and growing at 23 percent per year. Extrapolating this growth rate through 2015 provides a basis for the rough estimate in the text.

[2] Hilbert and Lopez (2011) estimate world computational power at 6.4 x 10^{18} instructions per second in 2007, and growing at 58 percent per year. Extrapolating this growth rate through 2015 provides a basis for the rough estimate in the text.

for particular circumstances, large numbers of memories of prior experiences can be searched, and a solution based on matching prior experience is used to guide response. When no matching prior memory exists, actions from similar prior memories can be interpolated, or human help can be requested, with the human-provided answer then recorded for future use both at that robot and at other robots as well.

The continuing progress on fast search algorithms for Internet information has accelerated memory-based approaches. Using cloud-based computing, the lookup of prior examples can exploit a large number of external computing resources in parallel with a relatively small amount of communication. But for memory-based methods to work, where do the memories of solutions come from?

Big Idea #2: High-Speed Sharing of Experiences

A single robot, using a memory-based method to implement additional degrees of autonomy, would, like a newborn child, quite probably take decades to learn to do anything useful. Indeed, the robot would be much slower than a human child, because even instincts would be missing.

But while the communications bandwidth inside of a human brain is high relative to that of a robot, human beings communicate externally with one another relatively slowly, at rates on the order of 10 bits per second. Robots, and computers in general, can communicate at rates over one gigabit per second—or roughly 100 million times faster. Based on this tremendous difference in external communication speeds, a combination of wireless and Internet communication can be exploited to share what is learned by every robot with all robots. Human beings take decades to learn enough to add meaningfully to the compendium of common knowledge. However, robots not only stand on the shoulders of each other's learning, but can start adding to the compendium of robot knowledge almost immediately after their creation.

It is unclear whether future robots will have high-performance on-board "brains" that cache memories from the cloud, using the Internet only when they aren't sure what to do, or whether high-speed internet communications will mean that most of a robot's intelligence (like most of the computation done by some modern video games) will be accomplished remotely using computational resources on the cloud. In either case, the capacities of fast communication and the Internet will catalyze increases in robot capability.

Big Idea #3: Learning from Imagination

Human beings often use imagination to practice and prepare for future circumstances. Similarly, a robot—or a cloud-computing robot "brain"—can use simulation to explore circumstances that may be faced by a robot in the future and to experiment with possible solutions, remembering only those that worked. Such simulations can be done without the need for any physical activity, and every robot's dreams will improve the performance of all robots.

Big Idea #4: Learning from People

Perception remains one of the most challenging components of robot autonomy. Recently, the ability of large datasets to catalyze perception has proven

to be quite powerful. Large datasets may also have significant utility in planning and control as well.

The online repository of visually recorded objects and human activity is a tremendous resource that robots may soon exploit to improve their ability to understand and interact with the world, including interactions with human beings. Social media sites uploaded more than 1 trillion photos in 2013 and 2014 combined (Meeker 2014, slide 62), and given the growth rate may upload another trillion in 2015. At present, about 300 hours of video are uploaded to the sharing site YouTube every minute, mostly showing people interacting with each other and the environment (see YouTube, https://www.youtube.com/yt/press/statistics.html). When 3D sensors become common, the richness of this data store will improve even further.

Most visual information on the Internet is of course not labeled, but clustering techniques can be used to identify similar components in images and videos—for example, in the way that similar faces are grouped now—so that when the meaning of one visual example is learned (perhaps by being labeled by a human), that information can inform the understanding of other images and videos. The utility of this resource is ripe for exponential growth.

Some Implications for the Economy and Workforce

While a Cambrian Explosion in robotics promises to improve the human condition dramatically, it also looms as a disruptive economic force, in part because of its much-discussed potential to make certain human jobs redundant. Yet there is reason to embrace the pending robotics revolution despite such concerns. Consider a Robinson Crusoe economy—a single person stranded on an island with no need for money. Would such a person benefit from a robot, or set of robots, to help with production? Of course. The fundamental economic insight is that robots generate wealth, but the traditional distribution of that wealth through human labor can become problematic. Here, I can offer the thoughts and speculations of a noneconomist on the potential impacts of robots on the ever-evolving labor force. I also discuss some alternatives to the distribution of wealth through labor. These include human services, distribution through capital, and something new—the personal preference information economy.

Output, Demand Satiation, and Human Adaptation

The traditional interaction between technology and the labor market has followed a pattern. Technology enabled an increase in output in certain areas. Demand for the goods produced in that area became at least somewhat satiated, but people were not satiated in their wants and instead soon discovered new areas of demand. Some human labor was displaced as technology expanded, but supply and demand in the labor market drove a series of transitions so that labor shifted to meet the new demand in other areas, and there was no sustained trend to greater unemployment over time. Instead, average wages increased because

technology lifted the productivity of labor. As one example, the Industrial Revolution in textile production created working conditions that were often brutal, but textile output rose and prices fell so that customers (often themselves workers) could afford to buy textiles that would formerly have been out of their economic reach. What consumers didn't spend on textiles could be spent in other sectors, including new sectors. Over time, working conditions in the textile industry improved and labor diversified.

This general pattern has repeated itself in many sectors of the economy over the last two centuries despite warnings every few decades that automation was about to cause mass unemployment. However, this time may be different. When robot capabilities evolve very rapidly, robots may displace a much greater proportion of the workforce in a much shorter time than previous waves of technology. Increased robot capabilities will lower the value of human labor in many sectors. Human abilities as suppliers, even in highly educated societies, evolve slowly. In other words, the increase in robot capabilities may be so rapid that many human workers may find themselves with little to sell.

In the longer run, the diversity and scale of human demand for goods and services has seemed insatiable—so that the labor demanded by the economy did not diminish. But as robot capabilities improve beyond a growing range of human capabilities, will this pattern continue to hold true? One can imagine a future in which many of the material goods that most people want are produced at low cost by advanced robots. Such an economy could evolve in a number of ways. But one possible outcome is that robots may do to many sectors of the economy what the Internet has done to the music business—that is, lead to an economy that pays superstar wages to a small number of exceptionally talented people while paying only a low level of income to many others. In the rest of this section, I explore a few possible ways out of this conundrum.

Human Services

Today, even if a machine-made product is superior in a number of dimensions to a hand-made one, the hand-made product often commands a premium price because it is more difficult to produce and involves the use of a precious commodity—the time of a skilled artisan. As one of many examples, live music continues to be a strong market even as the music recording industry has largely collapsed. This difference exists despite the fact that recorded music is often produced with far higher quality and fidelity than live music, and live music is often played in acoustic environments inferior to those in which music was recorded. However, the excitement of attending a live concert with a crowd of other people is a more valuable experience than listening alone to a recording or the radio. No one proudly wears T-shirts declaring when they listened to a recorded song.

Thus, some human services will probably continue to command a premium compared to robotically produced ones. The question is one of volume: can a future economy be based primarily on personal or hand-crafted goods when close-substitute robotically produced goods become very inexpensive? Concerts are popular, but the vast majority of the music to which people listen is recorded.

The Distribution of Robot Capital Ownership

Imagine a hypothetical economy in which everyone owned a robot and sent their robot to work in their stead. In such a world, the economy could proceed without a hitch, except that we would all have much more leisure time while our robotic stand-ins earned our keep. Of course, the matter of how to initially distribute, trade, and provide safeguards against bankruptcy for robot capital would have to be worked out. But essentially the present system of trading capital, where the intelligence (and significant luck) of investors determines who gets more and who gets less could provide at least some basis for distribution. The late James Albus, an engineer who also served as the head of the Intelligent Systems Division of the Manufacturing Engineering Laboratory at the National Institute of Standards and Technology, explored these issues and possibilities in depth in his 1976 book *Peoples' Capitalism: The Economics of the Robot Revolution*, and its 2011 successor *Path to a Better World: A Plan for Prosperity, Opportunity, and Economic Justice*.

The Personal Preferences Information Economy

In pre-mechanized economies, human beings were born with innate capital for producing economically valuable goods—their bodies. When technology lowered the value of mechanical labor, the economic value of bodies declined but the intrinsic capital value of human brains increased. If brains go the way of bodies, what inherent value will human beings have? Intriguingly, a new inherent human capital has arisen—personal preferences.

Internet companies that had their start producing computer tools like search, email, maps and others have monetized the personal preferences about their users gathered by the tools themselves—which are typically given away "for free." The gathered information is then sold to advertisers who use it to target individuals most likely to purchase specific goods. The business of these companies is fundamentally the arbitrage of personal preference information. Many people today don't realize the value of their personal preferences, although the substantial profits of the companies that gather and sell such information makes clear its value.

In a future robotic economy, various characteristics of bodies and brains may have much less economic value, but the inherently human value of personal preferences will remain. Were individuals not to surrender personal information so easily, one could imagine Internet companies playing a significant role in wealth distribution by regularly compensating individuals for the value of the information they provide about themselves.

What's Holding Back Robots?

The human brain does much more than store a very large number of associations and access useful memories quickly. It also transforms sensory and other information into generalizable representations invariant to unimportant changes, stores episodic memories, and generalizes learned examples into understanding. The key problems in robot capability yet to be solved are those of generalizable

knowledge representation and of cognition based on that representation. How can computer memories represent knowledge to be retrieved by memory-based methods so that similar but not identical situations will call up the appropriate memories and thoughts?

Significant cues are coming from the expanding understanding of the human brain, with the rate of understanding accelerating because of new brain imaging tools. Some machine learning algorithms, like the Deep Learning approached discussed earlier, are being applied in an attempt to discover generalizable representations automatically. It is not clear how soon this problem will be solved. It may only be a few years until robots take off—or considerably longer. Robots are already making large strides in their abilities, but as the generalizable knowledge representation problem is addressed, the growth of robot capabilities will begin in earnest, and it will likely be explosive. The effects on economic output and human workers are certain to be profound.

References

Albus, James S. 1976. *Peoples' Capitalism: The Economics of the Robot Revolution*. New World Books.

Albus, James S. 2011. *Path To A Better World: A Plan for Prosperity, Opportunity, and Economic Justice*. iUniverse.

Bauer, Harald, Jan Veira, and Florian Weig. 2013. "Moore's Law: Repeal or Renewal?" McKinsey Global Institute, December. http://www.mckinsey.com/insights/high_tech_telecoms_internet/moores_law_repeal_or_renewal.

Brooks, Rodney. 2015. In the "Panel Discussion: Artificial Intelligence and the Future of Humans and Robots in the Economy." Malcolm and Carolyn Wiener Annual Lecture, Council on Foreign Relations, February 27, 2015. http://www.cfr.org/technology-and-science/artificial-intelligence-future-humans-robots-economy/p36197.

Cisco. 2015. "The Zettabyte Era—Trends and Analysis." White Paper, May. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html.

Hilbert, Martin, and Priscila López. 2011. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science,* April 1, 33(6025): 60–65.

Kuffner, James J. 2010. "Cloud-Enabled Humanoid Robots." IEEE-RAS International Conference on Humanoid Robotics.

Li, Fei-Fei. 2014. "The Digital Sensory System: A Quest for Visual Intelligence in Computers." Presentation at "Stanford Engineering's EngX: The Digital Sensory System" conference, May 20. https://www.youtube.com/watch?v=ylVsqXzlJqA.

Meeker, Mary. 2014. "Internet Trends 2014–Code Conference." May 28. http://www.kpcb.com/blog/2014-internet-trends.

Miret, Santiago. 2013. "Storage Wars: Batteries vs. Supercapacitors." Berkeley Energy and Resources Collaborative, November 10. http://berc.berkeley.edu/storage-wars-batteries-vs-supercapacitors.

Moore, Gordon. 1965. "Cramming More Components onto Integrated Circuits." *Electronics Magazine,* April 19, 38(8).

Moore, Gordon. 1975. "Progress in Digital Integrated Electronics." *IEEE, IEDM Tech Digest.* pp. 11–13.

Parker, Andrew. 2003. *In the Blink of an Eye: How Vision Sparked the Big Bang of Evolution*. Basic Books.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 61–80

Promises and Perils of Pre-Analysis Plans[†]

Benjamin A. Olken

Imagine a nefarious researcher in economics who is only interested in finding a statistically significant result of an experiment. The researcher has 100 different variables he could examine, and the truth is that the experiment has no impact. By construction, the researcher should find an average of five of these variables statistically significantly different between the treatment group and the control group at the 5 percent level—after all, the exact definition of 5 percent significance implies that there will be a 5 percent false rejection rate of the null hypothesis that there is no difference between the groups. The nefarious researcher, who is interested only in showing that this experiment has an effect, chooses to report only the results on the five variables that pass the statistically significant threshold. If the researcher is interested in a particular sign of the result—that is, showing that this program "works" or "doesn't work"— on average half of these results will go in the direction the researcher wants. Thus, if a researcher can discard or not report all the variables that do not agree with his desired outcome, the researcher is virtually guaranteed a few positive and statistically significant results, even if in fact the experiment has no effect.

This is of course the well-known problem of "data-mining." If the researcher can choose which results to report, it is easy to see how results can be manipulated. Casey, Glennerster, and Miguel (2012), for example, demonstrate in a real-world economics context how researchers with opposite agendas could hypothetically string together two opposite but coherent sets of results by cherry-picking either positive or negative statistically significant results.

■ *Benjamin A. Olken is Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. His email is bolken@mit.edu.*

[†] To access the Appendix and disclosure statements, visit
http://dx.doi.org/10.1257/jep.29.3.61

doi=10.1257/jep.29.3.61

The parable of a nefarious researcher offers the most straightforward version of the data mining problem, but similar problems can arise in less-extreme forms. For example, real-world data are messy, and are often "cleaned" before analysis—for example, to remove data outliers like a person whose height is reported in the data as being 20 meters tall instead of 2.0 meters. However, in many cases the issue of whether to "clean" the data of certain observations will involve a judgment call, and the researcher will often know how including certain observations will tend to affect the final results. There are also many decisions to make about specifications: what regression form to use, what control variables to include, what transformations to make to the data, how to define variables, and so on (Leamer 1983). Even researchers who have the noblest of intentions may end up succumbing to the same sorts of biases when trying to figure out how, in the process of their analysis, to make sense of a complex set of results.

One potential solution to these issues is to pre-specify in a precise way the analysis to be run before examining the data. A researcher can specify variables, data cleaning procedures, regression specifications, and so on. If the regressions are pre-specified in advance and researchers are required to report all the results they pre-specify, data-mining becomes much less of a problem. In the "confirmatory" trials used for approval of pharmaceuticals by the Food and Drug Administration, pre-specified statistical analysis plans are required that explicitly spell out how data will be handled—and these analysis plans must be finalized and archived before researchers actually run regressions on the unblinded data (Food and Drug Administration 1998).

But pre-specifying analysis plans comes at a cost. A pre-analysis plan is relatively straightforward to write if there is a single, simple hypothesis, with a single, obvious outcome variable of interest. But in practice, most research is much more complicated than this simple ideal. In economics, the typical research paper is trying to elucidate or test various predictions from economic theory, rather than estimate a single parameter with a single hypothesis test. Most research papers test a large number of hypotheses. Hypotheses are often themselves conditional on the realizations of other, previous hypothesis tests: the precise statistical question a paper might tackle in Table 4 depends on the answer that was found in Table 3; the question posed in Table 5 depends on the answer in Table 4, and so on.

Pre-specifying the entire chain of logic for every possible realization of the data can quickly become an overwhelming task for even the most committed pre-specifier. And in practice, researchers often get ideas for new hypotheses from seeing realizations of the data that they did not expect to see. The most rigid adherents to pre-specification would discount any such results that were not rigorously specified in advance of the data. Usually, though, these later additions to the analysis would be allowed, but would be considered "exploratory"—that is, not of the same rigorous statistical standards as confirmatory trials.

In a world with unlimited resources and unlimited time to make decisions, one could imagine a sequence of studies on any single topic. Exploratory analysis would be used to generate hypotheses, and then in turn subsequent, separate pre-specified

confirmatory trials would be run to test those hypotheses more rigorously. Exploratory analysis from those trials could form the basis of future trials, and so on. In practice, though, there are time and particularly budgetary constraints—true everywhere, but particularly so in economics where the entire budget of the National Science Foundation for social and economic sciences—about $100 million in 2013 (National Science Foundation 2013)—pales in comparison with the billions spent annually on drug trials, where pre-specification is most rigorous. Such constraints mean that most of these follow-up confirmatory trials will never be done, and the "exploratory" analysis is all the community will have to go on. Thus, the question of how much to discount such exploratory analysis in assessing the results of studies—either for journal publications or as the basis of policy—is a substantive question of serious importance.

The purpose of this paper is to help think through the advantages and costs of rigorous pre-specification of statistical analysis plans in economics. I begin by laying out the basics of what a statistical analysis plan actually contains, so that those researchers unfamiliar with the issue can better understand how it is done. In so doing, I have drawn both on standards used in clinical trials, which are clearly specified by the Food and Drug Administration, as well as my own practical experience from writing these plans in economics contexts.

I then lay out some of the advantages of pre-specified analysis plans, both for the scientific community as a whole and also for the researcher. Even researchers with the noblest of intensions may end up succumbing to their biases when trying to figure out how to make sense of a complex set of results, and pre-analysis plans can also be a useful tool when research partners have strong vested interests. I also explore some of the limitations and costs of such plans. I then review a few pieces of evidence that suggest that, in many contexts, the benefits of using pre-specified analysis plans may not be as high as one might have expected initially. I suspect the possible explanations include that most researchers are not nefarious and that existing safeguards place limits on the ability of researchers to data mine. Such safeguards may include referee and editor preferences for robustness checks, open availability of data, the possibility of replication, and the structure imposed by economic theory.

Most of my examples will focus on the relatively narrow issue of pre-analysis for randomized controlled trials.[1] Such studies fit the idea of a pre-analysis plan well, because they are designed and in place for a time before the data become available. However, the issues and tradeoffs I will discuss potentially apply to other empirical research in economics, too. In principle, there is no reason, for example, that a researcher could not completely pre-specify a statistical analysis plan before

[1] The registration of trials is a separate, though related, issue. If researchers only report those trials that happen to have a particular result, then the sample of trials that readers see will be biased. One solution to this issue is to register trials before the results are known. The American Economic Association sponsors a registry for this purpose for social science trials (http://www.socialscienceregistry.org/). For clinical trials in medicine, the US National Institute of Health sponsors a similar registry (https://clinicaltrials.gov/), which to date includes over 170,000 studies.

downloading the US Current Population Survey, or any other pre-existing datasets. While such an approach would be possible with existing datasets (for example, Neumark 2001), there is no obvious before-and-after date when a pre-analysis plan would be formulated and then later carried out, so doing so becomes more complicated. While perhaps such an approach could be useful, as some have advocated (for instance, Miguel et al. 2014), it is not something I explicitly consider here.

What Is a Pre-Analysis Plan?

The Basics: What Features Should Statistical Pre-Analysis Plans Include?

Virtually all pre-analysis plans typically share a few common features, summarized in Table 1. In describing these features, I draw heavily on accepted practice in perhaps the most rigorous and heavily regulated setting where they are used: the "Statistical Principles for Clinical Trials" specified for full-scale confirmatory trials used by the US Food and Drug Administration (1998) to approve drugs and other medical products. I will also discuss how these approaches may need to be adapted to a social science context. The interested reader may also wish to consult Casey, Glennerster, and Miguel (2012), which also discuss related issues in framing pre-analysis plans in economics.

A primary outcome variable. Given that one of the key motivations of pre-specifying an analysis plan is to avoid temptations for data mining, a key decision that needs to be made is the primary outcome variable one plans to examine to judge the outcome of a project. The idea is to solve the multiple inference problem by designating in advance a single outcome metric to evaluate the study. In designating the primary outcome variable, one should be as precise as possible: not just the general topic one intends to study, but the precise variable definition one intends to use.

Designating a single primary outcome variable can turn out to be surprisingly hard. In medical clinical trials, conventions have evolved concerning how to evaluate many topics, thus allowing comparability across studies, but in social sciences, more choices are available to the researcher. For example, suppose you are designing a study to evaluate an after-school tutoring program for disadvantaged youth. Possible outcomes could include school dropout rates, attendance rates, test scores, juvenile delinquency, teen pregnancy, and others. The researcher must make a choice of which outcome to focus on. If the researcher does this right, he or she will have substantially increased the believability of the research. Of course, one must choose carefully: if the study designated test scores as its primary outcome variable, and found no impact on test scores, but instead found that the program improved school dropout rates by an economically meaningful and statistically significant amount, the logic of pre-analysis plans suggests that policymakers should be much less likely to rely on those results than if the researcher had designated dropout rates beforehand as the primary outcome variable.

If a researcher wants to designate more than one outcome variable as primary, there are two options. First, one can designate multiple co-primary outcome variables,

Table 1
Pre-Analysis Plan Checklist

Item	*Brief description*
Primary outcome variable	The key variable of interest for the study. If multiple variables are to be examined, one should know how the multiple hypothesis testing will be done.
Secondary outcome variable(s)	Additional variables of interest to be examined.
Variable definitions	Precise variable definitions that specify how the raw data will be transformed into the actual variables to be used for analysis.
Inclusion/Exclusion rules	Rules for including or excluding observations, and procedures for dealing with missing data.
Statistical model specification	Specification of the precise statistical model to be used, hypothesis tests to be run.
Covariates	List of any covariates to be included in analysis.
Subgroup analysis	Description of any heterogeneity analysis to be performed on the data.
Other issues	Other issues include data monitoring plans, stopping rules, and interim looks at the data.

but a researcher who chooses multiple hypotheses needs to adjust the statistical tests for each hypothesis to account for the multiple inference hypotheses. The simplest way to do this, known as a Bonferroni adjustment (Dunn 1961), simply divides the required p-value by the number of tests conducted: thus, if a study chooses three outcome variables and tests at the 5 percent significance level, one would require each test to have significance $0.05/3 = 0.0166$ before it would be viewed as statistically significant. There are other, more sophisticated ways to multiple-inference adjust that have less of an impact on statistical power, like the step-down approach (for example, Westfall and Young 1993). But the general principle is that each additional co-primary outcome comes at a meaningful cost in terms of statistical power.

Second, a researcher can aggregate the primary outcome variable into an index or composite variable. If variables have comparable scales, one can take a simple average. Otherwise, the most common approach in economics is to compute "average standardized effects," where one divides each variable by its standard deviation and then takes the average of these normalized variables (Kling, Liebman, and Katz 2007). The index approach can be more powerful than a joint hypothesis test because an index lines up the variables so that "better" results tend to be averaged together in the same direction, whereas a joint test is agnostic about the sign of different results. Alternatively, one can use principal components analysis, which looks at the covariance between the individual variables and weights them accordingly. These various techniques create a single hypothesis test, rather than multiple hypothesis tests, which improves power. The potential downside of an index approach is that, if one finds results, it is hard to know statistically precisely what is driving the results. Policymakers may find it difficult to act based on a change in an index number.

Secondary outcome variables. Many pre-analysis plans also specify secondary outcome variables, which are outcomes that may help shed light on the findings but would not themselves be "confirmatory." For example, if the Food and Drug Administration were considering whether to approve a drug, and a trial found meaningful results on a secondary outcome but not on a primary outcome, the drug would generally not be approved. In social science papers, secondary outcome variables often play a crucial role, because they illuminate the "mechanisms" or pathways that lie behind the results, which in turn helps guide the researcher in both enhancing understanding of the problem and in being able to say something sensible about external validity. Outside of regulatory contexts where secondary outcomes have a precise meaning (in particular, drug makers can be allowed to market a drug based on a proven secondary outcome if it is listed in advance and if they also found results on the primary outcome), researchers are in practice often somewhat laxer about multiple inference testing with secondary outcome variables. As I will discuss in more detail below, the pre-specification of secondary outcomes can become quite challenging in social science papers, because the set of secondary outcome variables to be examined depends on the results from primary outcome variables.

Variable definitions. A pre-analysis plan requires a *precise* variable definition. Continuing the earlier example, suppose that test scores are the primary outcome of interest. What test and test subjects are included? Will the outcome variable be the test score in levels or logs? Will it be in standard deviations, the percentile of the test score, a binary variable for passing the test, a binary variable for being above the 25th percentile, or the 50th percentile, and so on? Will the score be in levels or an improvement from a baseline? If there are multiple subjects, like math and reading, how will the scores be aggregated into a single outcome variable? Are there any rules for trimming or excluding outliers? A good rule of thumb is that if you gave the pre-analysis plan to two different programmers, and asked them each to prepare the data for the outcome variable, they should be both able to do so without asking any questions, and they should both be able to get the same answer.

Inclusion or exclusion rules. A precise set of rules lead to the "analysis set"—that is, the final set of data to be analyzed. As a general principle, of course, the analysis set should be as close as possible to the actual observations. However, if there are legitimate reasons to drop observations, they should be specified in advance in the analysis plan. Relatedly, one should discuss the plans for handling missing values and attrition, although a challenge is that one cannot always foresee the reasons one might want to exclude certain observations.[2]

[2] For example, in Kremer, Miguel, and Thornton's (2009) study of scholarships in Kenya, several schools withdrew from the study after a Teso-ethnicity school was hit by lightning and some Teso-ethnicity community members associated the lightning strike with the nongovernmental organization running the scholarship program. In some specifications, the authors restrict analysis to schools that did not withdraw due to this concern. A lightning strike seems like exactly the sort of legitimate reason one might want to exclude observations, but the possibility of lightning strikes and superstitious villagers would have been very hard to foresee in a pre-analysis plan.

Statistical model and covariates. An analysis plan should spell out the precise estimating equation and statistical model, including functional form and estimator (ordinary least squares, probit, logit, Poisson, instrumental variables, and so on). If fixed effects are going to be used, or comparisons to baseline values, or first differences of data, all this should be spelled out. The pre-analysis plan also states how standard errors will be treated (including any clustering, bootstrapping, or other techniques). If one is using nonstandard hypothesis tests, and in particular one-sided tests, it should be spelled out in advance.

Specification of the model should also be clear about which covariates should be included in regressions, because a typical study might collect tens or even hundreds of variables that could, potentially, be included as covariates. After all, researchers could potentially cherry-pick control variables to maximize statistical significance. Relatedly, it has become standard practice in most randomized controlled trials in economics to present a table showing that baseline covariates appear balanced across treatment and control groups. If the authors intend to present a balance test, it is also common sense to pre-specify in the analysis plan the variables that will be used to check covariate balance.

Subgroup analysis. Pre-specification of subgroup analysis matters because there are many possible ways of cutting the data into various subgroups—men versus women, old versus young, rural versus urban, and so on. Again, researchers could first do the analysis and then pick a subgroup with a statistically significant result, which is a frequent critique of some randomized trials in development economics (Deaton 2010). If heterogeneity analysis is likely to be important, pre-specification can be quite helpful to increase confidence in the results.

Other aspects. Other issues that are often considered in pre-analysis plans in the medical world include data monitoring plans, safety checks, stopping rules, and interim looks at the data. In particular, in medical trials one often checks the data in the middle of the trial to ensure that the outcome is not causing unexpected harm (in which case the trial might be stopped) and to learn whether results are so good that the trial can be declared a success early. A recent area of research has been to allow for adaptive trials, which are trials whose design evolves over time based on the data but according to pre-specified rules (Food and Drug Administration 2010). These issues can all be pre-specified in the analysis plan.

When Should You Write a Statistical Analysis Plan?

In the classic Food and Drug Administration model, the primary outcome and usually secondary outcomes would be specified in the formal trial protocol before the trial begins. However, the statistical portion of the pre-analysis plan can be finalized later—including issues such as covariates, regression specification, and handling of missing data—as long as it is written without ever unblinding the data, that is, without ever separating the data by treatment groups.[3]

[3] In fact, it is possible in rare circumstances to change the primary outcome variable of a trial once the trial has begun, if one can demonstrate that the original primary outcome variable no longer makes

Allowing researchers to design the statistical portion of their pre-analysis plans based on the blinded outcome data can be quite useful. In many cases in social science, the outcome variables that people study are sufficiently novel, and the data on relevant populations is sufficiently limited, that researchers have only limited information about the distribution of the variables when designing studies. For example, imagine that one of the variables in the study is the level of juvenile delinquency. Presumably, the researcher has some informed guess about the expected mean and standard deviation for this variable. But perhaps in this particular dataset, the standard deviation is much larger than usual. (Perhaps there was an unusual crime wave during the period of the study, or for some reason the study sample differs from the population.) Looking at the blinded data helps the researcher to discover if the outcome variables behave sensibly—that is, if they have reasonable relationships with each other and with the covariates—which helps to assure that the variables were measured well. If not, they can be excluded. Another use is to examine the blinded data to determine which covariates best predict the outcome variable, reducing standard errors by reducing the variance of the residuals. Especially when the outcome data or covariates are novel variables, it can be useful to examine the actual blinded data for this purpose.[4]

In my experience, it can be quite useful to write statistical programs, run them on the blinded data, and use the results to update the statistical analysis plan in the process. Indeed, one can generate a "fake" randomization—that is, one can run and rerun a randomization program with different starting seed values to generate the actual standard errors one would expect when running regressions. The exercise of writing the computer code and looking carefully at the data also forces the authors to make detailed choices about variable definitions and coding: for example, a researcher can make decisions about how to exclude outliers before knowing whether they are in treatment or control groups and how their exclusion will affect the results.

A trickier issue is the use of qualitative data, particularly for many social science trials, which are often not blinded (that is, both those administering the trial and the subjects know who is in the control group and who is in the treatment group). In this context, even if the statistical data is blinded, one may learn something from the qualitative findings of the trial. For example, one might observe that those in the after-school support program seem to be happy, so one might think to add subjective well-being measures to an analysis protocol. Even though

sense. For example, suppose that the primary outcome variable of a study was mortality, but the blinded data revealed that the overall mortality rate was much lower than expected and the trial was underpowered. It might be possible then to amend the trial protocol to change the outcome variable to be a combination of mortality and morbidity.

[4] Note that in doing so, it is often advisable to look at the complete, blinded data rather than looking at the control group and hiding the treatment group. There are a number of reasons for this. One reason is that in practice, researchers will have often seen summary statistics for the entire data: if one has, and also sees the control group, one can subtract to obtain the treatment group estimates. It is also easier to ensure that the data are not accidentally unblinded if the treatment and control assignments are kept entirely separate from the data one is using to construct the analysis plan.

this is based on qualitative observations, not the quantitative data, it has the same effect as looking at the unblinded data. For this reason, trial purists may prefer analysis plans to be finalized before the trial begins. The degree to which this makes sense depends on weighing the benefits from making sure this type of qualitative bias does not enter, against the costs in terms of missed advantages from trying out the analysis on blinded data, as discussed above.

What Do You Do With a Statistical Analysis Plan after You Write It?

Ideally, a pre-analysis plan should be added to a public archive. As discussed above, the American Economic Association operates a trial registry, where authors can also archive a statistical analysis plan, with specific timestamps marking exactly when it was registered. The timestamps can credibly convey to the reader that it was filed before all data was collected, for example, and that it was not modified later. The registry also allows authors to register a statistical analysis plan but not make it public until a later date (or not at all). In this way, authors who are concerned about others scooping their work could obtain the credibility benefits of pre-registration—that is, they could document to editors or referees that their analysis plan was pre-registered—while avoiding publicity about their work months or years before it is complete.

Benefits of Pre-Analysis Plans in Economics

The most obvious benefit from pre-specification is that a careful pre-analysis plan can address a substantial proportion of data-mining problems. For readers, referees, editors, and policymakers, knowing that analysis was pre-specified offers reassurance that the result is not a choice among many plausible alternatives, which can increase confidence in results.

However, a pre-analysis plan also offers some other useful benefits for researchers themselves, which are perhaps less obvious and therefore worth elaborating in further detail. The exercise of creating a pre-analysis plan can be useful for researchers to make sure that they think through, and collect, the data they need for the analysis. Beyond that, the act of commitment to an analysis plan per se offers some additional advantages.

First, pre-specified analysis plans allow researchers to take advantage of all the statistical power of well-designed statistical tests and to worry less about robustness to specifications. After seeing the results, it can be challenging for even well-intentioned researchers not to choose specifications that lead to more statistical significance—well-intentioned researchers might conclude, for example, that the specification that led to the smallest standard errors was the one that best fit the data, and it is hard to prefer intentionally a specification that makes ones' results look weaker. But given this, if specifications are not pre-specified, researchers will be required by referees and editors to report robustness results to a wide range of alternative specifications and will likely judge results by the average level of statistical significance across specifications rather than use the statistical significance from the preferred

specification. A pre-specified analysis plan could help discourage readers of the article—including journal referees—from expecting an endless series of robustness checks and accepting only those results that survive all possible permutations.

A second benefit to researchers, related to the first, is that pre-commitment can also potentially allow researchers to increase statistical power by using less-conventional statistical tests, if they really believe that such tests are appropriate in a given case, because they know that pre-committing to such a test means that they cannot be justly accused of cherry-picking the test after the fact. For example, convention typically dictates two-sided hypothesis tests, so that researchers can reject the null hypothesis of no effect of a program at the 5 percent level if the estimate is in the upper or lower 2.5 percent of the distribution under the null hypothesis. In practice, however, researchers are often interested in only knowing whether a program works or not, which could lead to a one-sided hypothesis test. Such a researcher might instead use a one-sided test rather than a two-sided test. Of course, a one-sided test has trade-offs, too. By committing to a one-sided test, researchers need to be prepared that even if they receive a very, very negative outcome—for example, an outcome in the bottom 0.005 percentile—they would interpret that outcome as no different from the null rather than report a negative and statistically significant result. The prospect of such a result is uncomfortable enough to cause most researchers to prefer two-sided tests.[5] Moreover, in addition to hypothesis tests, researchers often want to report confidence intervals. In a one-sided testing framework, the confidence interval has an infinite bound on one side, which may be less useful from a decision-making perspective. Clearly, there are often reasons for the conventional choices, like two-sided hypothesis tests, and researchers should proceed cautiously before pre-committing to alternatives.

A final major benefit to researchers is that pre-specification can be useful vis-à-vis their research partners. In practice, a substantial share of large-scale randomized controlled trials and other large-scale social science research is done in collaboration with partners who have some vested interest in a program's outcomes, like the government, a nongovernment organization, or a private sector program or firm. Even if sponsors don't have explicit rights of review of articles or research findings, researchers often develop close relationships with partners over time, which can potentially lead to awkward situations when results do not come out the way partners might have hoped. By creating an opportunity for researchers and partners to agree and commit before results are observed on how the program will be evaluated, pre-specification can provide researchers with protection from pressure from their partners to slant results in a more favorable light.

[5] An interesting hybrid alternative would be to pre-specify asymmetric tests: for example, to reject the null if the result was in the bottom 1 percent of the distribution or in the top 4 percent, or the bottom 0.5 and the top 4.5 percent, and so on. These asymmetric tests would gain much of the statistical power from one-sided tests, but still be set up statistically to reject the null in the presence of very large negative results. One could apply decision theory, where one specifies losses for each type of error, to determine the appropriate asymmetric approach to use. Although I have not seen this approach taken in economics, it seems like a potentially useful approach for researchers to consider.

Costs of Pre-Analysis Plans in Economics

When laid out in this way it seems hard to be against pre-analysis plans. After all, how can one argue against the idea that one should do hypothesis testing properly to get correct *p*-values, and that the research community should protect itself against data mining? However, restricting analysis to pre-specified hypotheses has some fairly important costs, which need to be weighed against the benefits.

One important challenge is that fully specifying papers in advance is close to impossible. Economics papers typically ask not just the result of a treatment, but also try to elucidate the mechanisms that underlie the treatment, such that the results quickly become too complex to map out in advance. For example, suppose that a paper has one main table and then ten follow-up tables of results, and each table can have three possible results—"positive," "zero," and "negative." In addition, suppose the question one would want to ask in each table depends on the outcome of the previous table. Pre-specifying the entire "analysis tree" would therefore involve writing out $3^{10} = 59{,}049$ possible regressions in advance. Even for the most dedicated pre-specifier, thinking through 59,049 possible regressions in advance would clearly be too taxing. It would also be inefficient—we would prefer that researchers spend their time and intellectual energy on those parts of the tree that are actually relevant rather than working down the branches of the tree that are meaningless.

Faced with this conundrum, researchers can take several possible tacks. First, they can try to pre-specify as much as possible. Many early pre-analysis plans in economics ended up voluminous, exceeding 50 pages, trying to pre-specify not just primary outcome variables but all of the secondary tests one would want to run conditional on results from the primary outcome variable. The result can be an unwieldy paper that reports all kinds of results that are not primarily of interest to the reader since they were relevant only conditional on realizations of the primary outcome variable that did not materialize.

More important, because researchers are spreading their thinking energy over the entire space of possible regressions they might want to run, they often do not focus on aspects of the space that end up being important, and they may miss out on important hypotheses. After all, scientific breakthroughs sometimes come from unexpected surprise results.[6] Limiting oneself only to the regressions that are pre-specified, and not including or severely discounting any additional analysis of the data inspired by surprise results, seems an inefficient way to learn the most from the data at hand.

[6] Limiting oneself strictly to pre-specified analysis at some point becomes absurd. Easterly (2012), for example, imagines what would have happened if Christopher Columbus had had to pre-specify an analysis plan for his 1492 voyage to the Indies: "(2) Columbus gets funding to test Going West approach [to reach the Indies]. (3) Rigorous Evaluation design filed in advance to test effect of Going West on Reach the Indies. (4) October 12, 1492. [Columbus discovers America.] (5) Rigorous Evaluation finds no effect of Going West approach on Reach the Indies. (6) Rigorous methodology means Evaluation not permitted to include any ex post outcomes of Going West not filed in advance in the design. (7) Going West approach declared ineffective, funding ends."

An alternative, more moderate approach is to focus the pre-analysis plan on a single primary outcome (or a narrow set of such outcomes), and then leave the remainder of the paper for exploring potential mechanisms as only exploratory and not pre-specified. Of course, this strategy would be explained as such in the pre-analysis plan and the article itself. In some sense, this is how pre-specification is supposed to proceed: a given trial is designed to test a single, well-specified hypothesis, and then the data is used in a variety of exploratory ways to come up with new hypotheses that are in turn then the subject of future pre-specified trials.

The problem with narrow pre-specification and extensive exploratory analysis is that, in practice, there are not enough resources to conduct repeated streams of separate trials simply to solve the pre-specification issue. Budgets for social science research are several orders of magnitude smaller than for medical research—and even in medicine, some journals would acknowledge that for many less-common areas, the exploratory results may be the only results that the scientific community will have.[7] The difference in magnitudes here is enormous: the registry for medical trials, http://clinicaltrials.gov, currently lists over 176,000 studies registered since the site was launched; by comparison, a reasonable estimate for the number of randomized controlled field experiments conducted in social science over a similar period is on the order of 1,000.[8]

This argument does not imply that researchers running any given trial would be better off by not pre-specifying analysis for that trial. But it does suggest that if journal editors were to restrict themselves to publishing studies based on the limited, pre-specified, confirmatory parts of analysis, and relegating exploratory analysis to second-tier status, a substantial amount of knowledge would be lost. We do not have near-infinite resources to run sequences of pre-specified trials iteratively, where each set of exploratory analysis from one trial was the subject of a subsequent, pre-specified confirmatory trial, and so it seems important to continue to allow researchers to publish, and the broader community to use, important results that were not necessarily pre-specified.

A related issue is that papers following rigorous pre-specified analysis plans may miss the nuance that categorizes social science research. Pre-analysis plans work particularly well for relatively simple papers: there was a trial of some particular

[7] For example, the total annual 2014 National Science Foundation budget for Social and Economic Sciences is $102 million. By comparison, the total National Institutes of Health budget in 2014 is approximately is $30.1 billion, and this does not include the billions spent by the private sector on clinical trials for pharmaceuticals and other medical products. While both of these budgets fund many activities that are not randomized trials, the difference in scale is remarkable.

[8] We do not have a precise number of trials in social science over this period. However, members of several of the largest organizations supporting such trials in development economics, the Abdul Latif Jameel Poverty Action Lab, Center for Effective Global Action, and Innovations for Poverty Action have each completed or are in the process of running about 500 trials since their respective founding in the early 2000s; since many of these trials are counted by several of these organizations, the total is likely closer to 750 or so. We do not have a formal count of other trials in economics outside these organizations (for example, trials run by the World Bank or organizations like MDRC are not included in these totals), but it seems safe to say that the total is on the order of a few thousand at most.

intervention, there is some key outcome metric to decide if it "works" or not, and the researcher compares that outcome across treatment and control groups. This framework naturally leads one to specify a primary outcome variable (the metric of whether the program "works" or not), and so on.

However, many empirical economics papers are instead seeking to test theoretical mechanisms to see whether they are borne out in practice. In many contexts, the point of the study is not just that this particular trial had this particular effect, but rather to show the existence in practice of a theoretically posited mechanism that may be of use elsewhere. Papers thus use a constellation of tests to elucidate economic mechanisms and test theories. While it may be possible to pre-specify complex papers, as discussed above, given the exponentially increasing challenges of pre-specifying complex analysis trees, pre-specification of analysis works best for simple setups, when there is a clear "primary outcome" or set of primary outcomes. One would not want the quest for pre-specification to come at the cost of writing only simple papers and losing the nuance that characterizes some of the best social science work.

A more prosaic but still important concern with requiring pre-analysis plans involves the intricacies of needing to monitor program implementation using unblinded data while at the same time finalizing the analysis plan based on blinded data—all with a limited staff. In principle, there are two distinct things one would like to do with the data while the trial is ongoing. First, as discussed above, one would like to look at the *blinded* data before finalizing the pre-analysis plan to improve the plan: for example, by checking means and standard deviations, doing data cleaning on the blinded data, or even just having more time to reflect on how to analyze the trial after the effort of launching the fieldwork has been completed. Second, one would also like to look at the *unblinded* data while the trial is ongoing to provide real-time feedback to implementing partners, ensure that implementation is going on correctly, and so on. For example, interim looks at the data can be used in a medical trial to see if a drug is causing an adverse reaction, to know if the trial should be stopped. One can similarly imagine that a business or government that is partnering with a social science researcher in a trial may require ongoing analysis of the trial to ensure that the experiment is not actively harming their business or program. Often follow-up trials need to be planned before a trial is complete, so interim looks at the data can be useful for that purpose as well.

In principle, in a really large research team, one could have two different sets of sub-teams, one looking at the blinded data during the trial and refining the analysis plan, and one looking at the unblinded data during the trial for management and safety purposes. In medicine, budgets are large enough that one can really have two completely different teams of people doing these tasks, with a firewall between them. For example, many medical trials have separate Data Monitoring Committees (DMCs) regularly inspecting unblinded data to verify safety and implementation—while keeping the unblinded data securely away from principal investigator's eyes. But many social science trials are on sufficiently tight budgets that having two separate teams of people for these two tasks isn't feasible. In that sense, requiring a

pre-analysis plan comes at a cost, since the researcher must forego one or the other of these during-trial activities.

A final, if perhaps less-persuasive, cost of pre-specification is that it prevents you from learning about your data as you analyze it. As all researchers who have worked with empirical data realize, a myriad of real-world issues arise: how should variables be defined, how to deal with outliers, and so on. In principle, perhaps, there is no reason that these issues cannot be sorted out on blinded data, and programs written in advance. In practice—much for the same reason that it is hard to think through every possible regression in advance—researchers frequently realize features of their data only during the process of analysis. For example, seeing surprisingly large standard errors on a regression may make authors realize that the distribution of a variable was more skewed or plagued with outliers than they had initially appreciated. Addressing these problems iteratively as they come up raises the possibility of data mining, but preventing researchers from dealing with these issues if they come up also may limit the amount we can learn from a given study.

These costs should not be necessarily viewed as dispositive, or arguing against pre-analysis plans in all cases. However, they do suggest that the degree to which requiring pre-analysis plans makes sense for the discipline depends on the extent to which the key problem—data mining—is actually a problem, an issue I explore in the next section.

Is There Much Need for Pre-analysis Plans in Practice?

How Bad Is the Problem?

The arguments in the previous section suggest that pre-specification of analysis has important benefits—preventing data mining and specification searching, limiting influence of partners, and so on. But imposing standards such that the only analysis that the scientific community trusts or that journal editors are willing to consider for publication is pre-specified also has costs. Authors may be limited in their ability to learn during the process of analysis, and as such will likely write papers of less-general interest focused only on those hypotheses that were pre-specified rather than on more potentially interesting findings discovered later.

The extent to which the community should reward pre-specification therefore depends, in practice, on how substantial the data mining concerns are. That is, many of the arguments in favor of pre-specification assume the worst about researchers: they are inherently biased and data mine as much as possible until they find results. But how common is the nefarious researcher in practice?

Several recent studies in social science suggest the problem is not as bad as the pessimists might believe. One recent study by Brodeur, Lé, Sangnier, and Zylberberg (forthcoming) tried to quantify the extent to which there is inflation of *p*-values through specification mining. The strategy was to look at the distribution of *p*-values in a wide range of studies, and to find out whether the *p*-values are bunched just below critical statistical significance values of 0.10 or 0.05. Brodeur

et al. examined all empirical regressions from the *American Economic Review*, *Journal of Political Economy*, and *Quarterly Journal of Economics*, between 2005 and 2011, examining over 50,000 regression results from 3,389 tables in 641 articles. They do find bunching of *p*-values, in a way that suggests that between 10 and 20 percent of all tests that show *p*-values in the conventional range of statistical significance between 0.05 to 0.0001 are in fact misallocated and instead should be in the range, not thought of as statistically significant, between 0.10 to 0.25. However, Brodeur et al. find no evidence of this problem arising in randomized trials, which suggests that at least as detected by their methodology, there is little bias problem in the context for which pre-analysis plans are most applicable.

Even to the extent Brodeur et al. (2013) do find excess bunching, their results imply that it may not be quantitatively as severe as one might have thought. Their results imply that out of 100 studies, instead of obtaining a nonsignificant result in 95 percent of studies where the null is in fact correct—as one would expect with a *p*-value of 0.05—we are in fact doing so for 92.75 percent of such studies.[9] Addressing this problem would be beneficial, but if it came at the cost of substantially excluding a variety of important and interesting findings that were discovered in after-the-fact analysis, it might not be worth the cost.

An alternative approach to searching for publication bias is to carry out the study again, in as similar a way as possible. Replication of large-scale field studies in economics is rare; in fact, given the costs of these studies and limited budgets, it probably makes sense in most cases to prioritize new experiments rather than funding replications of existing experiments. However, in social psychology where experiments can be conducted in the lab there have been some attempts to replicate main findings. A recent paper by Klein et al. (2014) reports an enormous effort (with more than 50 coauthors) to replicate 13 well-known psychology findings using labs around the world. Roughly speaking, they found that (depending on the standard applied) 10 or 11 of the 13 studies replicated well. Recall that even with correct 0.05 *p*-values, we would not at all be surprised if 1 out of the 13 (7.6 percent) failed to replicate. We would also not be surprised if some studies did not replicate given changes in subject pools, changes in experimenters, and so on. So on balance, while there appears to be evidence of a slight inflation of statistical significance, this replication-based approach suggests that in this context, major findings are holding up reasonably well.

Why Isn't the Problem Worse?

Economists and other social scientists may be closer to the world of correct *p*-values than to the world of the nefarious researcher who is cherry-picking results left and right. I suspect there are several reasons for this.

[9] To be concrete, let us suppose that 15 percent of tests should have *p*-values of 0.20 instead of 0.05. What would this mean for inference? It implies that the "correct" *p*-value, conditional on seeing a *p*-value of 0.05 and knowing that 15 percent of them should have *p*-values of 0.20, is $(0.85 \times 0.05) + (0.15 \times 0.20) = 0.0725$.

First, theory combined with experimental design provides some guidance that limits the degree to which researchers can engage in data mining. In many contexts, the primary outcome variable or variables for a given study will be fairly obvious. If you are studying an intervention to reduce teacher absence (as in Duflo, Hanna, and Ryan 2012), it is reasonably clear that you should show results on teacher absence and probably also results on student test scores. Any reasonable reader, referee, or journal editor would ask for such results if the authors did not report them. While there is some degree of manipulation researchers could do (for example, by reporting only math test scores and not language test scores), it is substantially limited by the expectations of readers concerning what outcomes the researchers should want to examine.

Second, authors are typically required to both show robustness and, for many journals, to make their data publicly available. Showing that main results are robust to a variety of specifications is statistically inefficient, because it means that papers are often judged by the average *p*-value across all specifications, rather than by a single, correctly specified *p*-value, but it has the advantage of making sure that authors are not systematically manipulating specifications to artificially improve their results. Making data available provides another check to make sure that researchers do not wildly mis-analyze their data. For example, the *American Economic Review*, and the other journals of the American Economic Association, along with *Econometrica,* the *Journal of Political Economy*, and the *Review of Economic Studies*, all require publication of data and programs for accepted articles.

Third, and perhaps most important, most academic researchers probably do not behave as the "nefarious" straw man I discussed in the beginning. To be sure, there are strong career and funding incentives, and everyone likes having strong and statistically significant results rather than statistically imprecise mush. But economics has no equivalent of the pharmaceutical trials where billions of dollars may depend on whether a single *p*-value is 0.049 or 0.051.

What Do Actual Papers Look Like?

To assess some of the challenges with pre-analysis plans in practice, I examined a set of recent papers that were using randomized controlled field trials. In particular, I looked at all such papers from the *American Economic Review*, *Quarterly Journal of Economics*, *Econometrica*, *Review of Economic Studies*, and *Journal of Political Economy* published from the start of 2013 until the middle of 2014: a total of 18 papers.[10] It is worth noting that none of these papers (as far as I could tell) had pre-analysis plans, which illustrates the degree to which pre-analysis plans are currently the exception, not the norm, in the economics profession.

For each of these papers, I examine the number of "primary" outcome variables and then the number of "conditional" tables of regressions, which potentially might have been specified in a different way if the primary outcome variables had

[10] The papers are listed in an online Appendix available with this paper at http://e-jep.org. I particularly thank John Firth for his help with this analysis.

realizations other than the ones that actually occurred. Since economists don't usually officially designate which outcomes are primary and which are secondary, and we cannot know for sure which tables would have been run conditional on the particular realization of outcomes and which would have been run regardless, this requires some judgment calls. Nevertheless, the exercise is useful to gauge some patterns and magnitudes.

First, these papers are complicated. The median paper has four treatment arms—three treatments groups and one control group—along with four main outcome variables. If we assume that each outcome variable could be positive, zero, or negative compared to the control group, that implies that each treatment arm has $3^4 = 81$ possible configurations of outcomes. Across three treatments, there are $81^3 = 531{,}441$ possible configurations of outcomes vis-à-vis controls. Second, it is common to look at secondary outcomes. The median paper in this group has 6.5 secondary outcomes, in addition to the primary outcomes. Third, I examine whether papers seem to be hovering near borderline statistical significance. If one was concerned that data mining was prevalent, one might expect most of the statistically significant *p*-values to be close to the 0.05 threshold.[11] However, these papers as a group are publishing statistically significant outcomes that are not close to the 0.05 threshold; they are much more statistically significant than that. Fourth, most of these papers use the robustness approach to convince readers that results are not spurious: specifically 10 of 18 papers show robustness tests to include controls of various types. Finally, 13 of the 18 papers examine subgroup heterogeneity.

This analysis suggests that complete pre-specification is not going to work without losing certain nuances that seem common in papers currently in top journals in economics. For example, supposing only one layer of conditionality, there are 531,441 possible combinations of primary outcome variables and results. Even if theory provides some guide for grouping these outcomes together, clearly the number of cases one would need to consider in writing a pre-analysis plan quickly becomes insuperable. Moreover, *p*-values are much more significant than 0.05, suggesting that fiddling around the margins is unlikely to be driving statistical significance in most of these studies. While the frequent use of heterogeneity analysis suggests that pre-specifying these issues may be important, overall these examples give some pause to the idea that requiring, or even strongly privileging, pre-specification for journal publication would on net improve the amount we learn from these trials.

[11] Specifically, for all statistically significant main outcomes (that is, all outcomes with *p*-values below 0.05), we calculate the *z*-statistic associated with it, and take the average. Across all significant outcomes in all papers, the average *z*-statistic is 3.18, which would correspond to a *p*-value of 0.0014. By comparison, if *p*-values were uniformly distributed between 0.00 and 0.05, one would expect an average *z*-statistic of 2.33, which would correspond to a *p*-value of 0.02; if there was substantial *p*-hacking, one might expect *p*-values closer to 0.05 and even lower average *z*-statistics. The reason it is not an average of 0.025 is because very low *p*-values have disproportionately high *z*-statistics, so the average *z*-statistic does not correspond to the average *p*-value.

Thoughts on the Way Forward

Economics papers tend to be complicated, and pre-specifying the entire chain of analysis is probably impossible for the median paper in economics. Forcing all papers to be fully pre-specified from start to end would likely result in simpler papers, which could potentially lose some of the nuance of current work. If economists were to exclude from publication or policy consideration all non-pre-specified, exploratory results in the name of increased transparency, we would be losing more than we would gain.

That said, in many contexts, pre-specification of one (or a few) key primary outcome variables, statistical specifications, and control variables offers a number of advantages. In cases where there is a partner with any kind of vested interest in the outcome, pre-specification of outcomes and analysis can be a huge advantage to all parties. Even when there is not a strong interested party, the rigor of researchers specifying a small number of primary outcomes in advance is a useful exercise that will help ensure that when data are analyzed, they know what to focus on. For the many decisions where there is no clear hard decision to make—what statistical model to use, what control variables to include, and so on—pre-specification frees the author from the need to report a large number of robustness checks and in so doing make their effective statistical power worse than it needs to be. Even if journals do not require pre-specification, individual researchers may choose to do so in order to enhance the credibility of their results, and mechanisms like the AEA registry that allow them to commit publicly to pre-registration can be useful to allow them to do so.

■ *I thank Abhijit Banerjee, Paul Catalano, Esther Duflo, Amy Finkelstein, Marc Fisher, Rachel Glennerster, Lisa LaVange, Heather Lanthorn, Edward Miguel, Brian Nosek, Sharon-Lise Normand, Robert O'Neil, Uri Simonsohn, Robert Temple, Marta Wosinska, and participants at the Berkeley Initiative for Transparency in Social Sciences conference for many helpful discussions on these topics; Gordon Hansen, Enrico Moretti, and Timothy Taylor for helpful editorial suggestions; and John Firth for comments and research assistance. I also thank my many coauthors with whom I have worked on preparing these analysis plans from scratch for our research. The views in this paper are those of the author alone and do not represent any of the individuals acknowledged here or their respective institutions.*

References

Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household Bargaining and Excess Fertility: An Experimental Study in Zambia." *American Economic Review* 104(7): 2210–37.

Avvisati, Francesco, Marc Gurgand, Nina Guyon, and Eric Maurin. 2014. "Getting Parents Involved: A Field Experiment in Deprived Schools." *Review of Economic Studies* 81(1): 57–83.

Balafoutas, Loukas, Adrian Beck, Rudolf Kerschbamer, and Matthias Sutter. 2013. "What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods." *Review of Economic Studies* 80(3): 876–91.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics* 129(2): 697–752.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128(1): 1–51.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. Forthcoming. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics.*

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions." *Econometrica* 82(4): 1273–1301.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *Quarterly Journal of Economics* 127(4): 1755–1812.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment." *Quarterly Journal of Economics* 128(2): 531–80.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics* 128(3): 1169–218.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48(2): 424–55.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128(4): 1499–545.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102(4): 1241–78.

Dunn, Olive J. 1961. "Multiple Comparisons among Means." *Journal of the American Statistical Association* 56(293): 52–64.

Dupas, Pascaline. 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica* 82(1): 197–228.

Dupas, Pascaline, and Jonathan Robinson. 2013. "Why Don't the Poor Save More? Evidence from Health Savings Experiments." *American Economic Review* 103(4): 1138–71.

Easterly, William. 2012. "If Christopher Columbus Had Been Funded by Gates." Blog post, NYU Development Research Institute, October. http://www.nyudri.org/2012/10/15/if-christopher-columbus-had-been-funded-by-gates/.

Eriksson, Stefan, and Dan-Olof Rooth. 2014. "Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence from a Field Experiment." *American Economic Review* 104(3): 1014–39.

Feigenberg, Benjamin, Erica Field, and Rohini Pande. 2014. "The Economic Returns to Social Interaction: Experimental Evidence from Microfinance." *Review of Economic Studies* 80(4): 1459–83.

Field, Erica, Rohini Pande, John Papp, and Natalia Rigol. 2013. "Does the Classic Microfinance Model Discourage Entrepreneurship among the Poor? Experimental Evidence from India." *American Economic Review* 103(6): 2196–226.

Food and Drug Administration. 1998. "Guidance for Industry: E9 Statistical Principles for Clinical Trials." US Department of Health and Human Services. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf.

Food and Drug Administration. 2010. "Adaptive Design Clinical Trials for Drugs and Biologics." Draft Guidance. U.S. Department of Health and Human Services. http://www.fda.gov/downloads/Drugs/.../Guidances/ucm201790.pdf.

Jessoe, Katrina, and David Rapson. 2014. "Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use." *American Economic Review* 104(4): 1417–38.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. 2014. "Agricultural Decisions after Relaxing Credit and Risk Constraints." *Quarterly Journal of Economics* 129(2): 597–652.

Klein, Richard A. et al. 2014. "Investigating Variation in Replicability: A "Many Labs" Replication Project." *Social Psychology* 45(3): 142–52.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83–119.

Kremer, Michael R., Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *Review of Economics and Statistics* 91(3): 437–56.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment." *Quarterly Journal of Economics* 128(3): 1123–67.

Leamer, Edward E. 1983. "Let's Take the Con out of Econometrics." *American Economic Review* 73(1): 31–43.

Miguel, E., et. al. 2014. "Promoting Transparency in Social Science Research." *Science* 343(6166): 30–31.

National Science Foundation. 2013. "Social, Behavioral, and Economic Sciences." https://www.nsf.gov/about/budget/fy2013/pdf/10-SBE_fy2013.pdf. In *National Science Founation FY 2013 Budget Request to Congress.* https://www.nsf.gov/about/budget/fy2013/.

Neumark, David. 2001. "The Employment Effects of Minimum Wages: Evidence from a Prespecified Research Design." *Industrial Relations* 40(1): 121–44.

Tarozzi, Alessandro, Aprajit Mahajan, Brian Blackburn, Dan Kopf, Lakshmi Krishnan, and Joanne Yoong. 2014. "Micro-loans, Insecticide-Treated Bednets, and Malaria: Evidence from a Randomized Controlled Trial in Orissa, India." *American Economic Review* 104(7): 1909–41.

Westfall, Peter H., and S. Stanley Young. 1993. *Resampling-based Multiple Testing: Examples and Methods for* p*-value Adjustment.* Hoboken, NJ: John Wiley & Sons.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 81–98

Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible

Lucas C. Coffman and Muriel Niederle

The social sciences—including economics—have long called for transparency in research to counter threats to producing robust and replicable results (for example, McAleer, Pagan, and Volker 1985; Roth 1994). Recently, the push for transparency has focused on a few specific policies. In this paper, we discuss the pros and cons of three of the more prominent proposed approaches: pre-analysis plans, hypothesis registries, and replications. While these policies potentially extend to all different empirical and perhaps also theoretical approaches, they have been primarily discussed for experimental research, both in the field including randomized control trials and the laboratory, so we focus on these areas.

A pre-analysis plan is a credibly fixed plan of how a researcher will collect and analyze data, which is submitted before a project begins. Pre-analysis plans have been lauded in the popular press (for example, Chambers 2014; Nyhan 2014) and across the social sciences (for example, Humphreys, de la Sierra, and van der Windt 2013; Monogan 2013; Miguel et al. 2014). We will argue for tempering such enthusiasm for pre-analysis plans for three reasons. First, recent empirical literature suggests the behavioral problems that pre-analysis plans attenuate are not a pervasive problem in experimental economics. Second, pre-analysis plans have quite limited value in cases where more than one hypothesis is tested, piloted, or surveyed, and also where null results may not be reported. However, in very costly one-of-a-kind field experiments, including heroic efforts as the Oregon

■ *Lucas C. Coffman is Assistant Professor of Economics, Ohio State University, Columbus, Ohio. Muriel Niederle is Professor of Economics, Stanford University, Stanford, California. Niederle is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are coffman.155@osu.edu and niederle@stanford.edu.*

http://dx.doi.org/10.1257/jep.29.3.81 doi=10.1257/jep.29.3.81

health insurance study or Moving to Opportunity (Finkelstein et al. 2012; Katz, Kling, and Liebman 2001), they can be valuable. Third, pre-analysis plans may discourage the use of novel research designs and hence inhibit studies of robustness of previous findings.

Hypothesis registries are a database of all projects attempted. The immediate goal of this mechanism is to alleviate the "file drawer problem," which is that statistically significant results are more likely to be published, while other results are consigned to the researcher's "file drawer." This promising concept will not necessarily limit the number of times a hypothesis is tested, but instead simply give us a more accurate understanding of that number. One trade-off we foresee and discuss for registries is the benefit of eliciting precise, helpful descriptions of a project versus protecting researchers' intellectual property before it is published.

Finally, we evaluate the efficacy of replications. We argue that even with modest amounts of researcher bias—either replication attempts bent on proving or disproving the published work—or modest amounts of poor replication attempts—designs that are underpowered or orthogonal to the hypothesis—replications correct even the most inaccurate beliefs within three to five replications. We offer practical proposals for how to increase the incentives for researchers to carry out replications. We propose a journal of replication studies that accepts meaningful, well-designed replication attempts, failed or successful. In addition, we believe that other journals should enforce a norm of citing replications alongside the original result.

Pre-Analysis Plans

A pre-analysis plan requires researchers to register—in advance of carrying out the study—the hypotheses they plan to investigate and how they want to test their hypotheses. For empirical papers, the latter typically consists of a data collection protocol combined with a plan on how to analyze the data. A pre-analysis plan has at least three goals. First, pre-analysis plans limit the freedom of researchers concerning which hypothesis to investigate. A researcher will not be able to consider, say, ten different hypotheses using the same dataset and then publish a paper discussing only the one hypothesis that turned out to be statistically significant. Second, the researcher is restricted on how to test the hypothesis. The researcher cannot try many different specifications and focus only on the one with the control variables that provide the most satisfactory result. Third, the researcher often also precommits to a data collection plan. In particular, the researcher cannot stop collecting data only when a desired level of statistical significance has been reached. Hence, a pre-analysis plan reduces the ability of a researcher to cherry-pick hypotheses, data analyses, or a good dataset. The result is that a pre-analysis plan should increase the probability that a published positive result is true. Casey, Glennerster, and Miguel (2012) are typically credited for the first pre-analysis plan in economics, and they offer a fuller discussion of potential benefits.

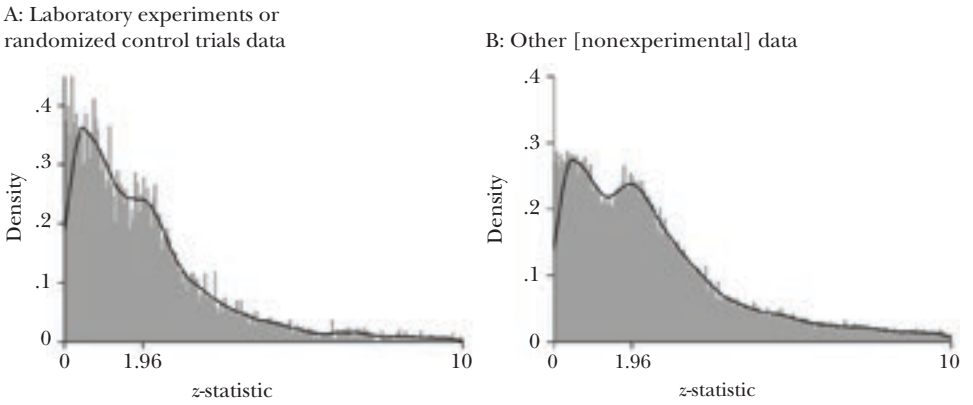
A Need for Pre-Analysis Plans?

Before discussing the pros and cons, we review the evidence on the need for pre-analysis plans. Their rise to prominence has, at the least, been facilitated by recent, troubling findings in other social sciences that suggest false positives may be more pervasive than implied by conventional levels of statistical significance. For example John, Loewenstein, and Prelec (2012) show evidence of the ubiquity of questionable research practices in psychology, while Simmons, Nelson, and Simonsohn (2011) show how these practices dramatically increase the incidence rate of false positives. Moreover, the questionable practices at the center of these papers are precisely the behaviors pre-analysis plans are meant to quash. Simonsohn, Nelson, and Simmons (2014) analyze the pattern of significant results to assess whether *p*-hacking (manipulating *p*-values) is a pervasive problem in psychology. Using papers published in the *Journal of Personality and Social Psychology*, a top psychology journal, their findings suggest that *p*-hacking is indeed pervasive for papers that report results *only* with a covariate, though not for other papers. Such research suggests there is a problem in some social sciences, and pre-analysis plans could help to provide a solution.

For evidence leaning in the other direction, Brodeur, Lé, Sangnier, and Zylberberg (forthcoming) provide the first analysis of whether *p*-hacking through such questionable research practices is a substantial problem in applied economics and whether this problem is indeed more pervasive in experimental economics, the area in which researchers control the data collection process. They analyze every *z*-statistic reported in the *American Economic Review, Journal of Political Economy,* or *Quarterly Journal of Economics* between 2005 and 2011. A *z*-statistic is a measure of how likely a result is due to chance rather than a true finding, where the higher the absolute value of the *z*-statistic, the lower the associated *p*-value. Figure 1 shows their figures with the distribution of *z*-statistics for all experimental work, including both laboratory and field studies, in the left panel, and all other empirical papers in the right panel.

In the absence of *p*-hacking, one would expect a perfectly smooth distribution of *z*-statistics, with perhaps one peak due to a threshold *z*-statistic for publication. In the presence of authors *p*-hacking to get *p*-values just below some desired thresholds, especially 0.05 (or a *z*-statistic just above 1.96), the distribution would have two peaks. This is because results that "just" fall short of a significance threshold are *p*-hacked to provide "nicer" results. This in turn generates "missing" *z*-statistics and hence a valley between the two peaks, as shown by the camel-shaped pattern in Figure 1 reproduced from Brodeur et al. (forthcoming). Visually, the distribution of experimentally produced *z*-statistics on the left-hand panel is single-peaked with a slight second bump, while the nonexperimental distribution has two sharp peaks. The analysis by the authors backs up the visual. With 122 papers in their dataset for experimental papers, they are not able conclude at a suitable level of statistical significance that this group of papers exhibits signs of *p*-hacking. Though pre-analysis plans could apply to other empirical work (in fact, Brodeur et al. find significant *p*-hacking on nonexperimental papers as shown in

Figure 1
Evidence of *p*-hacking



Source: Figures 6e and f from Brodeur, Lé, Sangnier, and Zylbergerg (forthcoming).
Notes: Displays distribution of *z*-statistics reported in all papers appearing in either the *American Economic Review*, *Journal of Political Economy*, or *Quarterly Journal of Economics* between 2005 and 2011. Experiments, both lab and field, are in the left panel; all other papers in the right panel.

the right panel in Figure 1), or even theoretical work, it is worth noting that the push for pre-analysis plans is happening precisely within experimental fields. For example, the Social Science Registry, run by the American Economic Association, explicitly states it is a "registry for randomized control trials."

There are at least two caveats to the null result found by Brodeur et al. (forthcoming). First, the dataset comes from the top three journals in economics. Perhaps *p*-hacking is more pervasive elsewhere. Second, experimental economists may have other tools at their disposal for producing false positives, just not the tools that are targeted by pre-analysis plans.

Benefits of Pre-Analysis Plans

How much does a pre-analysis plan increase the probability that a statistically significant result is indeed true? No data exist to address this question. However, we can obtain a theory-driven estimate for this question using the framework of Ioannidis (2005). Our goal is to compute the probability that a published, positive result is true, the "positive predictive value." Our estimate is built on five parameters.

The first parameter α is the statistical significance threshold for a positive result. Here, we will use $\alpha = 0.05$.

The second parameter is the "power" of the study. Say that β is the "Type II" error, which is the probability that a study will fail to detect an effect when an effect actually exists. A smaller β means a more powerful study. The power of the study is typically expressed as $1 - \beta$, so that a smaller β leads to a larger number. Here, we set $\beta = 0.2$ and so $1 - \beta = 0.8$.

The third parameter π is the proportion of studies that are testing true hypotheses (or the expected probability of a hypothesis being true). Rather than try to pin down this value, we experiment with a range of values: 0.3, 0.5, 0.7, and 0.9.

The fourth parameter is u, the study bias, which is the probability with which a study that would have been reported false without any bias is instead reported positive (for any reason). Practices that affect u can operate by a variety of mechanisms. For example, one approach is continuing to add more subjects to an experiment, or perhaps extending the sample, until a positive result is reached (Simmons, Nelson, and Simonsohn 2011). Another way to affect u is through channels having to do with how a given dataset is analyzed. For example, a researcher may have a lot of freedom in deciding which control variables to use in what combinations and can try these out until a positive result is achieved. One primary goal of a pre-analysis plan is that it would reduce u. For our illustrative calculations, we consider $u = 0.25$, 0.10, or 0.01. Though this is merely guesswork, perhaps a value of 0.10 can be thought of as corresponding to some restriction due to a pre-analysis plan, and 0.01 a very restrictive pre-analysis plan.

The final parameter is k, the number of substitute studies that were (or could be) investigated. To be precise, we assume that out of k possible investigations, only the first positive one is reported, and all others are either never investigated or simply never reported. We will explore values of $k = 1$, 10, and 25. One value of a pre-analysis plan is that it restricts the researchers' ability to consider several (perhaps not necessarily completely independent) hypotheses with the same data, and hence, within a given dataset, forces k to be one. (Of course, the value of such a restriction relies on the researcher not writing, say, ten pre-analysis plans for the same dataset with one hypothesis each.) There are, however, many ways in which k can be bigger than 1 other than the case of multiple hypotheses to be tested with a single dataset.

One way to have k bigger than one is pointed out by Ioannidis (2005): suppose multiple researchers investigate the same hypothesis, some of them do not get a statistically significant result, and the first one to do so is published (or written up), and the future researchers do not investigate the same question after the first positive result is published. There are, however, some projects for which numerous substitute studies are less likely. Large field experiments, like the Oregon health insurance experiment or Moving to Opportunity, may arguably be the only test of their respective hypotheses and may be for some time (Finkelstein et al. 2012; Katz, Kling, and Liebman 2001).

However, what constitutes "substitute studies" should be much broader than is commonly recognized. For example, a researcher could work on multiple distinct projects, each testing a different (though, for the sake of the argument) equally likely hypothesis. The time-limited researcher then decides to only write up the first project with a positive result and lets others languish and get filed away.

Another way in which k is potentially greater than one is if a researcher runs pilot studies to assess which hypothesis may be most likely to yield a statistically significant result. These pilot studies can be informal. Perhaps a researcher runs a

large-scale survey to understand what is driving a particular phenomenon, but only runs an experiment on the most promising outcome from the survey. Or perhaps the researcher could simply run thought experiments about different scenarios or experimental paradigms, and dismiss those that would not likely yield a positive result. For example, consider a field study or an experiment investigating a specific hypothesis. The researcher then has to find an environment, or a task, or a specific game in which to investigate the hypothesis. In making this choice, either with the aid of piloting or thought experimenting, the researcher has dismissed many other possible tests using different samples, environments, and tasks. The issues that arise in having pilot studies that are reported have received some attention in experimental economics (for example, Roth 1994), and pilots run inside the researcher's head run into similar problems. While ten such pre-tests (actual tests, pilots, or even thought about designs) are clearly not ten independent tests of the same hypothesis, it is also clear that they are not the same as just testing one hypothesis.

Using all those parameters that affect the probability that a published, positive result is true, we trivially extend the Ioannidis (2005) results to bring together u and k into the same equation and obtain:

$$\textit{Positive Predictive Value-} = \frac{[1 - \beta^k(1 - u)^k]\pi}{[1 - \beta^k(1 - u)^k]\pi + (1 - \pi)[1 - (1 - \alpha)^k(1 - u)^k]}.$$

To obtain an intuitive feeling for how this equation works, consider first the situation when the parameter u for the study bias is zero and k for the close substitute studies is equal to 1. Then the numerator reduces to $[1 - \beta]\pi$, which is the power of the study multiplied by the share of times π that a study is testing a true hypothesis. For example, if $\pi = .5$ and the power of the study is .8, then the study will confirm the true result 40 percent of the time. Adding back the parameters u and k means that β, the measure of "Type II" error, is now multiplied by a $(1 - u)$ term, capturing how study bias can diminish the probability that a positive finding is indeed a true result. The k term in the exponents means that as the number of substitute studies rises, a false hypothesis has to come out negative for *every* test of that hypothesis for no false publications to arise. (This is why these terms are raised to the power of k.)

Now consider the denominator of the equation in this same case, where the parameter u for the study bias is zero and k for the close substitute studies is equal to 1. With this simplification, the denominator becomes $[1 - \beta]\pi + (1 - \pi)[1 - (1 - \alpha)]$. The first term, as in the numerator, shows the power of the study multiplied by the chance π that the studies are testing true hypotheses. In the second part of this expression, $1 - \pi$ is the proportion of studies that are not testing true hypotheses, and the rest of this part of the expression simplifies to the statistical significance parameter α, which is of course the chance that even though a hypothesis is not true, it is accepted anyway. Again, adding back the parameter u, gives us how study bias can diminish the meaningfulness of a positive result, while adding the k term in the exponents means that bias and the particular level of statistical significance becomes exponentially less important as the number of studies

Table 1

How Reducing Within-Study Bias Affects Probability that a Published Positive Result Is True (PPV), by Number of Substitute Studies and Expected Probability That a Hypothesis Is True

Number of substitute studies:		*1 study*		*10 studies*		*25 studies*	
Expected probability of true hypothesis	***Bias***	*PPV*	*ΔPPV (from row above)*	*PPV*	*ΔPPV (from row above)*	*PPV*	*ΔPPV (from row above)*
0.30	**0.25**	0.56	–	0.31	–	0.30	–
	0.10	0.71	0.15	0.35	0.04	0.30	0.00
	0.01	0.86	0.14	0.52	0.17	0.37	0.07
0.50	**0.25**	0.75	–	0.51	–	0.50	–
	0.10	0.85	0.10	0.56	0.05	0.50	0.00
	0.01	0.93	0.08	0.71	0.16	0.58	0.08
0.70	**0.25**	0.87	–	0.71	–	0.70	–
	0.10	0.93	0.06	0.75	0.04	0.70	0.00
	0.01	0.97	0.04	0.85	0.11	0.76	0.06
0.90	**0.25**	0.96	–	0.90	–	0.90	–
	0.10	0.98	0.02	0.92	0.02	0.90	0.00
	0.01	0.99	0.01	0.96	0.04	0.93	0.03

Note: A significance level of 0.05 and power of 0.8 is used throughout; "PPV" refers to the "positive predictive value" as in Ioannidis (2005), which is the probability of a result being true given a positive result.

increases because every test of the hypothesis would have to come out "wrong" for the false positive to remain.

Table 1 uses the formula to compute positive predictive values given the parameters above. We compute the change in the probability that the positive result is correct as we reduce the research bias for any given k, the number of substitute studies.

The results in Table 1 make clear that a pre-analysis plan that reduces the chance a researcher can "generate" a false positive from 25 to 10 percent is most effective when $k = 1$ (there are no, and never will be any, substitute studies) and the prior for the hypothesis to be correct is low. In cases where there are, or ever will be, substitute studies, pre-analysis plans are most helpful when they are very restrictive—that is, the bias is reduced almost to zero. In those cases, the reduction from 10 to 1 percent is the most important in affecting the posterior that a hypothesis is actually true after a positive paper. For hypotheses that will be tested many times because a large number of substitute studies k are possible, reducing the bias variable u has relatively little effect, unless that reduction is nearing a full elimination of bias.

The results suggest that if a paper is going to be the only attempt of a hypothesis, which might be true of many large and expensive field experiments, employing

a pre-analysis plan to reduce bias can be a very fruitful endeavor. However, if the hypothesis being tested is in a lower-cost environment where we might expect several tests, the gains from utilizing a pre-analysis plan are small enough that the potential costs are worth more consideration.

Finally, it is worth considering the absolute levels of positive predictive value (PPV) throughout the table, because other than projects testing a hypothesis with a high expected probability of being true or when the hypothesis will only be tested once ever, the absolute levels of positive predictive value (the probability of a result being true given a positive result) are disturbingly low. Even if the pre-analysis plan is so restrictive that the chance a researcher can bias the results is basically eliminated (with a value of 0.01), the increase in the posterior probability that a hypothesis is true after a positive result is disappointingly small. When there is no competition for the result, a prior of 0.30 would be updated all the way to 0.86 if a paper found a positive result and there was a very restrictive pre-analysis plan. However, if there are ten "substitute studies," the posterior after a positive result is only 0.52 even with a very restrictive pre-analysis plan. When there are 25 "substitute studies," this number drops to 0.37. When lots of substitute studies are available, and only the first one to find a statistically significant result is published, such a finding does not increase the positive predictive value to acceptable levels.

Costs of Pre-Analysis Plans

A common criticism of pre-analysis plans is that they inhibit exploratory work (for example, Gelman 2013). Without the autonomy to reoptimize research after it has begun, working in areas with many unknowns becomes a risky endeavor. A researcher carrying out a field experiment, for example, often is not armed with confident priors about which projects will be successful, which treatments to run within a project, what analysis will be most appropriate, what subpopulation will most respond to the treatment, and so on. When an economist obtains a new, rich dataset, we do not want to handcuff the analysis to a specific question. We want the researcher to report, with appropriate caveats, all that can be learned from the data. This is why we typically give researchers the freedom to pursue the most interesting follow-up, performing the analysis that best fits the patterns of the data as they emerge.

However, we also know that allowing empirical or fieldwork such degrees of freedom can produce high false positive incidence rates (as in Simmons, Nelson, and Simonsohn 2011). We can combat this, while allowing leeway to investigators while the research is in progress, in two ways. First, we can allow the researcher to offer reasons in defense of the reasonableness of, say, add-on treatments, language changes, or a unique method for analyzing the data. Audience members, anonymous referees, and readers can determine if these add-ons seem reasonable. Second, we can use robustness tests, in which important and/or surprising results should be replicated with a variety of modest alterations whenever possible. Although pre-analysis plans may help reduce the proportion of results that are false positives, pre-analysis plans do not help us learn about the robustness of results.

Miguel et al. (2014) rightly point out that pre-analysis plans can encourage exploratory work by lending credibility to surprising findings. A researcher who has set the hypothesis in stone ahead of time cannot be accused of making up that hypothesis only after the statistical analysis was done. Likewise, if a researcher plans to use statistical techniques that might be viewed as suspect data mining (for example, by analyzing subgroups or removing certain outliers), the researcher can pre-register those plans and avoid distrust. However, in these cases the investigator has a clear sense of direction and methods. But as noted above, doing research in new areas often does not come with this luxury.

On the other side, the rigidity of pre-analysis plans may also motivate researchers to know more about their design before they start. For example, this may increase the rate by which researchers pre-test their designs, or it may also increase the temptation to use only very minor deviations from existing designs. Results from known designs will be less surprising on average, lending themselves more readily to a pre-committed analysis plan, but also reducing what we learn about the context-specificity of the original result. Finally, the costs for exploratory work may be increased relative to somewhat more derivative work as a researcher may be reluctant to head into uncharted territory if the researcher has to commit to a rigid pre-analysis plan beforehand.

Hypothesis Registries

When a hypothesis is registered, it does not necessarily lay out, or commit to, any specifics regarding data collection or method of analysis (though these can be included). Here, we consider a hypothesis registry simply to be a publicly available database of well-defined hypotheses submitted before any attempt at data collection or analysis was made. This mechanism is rightfully gaining steam in economics. The American Economic Association runs a hypothesis registry website for randomized controlled trials with well over 300 studies registered at http://socialscienceregistry.org. This approach seems relatively popular among development economists. In addition to the AEA registry, many organizations enabling research in developing countries have similar registries, including the Jameel Poverty Action Lab at http://www.povertyactionlab.org/Hypothesis-Registry and the International Initiative for Impact Evaluation (3IE) registry at http://www.3ieimpact.org/en/evaluation/ridie/ for the 3IE registry. Also, outside of the social sciences, for some time now and preceding most of these social science registries, all US clinical trials have had to be pre-registered.

The Need for Hypothesis Registries

Most prominently, hypothesis registries will help eliminate the file-drawer problem, in which null results are more likely to remain unpublished. Empirical studies on the extent of the file-drawer problem have been difficult, though Franco, Malhotra, and Simonovits (2014) recently studied experiments on Time-Sharing

Experiments in the Social Sciences (TESS). According to the TESS website (http://www.tessexperiments.org/, accessed June 26, 2015); "Investigators submit proposals for experiments, and TESS fields successful proposals for **free** on a representative sample of adults in the United States . . . a highly-respected Internet survey platform." To run an experiment on TESS, researchers apply with a proposal, which is then peer-reviewed. The 249 studies that were conducted on TESS between 2002 and 2012 provide a unique sample of studies whose pre-data collection design is publicly available. We have access to the file drawer for TESS studies.

Franco et al. (2014) show that strong results have a 60 percentage point higher likelihood of being written up, and about a 40 percentage point higher chance to be published than null results. Given that the sample is somewhat unique in that it consists of vetted studies, the results suggest that the file-drawer problem may be quite substantial.

Benefits of Hypothesis Registries

Hypothesis registries provide data on the number of previous attempts at establishing a certain hypothesis, even those that ended up as null results and were not published. In this way, they offer a better sense of the lower bound on the number of substitute studies for a given hypothesis. Though a registry would not directly decrease the number of substitute studies, it would give us a better sense of the number of substitute studies run for a given class of hypotheses. Hence, the registry would not necessarily increase the probability that a published result is true, but it would give us a better idea of what that probability is.

Additionally, in equilibrium, the registries could reduce the number of substitute studies run. For example, if having a high registered-hypotheses-to-published-results ratio becomes a negative mark on a researcher's resume, researchers may take measures to ensure higher power when designing a study.

Costs of Hypothesis Registries

Hypothesis registries are a useful idea that seems likely to spread. Here we list a few possible downsides, which should help to clarify how information from registries should be consumed and perhaps also to shape the design of such registries.

First, many researchers would not feel comfortable sharing the details of their hypothesis and design before they have published their work. Though this may be less of a concern for projects with higher fixed costs, such as experimental fieldwork, the concern becomes more acute for lower-cost, quicker-turnaround work. Consequently, lest we encourage vague, unhelpful (and hence unstealable) registered hypotheses, each registered item would need a predetermined privacy period before it was made public. If we were to afford the authors a time period within which they have a fair chance to publish their work, this period will be measured in years, perhaps even five years or more. As a result, it seems that in designing hypothesis registries, we must choose between knowing what is in the file drawer only with a substantial lag, or ending up with a registry that is frustratingly vague.

Second, if a listing in a hypothesis registry does not result in a published paper, it would not be clear why. In some cases, perhaps, the research budget ran out or the researcher turned to other topics, and so the paper was never written. Even if we managed to require the researcher to report results back to the registry, it would not be easy to infer why the paper was rejected for publication. Maybe the setup was simply a poor test of the hypothesis. Perhaps the project did not obtain a statistically significant result, and journal referees viewed it as not worth publishing. A lack of publication of a registered hypothesis does not reveal whether the hypothesis was rejected, or poorly tested, or some mixture of the two. (A similar issue arises with pre-analysis plans, when no published paper later results.)

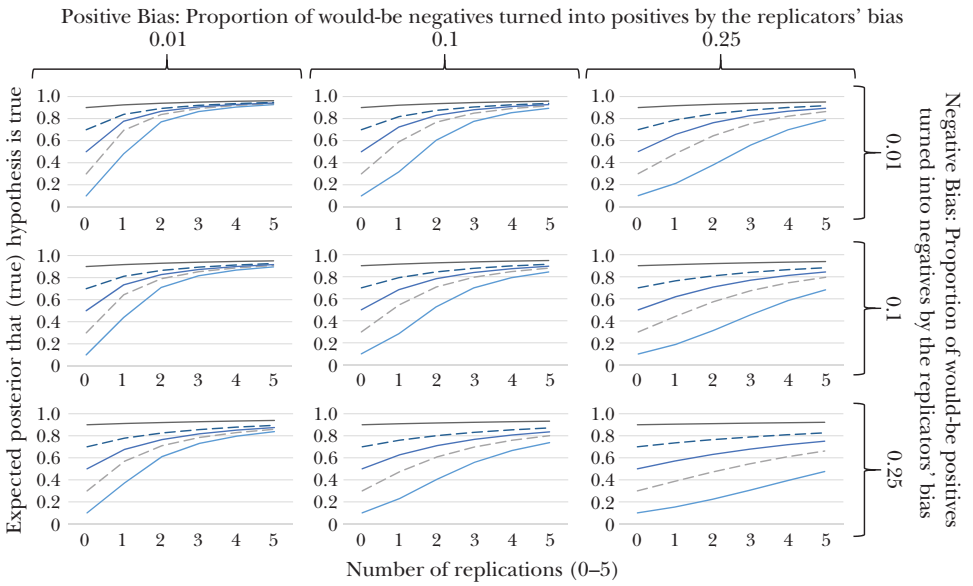
Third, the hypotheses in the registry would not necessarily be organized in a helpful way, and, as with Google Scholar and other literature search tools now, navigating the registry for work related to a specific hypothesis would not be straightforward. Different fields use different keywords. Some entries might be vague. Some might be in their privacy period. This problem is in contrast to replications, discussed in the next section, where a natural self-organization exists: once you knew the original work, describing many subsequent tests and how they relate to variations in data or statistical specification would be straightforward.

These drawbacks suggest that hypothesis registries will need to think seriously and evolve useful rules and standards in several areas: privacy periods; a required level of specificity; and figuring out a flexible and serviceable organizational mechanism.

Replications

The power of replications in a series of studies is perhaps best illuminated by the ultimatum game literature, started by Güth, Schmittberger, and Schwarze (1982). An ultimatum game has two players: a Proposer and a Responder. The experimenter provides a stake. The Proposer suggests how the stake should be split. If the Responder accepts the proposal, then both players receive what the Proposer suggested. If the Responder does not accept the division, both players receive zero. The straightforward game theory prediction is that a logical Proposer will offer the Responder the smallest possible slice of the overall stake, and the Responder will accept—because the alternative is to receive nothing at all. However, Güth, Schmittberger, and Schwarze find that Proposers ask for much less than nearly all of the stake, and that Responders reject many offers, preferring to receive zero rather than what they view as an unfair offer. Many follow-up studies have tested these results in various environments and cultures, and as a result, the original results have been replicated hundreds or even thousands of times, testing both whether the original results were a chance draw and how robust the results are to contextual changes. We know with considerable confidence that ultimatum game offers are indeed robustly closer to half of the stake than to zero, and that many offers of positive amounts are rejected. Subsequent work has shown conditions which may lead to a larger

Figure 2
Expected Posterior of True Hypothesis after *n* Replications, by Different Researcher Biases



Source: Authors.
Notes: All nine figures report expected beliefs in a hypothesis after a given number of replications, taking prior belief as given. Calculations assume power of 0.8, false positive rate 0.05 (for zero researcher bias), and all hypotheses are true.

acceptance of lower offers (for example, larger stakes), the importance of fairness beyond ultimatum games, as well as some conditions necessary for fairness motives to play a large role. For surveys of this literature, see Roth (1995) on bargaining and Cooper and Kagel (forthcoming) on fairness and other-regarding preferences.

One way to evaluate the upside of replications is to consider how speedily beliefs converge to the truth. Suppose a study finds a statistically significant result, and further suppose that the hypothesis is actually true. How much more confident do we become that the result is correct after one replication? Five replications? How does this conclusion depend on our prior beliefs in the hypothesis and upon how the replication attempts are carried out?

Figure 2 shows how beliefs are expected to converge to the truth for a true hypothesis. Each line takes a given prior as its starting point, shown on the vertical axis. One could consider this starting point to be the probability a published positive result is true based on calculations like those from Table 1. However, because priors are an open-ended term, the priors on the vertical axis could also represent beliefs in a hypothesis at any point in time, after several papers or replications. Before accounting for researcher bias, the figure uses the same standard estimates for statistical power (0.8) and level of significance ($\alpha = 0.05$) used earlier.

The top left graph in Figure 2 shows how quickly this convergence happens for almost unbiased replications. Each line takes as given the prior belief that the hypothesis is true and subsequently tracks how beliefs increase in expectation with each given replication. Even for dramatically low prior beliefs, posteriors increase rapidly. A prior belief of only 0.30 that the hypothesis is true (equal to the lowest probability a published positive result is true in Table 1) is corrected upwards to 0.84 after only two replications and to 0.89 after three. In this case, most of the convergence typically happens within two or three replications and the value of additional replications (under these assumptions) is much smaller thereafter.

However, there are at least two reasons for concern that the replications themselves will not be unbiased. First, researchers may be motivated (for a variety of noble and ignoble reasons) to prove or disprove a published result, and thus such motivations can artificially increase the rate of the desired outcome. Second, a failure to replicate a result can arise out of a poor test of the original hypothesis. For example, perhaps the follow-up experiment may be underpowered, or it may have a design somewhat orthogonal to the original hypothesis. In either case, a negative outcome is hardly dispositive of the veracity of the published result. What constitutes a fair replication of the original result is a question worthy of its own literature (as a starting point on this issue, see Brandt et al. 2014; Coffman and Niederle in preparation). We will focus on how poor replications may diminish the beneficial effect of replications on belief-updating.

Here, we model the bias operating in replication studies as a proportion of positive (negative) results being flipped to negative (positive), compared to if the experimental replications had been run well, honestly, and so on. Incidences of poorly run experiments, either underpowered or orthogonal, are modeled as the results being reversed from positive to negative. Figure 2 illustrates how bias in replications affect the informational value of replications. Going from left to right, Figure 2 increases the proportion of would-be negative to positive outcomes ("positive bias") from 0.01 to 0.10 to 0.25 (and increases "negative bias" going from top to bottom). As one would expect, adding such biases decreases the signal-to-noise ratio of a replication, and posterior beliefs that the hypothesis is true converge to the truth less quickly.

Without making any claims about what bias rates the replications have or should have, these kinds of calculations suggest two clear takeaways. First, for modest bias rates (say, 10 percent and below), we can expect posteriors not too distant from the truth after three to five replications. Second, the usefulness of replications is greater if their bias is modest, and for this, pre-analysis plans can be a highly useful tool. If one-quarter of positive results that are true are reversed as in the bottom graph, it may be some replications are not more valuable than their costs. However, if pre-analysis plans can help to minimize these biases, even if just to 10 percent, it would seem that replications can be a valuable tool. Moreover, a main potential downside of pre-analysis plans—that is, inhibiting discovery—is a non-issue with replications. When replicating, there are fewer unknowns about the design and the results, so the researcher needs less flexibility. Even though pre-analysis plans may not be appropriate for all work, they may prove invaluable for replication studies.

Of course, in thinking about the value of replications, the financial costs of replications will vary widely. Replication for a nonexperimental economic study—say, using different data to test a certain hypothesis—has a relatively low cost. A typical experimental economics study in a laboratory context can cost about $5,000 in subject payments and be done in a few months. A randomized control trial in a developing country can cost 20 times that in staff salaries alone and require several years to complete. However, the total cost of replications for a specific project, at least, are somewhat known. The cost for each replication can be inferred from the initial project, and Figure 2 suggests roughly three to five replications need to be done. These cost estimates can be judged relative to the importance of the result. Of course, the cost estimates given here do not include the opportunity cost of the time of researchers involved in replication studies and whether researchers perceive replication as a worthwhile use of their time.

A Proposal for Incentivizing Replications

At present, replications are relatively scarce, which suggests that researchers have little incentive to replicate previous studies. Here, we present a modest proposal as a first step towards thinking about how to motivate more replications. The incentive for replications would be built on two of the currencies of our industry: publications and citations. We hope to promote both what can be called "exact replications"—that assess whether the initial result is likely to be true, or whether the initial study was a chance draw of the data—and also work that considers variations of the initial design or mode of inquiry to understand the robustness of results. Our proposal has two components: 1) an outlet for replication studies; for now, we will refer to this as the *Journal of Replication Studies*; coupled with 2) a plea to referees of other journals to require citations of replications alongside the citation of the original paper.

A *Journal of Replication Studies* has three purposes. First, such a journal would offer an outlet for publication to meaningful, well-designed, and well-run replications. Though many journals accept replication attempts, authors often (and probably correctly) fear that the odds of publication are substantially lower for nonoriginal work, leading to replications never being produced in the first place. A dedicated journal would alleviate these concerns by agreeing to judge a submission based on whether it was a good replication, regardless of the findings and degree of originality.

Second, the journal could perhaps signal what articles are higher priority for replication attempts. One could imagine that the editorial board, or the board of specific organizations (maybe the Economic Science Association for experimental economics, Bureau for Research and Economic Analysis of Development for development economics, and so on) could publish a list of papers for which such a replication exercise would be more likely to result in a publishable paper. On the one side, deciding on such a list might be politically difficult. On the other side, targeting replications to industry agreed-upon published results, rather than

towards personal disagreements or even witch hunts, could help to increase the value and visibility of replications. Also, having a list of papers that are high priority for replication can provide a greater incentive for more replications.

Third, the *Journal of Replication Studies* could also collect replications (failed or not) that exist within other original papers.[1] Suppose a researcher writes a paper that builds on an important result. In doing so, the researcher also replicates the original study and ultimately publishes the paper in a different journal. It could be valuable for a *Journal of Replication Studies* to publish a shorter paper, almost an extended abstract, describing the results of the replication and referring to the longer version of the paper. In this way, a dedicated journal could become a one-stop shop for a record of replications at least within a certain field, whether the replications were failed or successful.

While we are aware that most researchers will not receive tenure based on papers published in the *Journal of Replication Studies*, it is also the case that many universities judge a tenure case not only based on the best three to five papers, but also on the number of publications. It may well be that a couple of publications in the *Journal of Replication Studies* provide a useful "surrounding cast" to the "main portfolio" to push a candidate over the tenure bar. It could also be a great exercise for, say, third-year graduate students to attempt a replication of recent important work.

The second part of our proposal seeks to ensure that replications will be cited, and hence increase the visibility of researchers willing to replicate papers. To do this, a norm must be enforced *at other journals*: if a submission cites an original paper that has replication attempts, the author agrees also to cite replications that appeared in the *Journal of Replication Studies*, and the editors and referees agree to enforce this norm. While we understand that journal space is expensive, and not all journals will feel that they can justify publishing a series of replications of earlier papers in their own pages, journal space is not so expensive as to rule out adding a citation noting that a study was replicated by *X* or failed to be replicated by *Y*. We can imagine various conventions that might arise in reference lists of journal articles, where the citation to an empirical article might be followed by NR for "not replicated," and citations to a replication study might be followed by R+ for "replicated and confirmed" or R− for "replicated and did not confirm." Such citations will properly strengthen (or weaken) the citation made.

We readily acknowledge that the details of this proposal could use some additional consideration. But the overall goal here is worth remembering: with greater professional incentives for replication, economists can properly test, and re-test, our most important and influential findings, which should over time leave us with greater confidence in the veracity of the results.

[1] We thank Katherine Coffman for the suggestion.

Conclusion

In this paper, we discussed the costs and benefits of different institutions for increasing our ability to estimate the likelihood that empirical results are true. We paid particular attention to pre-analysis plans and replication attempts.

Contrary to popular belief, pre-analysis plans do not always offer dramatic decreases in the false positive rate. They seem to be most effective in reducing bias for work where there are few other substitute studies—expensive fieldwork is a likely candidate—and when pre-analysis plans are very restrictive, effectively reducing researcher biases close to zero. We conclude that if pre-analysis plans have a downside, like inhibiting exploratory work, or placing a greater burden on young and less-experienced researchers, the results suggest pre-analysis plans should be limited to costly, one-time studies. However, pre-analysis plans are likely a great tool for replication studies: in replication studies, there is no risk of deterring creative work, and reducing researcher bias in replications greatly increases their informational value. When possible, replications can not only sniff out false positives but also provide data on the robustness of results to their contexts. Improving the professional incentives of researchers to carry out replications should be a priority.

We therefore hope that as a profession we move towards valuing replications and robustness checks of positive results. We think that false positives are basically unavoidable in a young field like economics, where researchers may investigate quite different hypotheses from one another. If a result is deemed important, it should be important enough to warrant some replications that can elevate, to meaningful levels, the posterior that the hypothesis is actually true.

References

Brandt, Mark J., Hans Ijzerman, Ap Dijksterhuis, Frank J. Farach, Jason Geller, Roger Giner-Sorolla, James A. Grange, Marco Perugini, Jeffrey R. Spies, Anna van 't Veer. 2014. "The Replication Recipe: What Makes for a Convincing Replication?" *Journal of Experimental Social Psychology* 50: 217–24.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. Forthcoming. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics*.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *Quarterly Journal of Economics* 127(4): 1755–1812.

Chambers, Chris. 2014. "Psychology's 'Registration Revolution.'" *The Guardian*, May 20.

Coffman, Lucas C, and Muriel Niederle. In preparation. "Exact and Robust Replications: A Proposal for Replications."

Cooper, David, and John H. Kagel. Forthcoming.

"Other Regarding Preferences: A Selective Survey of Experimental Results." *Handbook of Experimental Economics* vol. 2, edited by John H. Kagel and Alvin E. Roth.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127(3): 1057–1106.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science*, September 19, 345(6203): 1502–05.

Gelman, Andrew. 2013. "Preregistration of Studies and Mock Reports." *Political Analysis* 21(1): 40–41.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior & Organization* 3(4): 367–88.

Humphreys, Marcatan, Raul Sanchez de la Sierra, and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21(1): 1–20.

Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Med* 2(8): e124. doi:10.1371/journal.pmed.0020124.

John, Leslie K., George Loewenstein, and Drazen Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science* 23(5): 524–32.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics* 116(2): 607–54.

McAleer, Michael, Adrian R. Pagan, and Paul A. Volker. 1985. "What Will Take the Con Out of Econometrics?" *American Economic Review* 75(3): 293–307.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, and M. Van der Laan. 2014. "Promoting Transparency in Social Science Research." *Science*, January 3, 343(6166): 30–31.

Monogan, James E. 2013. "A Case for Registering Studies of Political Outcomes: An Application in the 2010 House Elections." *Political Analysis* 21(1): 21–37.

Nyhan, Brendan. 2014. "To Get More out of Science, Show the Rejected Research." *New York Times*, September 18.

Roth, Alvin E. 1994. "Lets Keep the Con Out of Experimental Econ.: A Methodological Note." *Empirical Economics* 19(2): 279–89.

Roth, Alvin E. 1995. "Bargaining Experiments." *The Handbook of Experimental Economics*, edited by John H. Kagel, and Alvin E. Roth, pp. 253–48. Princeton University Press.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22(11): 1359–66.

Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons. 2014. "*P*-Curve: A Key to the File-Drawer." *Journal of Experimental Psychology: General* 143(2): 534–47.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 99–120

Law, Regulation, and the Business Climate: The Nature and Influence of the World Bank Doing Business Project[†]

Timothy Besley

The importance of a well-functioning legal and regulatory system in creating an effective market economy is now widely accepted. After all, a poor contracting and regulatory environment can raise the cost of doing business with knock-on effects to employment, output, investment, productivity, and living standards. But how to measure or even to conceptualize differences in the business climate is far from settled. Should the focus be on a few specific indicators or many? Can whatever indicators are chosen be usefully compared across time and countries? Can such data be updated in a timely way as policy reforms occur?

One flagship project that tries to measure the environment in which businesses operate in countries across the world is the World Bank's Doing Business project, which was launched in 2002. At its core, this project gathers quantitative data to compare regulations faced by small and medium-size enterprises across economies and over time. The centerpiece of the project is the annual Doing Business report. It was first published in 2003 with five sets of indicators for 133 economies, and currently includes 11 sets of indicators for 189 economies. The report includes a table that ranks each country in the world according to its scores across the indicators.

The Doing Business project has become a major resource for academics, journalists, and policymakers. The project also enjoys a high public profile with close to ten million hits on its website each year, making it one of the most prominent knowledge products produced by the World Bank. When Narendra Modi was elected Prime Minister of India, he explicitly targeted achieving 50th place in the ranking as

■ *Timothy Besley is School Professor of Economics and Political Science, London School of Economics, London, United Kingdom and W. Arthur Lewis Professor of Development Economics. His email address is t.besley@lse.ac.uk.*

[†] To access the Appendix, Data Appendix, and disclosure statement, visit http://dx.doi.org/10.1257/jep.29.3.99

doi=10.1257/jep.29.3.99

a benchmark for his administration—which would mean an improvement of almost 100 places compared to India's recent rankings (for example, Buerkle 2015). In 2012 Russian President Vladimir Putin set the goal of improving its Doing Business ranking to twentieth by 2018 (as reported in Adelaja 2012). Many countries are keen to promote their achievements in moving up the rankings in trying to attract investors, which is acknowledged in government export promotion strategies. For example, the UK government mentions Peru's ranking of 43 on "ease of doing business" prominently in its assessment of its business climate (UK Trade & Investment 2014). The project has passed from being a data source and research tool to playing a role in the political economy of development policy.

Leading academic economists have been involved in the Doing Business report from the start, both on the design of indicators and in using the results in research, and so in general terms, the report reflects the broad direction of mainstream thinking in development economics. Thus, it is now common to talk about the institutional underpinnings of development and the quality of the state in supporting development (for example, Besley and Persson 2011; Acemoglu and Robinson 2012). The Doing Business report collects data for a terrain over which there had been only scant knowledge previously, and so it is no surprise that academic researchers and policy analysts have taken the data to heart. Since 2003, over 2,000 research articles have been published in peer-reviewed academic journals using this data, with more than 5,000 working papers being posted online.

With such interest, it's no surprise that the Doing Business report has come under intense scrutiny. In 2012, following discussions by its board, the World Bank commissioned an independent review to evaluate the project (see http://www.dbrpanel.org/), on which I served as a member. In broad terms, the Doing Business report has been criticized for the way in which the data are collected and whether they reflect the business and regulatory environment accurately. Concerns were raised about whether the construction of the survey fostered a "deregulation bias." A measure of labor market regulation was a particular focus of concern, although this measure had already been removed from the set of measures used to determine aggregate rankings. Particular attention has focused on whether it is valid to collect the separate rankings into an aggregate ranking. A number of countries objected to being ranked at all.

Of course, alongside the specifics of the Doing Business report data, there are the usual concerns about the use of data that permeate empirical research in economics. Some researchers have used the data as a right-hand-side variable to "explain" outcomes of interest. Others put the data on the left-hand-side and ask how politics and institutions influence the business climate. In all cases, the usual concerns apply as to what inferences about causality can be drawn from such exercises. There is also a concern about the mapping from the Doing Business indicators to the conceptual categories that economic theory suggests ought to be important. Because so many researchers appear to equate "empirical evidence" with interpreting regression coefficients, this point merits discussion.

In the next section of this paper, I will describe how the Doing Business project works and illustrate its modus operandi with some of the key findings of the 2015

report. I address what is valuable about the project before turning to the criticisms of it. I then discuss some wider political economy issues illustrated by the report.

The main message of the paper is that, even with all of its imperfections, data collection of the kind undertaken by the Doing Business project is an integral part of the political economy of policymaking. The indicators try to get at features of the policy climate that many economists have been arguing are vital to economic progress but where no internationally comparable data was previously available. The story of the Doing Business project is one where a particular worldview can become influential and the impact of economic ideas enhanced through the collection of data.

How the Doing Business Project Works

It is useful to begin with a broad understanding of how the Doing Business project works. The data collection surveys law firms, with around 10,000 questionnaires being fielded across the participating countries. Data are collected in a questionnaire concerning 11 specific topics:

1) Starting a Business is a measure of the procedures, time, cost, and minimum capital required to start a new business.
2) Dealing with Construction Permits is a measure of the procedures, time, and cost required to build a warehouse.
3) Getting Electricity is a measure of the procedures, time, and cost required for a business to obtain a permanent electricity connection for a newly constructed warehouse.
4) Registering Property is a measure of the procedures, time, and cost required to register commercial real estate.
5) Getting Credit assesses the strength of the Legal Rights index, which measures the degree to which collateral and bankruptcy laws protect the rights of borrowers and lenders, and the depth of the Credit Information index, which measures the sharing of credit information.
6) Protecting Investors measures the extent of disclosure and director liability, and the ease of shareholder lawsuits.
7) *Paying Taxes* measures the number of taxes paid, hours per year spent preparing tax returns and the total tax payable as a share of gross profit.
8) Trading Across Borders is a measure of the number of documents, cost, and time required to export and import goods.
9) Enforcing Contracts is a measure of the procedures, time, and cost required to enforce a debt contract.
10) Resolving Insolvency is a measure of the time, cost, and percentage recovery rate involved with bankruptcy proceedings.
11) Employing Workers is a measure of the ease with which workers can be hired or made redundant and the rigidity of working hours, although this index is no longer used in the aggregate rankings.

As far as possible, the data collection is based on a reading of the laws and regulatory provisions. However, some assessment is inevitably subjective and reflects custom and practice in implementing the law. For example, the number of processes involved in starting a business can be interpreted by looking at actual laws. However, a question about the time that it takes to start a business is not asking about a schedule laid down in a specific statute, and can be regarded as a de facto measure of what commonly happens.

The indicators are intended to be comparable across countries, which is facilitated by basing the data collection on a precisely defined hypothetical enterprise and the circumstances that it faces. The central case is a firm with at least 60 employees, which is located in the country's largest business city. It is a private, limited-liability company and does not operate in an export-processing zone or an industrial estate with special export or import privileges. It is 100 percent domestically owned, and exports constitute more than 10 percent of its sales. While this detailed hypothetical makes comparisons feasible, the importance of limited-liability companies varies a lot from country-to-country, and there is an issue of how relevant the indicators are across all sectors of the economy and all types of enterprises. For example, the findings would not apply automatically to the agricultural sector, which is a large part of the economy in many low-income countries. (In fact, the Doing Business team has now started a separate project on agriculture.)

After receiving the survey responses, an "in house" team at the World Bank cross-checks the responses against the relevant laws and regulations. After internal consultation, an index score is created on each dimension. The result is the "raw" data, which are available on the website in disaggregated form. It is a fair critique that the way in which the questionnaires are processed to get a final score is not altogether transparent. But in this way, the Doing Business project is not unusual. It is not clear how to report information from this kind of survey in a meaningful way, and one can argue that the honesty of the responses in certain countries might suffer if the original data were widely available.

For each of the 11 dimensions in the data, an aggregate score is created by taking a simple unweighted average of the ranks of the underlying indicators, which leads to a cross-country ranking within each of the 11 topics. To obtain an overall Doing Business aggregate ranking, the report calculates a percentile for each country for ten of the topics (the Employing Workers category is excluded). These percentiles are aggregated to obtain the Ease of Doing Business ranking. These are the headline rankings that receive so much attention in media coverage.

The Doing Business report now also measures the distance from the frontier to gauge how far countries are from best practice. The benchmark for this exercise is the best performance observed on each Doing Business topic across all economies and years since 2005. The score lies on a scale between 0 and 100. A perfect score of 100 would require that the economy is on the frontier in every one of the 10 dimensions that go into the ranking. A 75, for example, implies that an economy is 25 percentage points away from the frontier.

Table 1

Distance to the Frontier Score from the 2015 Doing Business Report

(100 minus the distance to the frontier score gives the percentage points away from the frontier; higher is better)

Top 10	*Country*	*Distance to Frontier score*	*Bottom 10*	*Country*	*Distance to Frontier score*
1	Singapore	88.27	180	Haiti	42.18
2	New Zealand	86.91	181	Angola	41.85
3	Hong Kong	84.97	182	Venezuela	41.41
4	Denmark	84.20	183	Afghanistan	41.16
5	South Korea	83.40	184	Democratic Republic of Congo	40.60
6	Norway	82.40	185	Chad	37.25
7	United States	81.98	186	South Sudan	35.72
8	United Kingdom	80.96	187	Central African Republic	34.47
9	Finland	80.33	188	Libya	33.35
10	Australia	80.60	189	Eritrea	33.16

Source: World Bank, *Doing Business 2015* report available at http://www.doingbusiness.org/reports/global-reports/doing-business-2015.

Notes: The distance to the frontier score is calculated relative to the best performance observed on each Doing Business topic across all economies and years since 2005. A perfect score of 100 would require that the economy is on the frontier in every one of the 10 dimensions that go into the ranking. Zero represents the lowest performance. 100 minus the distance to the frontier score give the percentage points away from the frontier. A 75, for example, implies that an economy is 25 percentage points away from the frontier. For more information, see http://www.doingbusiness.org/data/distance-to-frontier.

The aggregate output from the *Doing Business 2015* report is illustrated in Table 1, which lists the top ten and bottom ten countries along with the distance to frontier scores. There are no big surprises. Three countries from Scandinavia are in the top ten. A number of countries from sub-Saharan Africa are in the bottom ten, along with Haiti, Libya, and Afghanistan. The overall rank and distance from the frontier score of the BRIC countries are: Brazil (120, 58.01), Russia (62, 66.66), India (142, 53.97), and China (90, 62.58).

Each Doing Business report documents frequent reforms. For example, between June 2013 and June 2014, the report lists 230 reforms, with sub-Saharan Africa accounting for the largest number. A trend in recent years has been towards collecting subnational indicators, to represent better the heterogeneity in some countries. For example, the 2015 report included data for two cities, rather than just the largest business city, for 11 economies: Bangladesh, Brazil, China, India, Indonesia, Japan, Mexico, Nigeria, Pakistan, Russia, and the United States.

The underlying indicators reveal striking differences across countries. For example, the number of days that it takes to start a business according to that indicator is 144 days in Venezuela, 90 days in Zimbabwe, 25 days in Argentina, and 2.5 days in Singapore. There is some evidence of within-country variation where this

information is collected, although the within-country variation tends to be small compared to the between-country variation.

The ranks on different dimensions of the Doing Business indicators tend to be positively correlated with (Spearman) rank correlations across the indicators typically between 0.3 and 0.6. That said, it is not hard to find cases where countries have quite different rankings across dimensions. For example, China is ranked at 35 for enforcing contracts but at 128 for starting a business, while Egypt has the opposite rank difference with 152 on enforcing contracts and 73 on starting a business. Such differences reinforce the need to look beyond the aggregate measures and to drill down into the specific performances across the indicators.

Overall, country rank in the Doing Business report tends to be strongly correlated with measures of development success, as well as with income per capita and with other standard measures of institutional quality, but this gives little insight into the direction of causation. Instead, this is likely to be an instance of what in Besley and Persson (2011) we have called "development clustering," the observed phenomenon that most dimensions of development move together. If the exercise is valuable for monitoring progress relative to using more standard measures of institutional quality and prosperity, it is because the specific indicators are worth exploring dimension-by-dimension.

The Value of the Exercise

The Doing Business project provides a unique perspective. But it is important for those who use the data to be familiar with how they are collected, rather than blindly downloading them and running regressions. The data are quite unique: there is no other comparable project in terms of scale or scope. Thus, the Doing Business report has the capacity to cast light on dimensions of policymaking that were not covered in previous datasets. The current chief economist of the World Bank, Kaushik Basu, states the case in his forward to the 2015 edition of the report:

> The public discourse on economic policy is overwhelmingly focused on fiscal measures, monetary interventions, welfare programs, and other such highly visible instruments of government action. Thus when an economy does poorly, a disproportionate amount of our debate centers on whether or not it needs a fiscal stimulus, whether there should be liquidity easing or tightening, whether its welfare programs have been too profligate or too paltry and so on. What gets much less attention but is equally—and, in some situations, even more—important for an economy's success or failure is the nuts and bolts that hold the economy together and the plumbing that underlies the economy. . . The World Bank Group's *Doing Business* report is an annual statement of the state of the nuts and bolts of economies around the world and, as such, is one of the most important compendiums of information and analysis of the basis of an economy's effective day-to-day functioning and development.

Prior to the Doing Business project, little was known about many aspects of the business climate. For example, most countries did not collect information on the time and effort that goes into starting a business, let alone produce such information in a way that could be compared internationally. Other areas where the incremental gains from the project in terms of bringing knowledge of cross-country differences into the daylight have been especially large include the areas of creditor rights and property registration.

Moreover, postponing for a moment the controversies about what is measured and how it is measured, it is important to note that the survey is conducted in a methodical and transparent way. The Doing Business indicators are updated annually, which means that policy reforms are soon reflected in the indicators. This feature is attractive to both researchers and also to policymakers, who can see their reform efforts translated quickly and directly into changes in the index. A consistent methodology has been pursued throughout the project, which means that country-level performances can be tracked over time. The Doing Business report is a living project that seeks to learn from criticisms with frequent efforts to improve the methodology and to expand the domain of the indicators. The methods used are clearly articulated and documented on the website along with the historical data. When indicators are changed, there are careful efforts to maintain comparability and to explain such changes. This approach contrasts with many other widely used databases that are used extensively by economists, such as the International Country Risk Guide, where data time series that are comparable before and after methodological changes can be hard to come by. The Doing Business project is also quite up-front about the limitations of its approach. For example, Table 2.3 in the 2015 report discusses the pros and cons of focusing on a standardized type of firm in the formal sector of the largest city as well as using opinions from lawyers to assess the environment for doing business in a country.

Of course, the fact that what is reported in the Doing Business data is clear and transparent should not be confused with the claim that it can tell us about the right policy mix for any particular economy. The parallel with more standard macroeconomic data is instructive. Measuring a fiscal deficit does not, by itself, yield any automatic conclusions for the best path for levels of taxing and spending. Moreover, details of specific types of taxes and spending and how they are implemented can be important. That said, nobody would wish to debate fiscal policy without some measure of the budgetary position in hand. Moreover, comparable data on international experiences using common methods of measurement are helpful in policy debates, not least in trying to draw lessons from a range of experiences.

In this spirit, the key question is whether the Doing Business data provides useful information that is relevant to real policy debates concerning the environment in which businesses operate. Policymakers in China or Brazil or Egypt have good reasons to be interested in how economies like Singapore or Sweden approach business regulation without deciding blindly that they should copy these practices. The Doing Business rankings provide a way into this question—the basis

for beginning a dialogue about policy reform. This discussion was not happening on a systematic basis before the Doing Business project came along.

Moreover, there is a valuable contribution to democratic debate made by the Doing Business data, which can be downloaded and read by citizens and policymakers of any country who wish to know how their country performs and to question whether this performance is justified. More generally, the report can be thought of as tool of "yardstick competition" between governments: that is, citizens use information that is available through the media to hold their governments to account based on performance comparisons (for example, Besley and Case 1995; Salmon 1987). After all, if the Doing Business comparison is not useful in some cases, it can be always be set aside. No country or politician or citizen is obliged to take notice of it.

That said, there is a concern that international pressure can be brought to bear when international aid is at stake and when the Doing Business indicators are used in policy dialogue or as a form of conditionality. This ties into wider debates about whether aid or conditions on aid are an affront to national sovereignty in general (for example, see Easterly 2013 for discussion), and is not a concern specific to the Doing Business project.

The World Bank has a well-established role in relation to data and statistics and, under its open data initiative, now makes much data freely available. These efforts by the World Bank acknowledge the key role that data play in development debates. Good examples of World Bank–supported initiatives are the World Bank's Living Standards Measurement Studies and the World Bank Enterprise Surveys. The collection of data can be valuable to citizens and civil society, even if some of the findings can in some cases embarrass the government of a country by highlighting policy failures.

Even the debates about the validity of the Doing Business indicators (discussed in the next section) point out the usefulness of the project; after all, without the project, such debates would have been based only on uninformed conjecture. Taken as a whole, the main achievement of the Doing Business project has been to shed light and create a more informed debate on a range of differences in laws and regulations across countries in areas where little was known on a systematic basis before the project began.

Criticisms and Caveats in Context

Concerns about the use of the Doing Business data and the indicators on which it is based fall into three main categories: 1) the way in which the project works, including some defects that are inherent in the design; 2) the validity of the indicators for policy choices or outcomes; and 3) the underlying objectives and motives of those who are designing the indicators. Let us consider these in turn.

The Nature of Exercise

The Doing Business report focuses on formal and legal requirements as assessed by legal professionals in that country. Given the specified type of reference

firm for which the data are collected, we would not expect the data to correlate with the experience of all firms in the economy.

A tradeoff arises here. On one side, given the differences in industrial structure and corporate forms across economies, the idea of a standardized firm is necessary for international comparisons, even if it limits what can be learned about the broader business environment. On the other side, the extent to which the specific reference frame created for the standardized firm indicates how firms in general experience the business climate is far from clear. For example, in many developing countries, large swathes of economic activity are conducted under the radar in the informal sector, where the role of formal rules and legal procedures is murky at best. For example, Schneider (2002) estimates that the median level of formal economic activity in the countries in his sample is around 67 percent, while in the bottom quartile of the countries in the distribution, around 49 percent of economic activity is formal.

A robust finding in the Doing Business report is that the countries which have a higher rank tend to have smaller informal sectors. This pattern suggests that the choice to become a formal firm may be a key margin affected by business regulation and formal laws. Of course, the Doing Business indicators may offer some insight into why some forms of formal enterprise are discouraged. Indeed, implicit within the Doing Business approach is the plausible belief that, in the end, it is likely to be the development of larger, formal sector firms that will be engines of employment creation and poverty reduction. But for that very reason, the way in which business conditions affect the extensive margin between whether firms choose to be formal and informal may be more important than how such rules affect the behavior of the formal sector taken alone.

Firms will experience the business climate differently. Firm-level surveys done at the national level, such as the World Bank Enterprise Surveys (at http://www.enterprisesurveys.org/), offer a useful complement to the Doing Business approach to examining regulation. They allow the range of firms surveyed to be broader than the stylized type of firm towards which the Doing Business indicators are targeted: for example, only about half of the firms in the Enterprise Surveys are privately held, limited liability companies, and the proportion in the data varies significantly by country. Such detailed surveys are expensive to implement and cannot be conducted annually for a broad range of countries. But they can ask useful questions about the experiences that enterprises have in dealing with government and in turn connect these experiences to measures of firm-level performance.

Table 2 describes how certain questions in the Enterprise Surveys are correlated with comparable questions in the Doing Business report. It reports whether correlations are positive or negative and includes an asterisk if they are statistically significant at 5 percent. The Enterprise Survey question used in each regression is given at the start of each panel. We then list the Doing Business indicators that were selected to correspond best to this question. Next we report the sign of each correlation with that indicator conditioning on year dummies and GDP per capita. The first column considers conditional correlations across all firms. The remaining columns look at whether these correlations are more robust for those firms in the

Table 2
Enterprise Survey Correlations with Doing Business Indicators

			Firm size (number of workers w)			*Firm in capital city?*		*Legal structure: limited liability?*	
		All firms	*Small: $w \leq 20$*	*Medium: $20 < w \leq 100$*	*Large: $w > 100$*	*Yes*	*No*	*Yes*	*No*
Enterprise Survey Question: *Biggest obstacle for firm is finance?* [no = 0, yes = 1]									
Doing Business topic: Getting credit	Strength of legal right index	+	+	+	−	+	+	+	+
	Depth of credit information index	−	−	−	−	−	−	−	−
	Public credit registry coverage, % of population	+	+	+	+	+(*)	+	+	+
	Private credit registry, % of population	−	−	−	+	+	−	−	−
Enterprise Survey Question: *How much of an obstacle is access to finance?* [High: More severe obstacle]									
Doing Business topic: Getting credit	Strength of legal right index	−	−	−	−	−	−	−	−
	Depth of credit information index	−(*)	−(*)	−(*)	−	−	−(*)	−	−(*)
	Public credit registry coverage, % of population	+	+	+	−	+	+	+	+
	Private credit registry, % of population	+	+	+	+	+	+	+	+
Enterprise Survey Question: *Biggest obstacle for firm is business regulation?* [no = 0, yes = 1]									
Doing Business topic: Starting a business	Procedures (number)	+(*)	+(*)	+(*)	+	+(*)	+(*)	+(*)	+
	Time (days)	−	−	−	+	−	−	−	+
	Cost of income per capita	−	−	+	−(*)	−	+	+	−
Enterprise Survey Question: *Biggest obstacle for firm is courts?* [no = 0, yes = 1]									
Doing Business topic: Enforcing contracts	Time (days)	+	+	+	−	+	+	−	+
	Cost (% of claim)	−	−	−	−	−	−	+	−
	Procedures (number)	−(*)	−	−	−	+	−(*)	−	−
Enterprise Survey Question: *How much of an obstacle is the court system?* [High: More severe obstacle]									
Doing Business topic: Enforcing contracts	Time (days)	+	+(*)	+	+	+	+(*)	+	+
	Cost (% of claim)	−	−	−	−	+	−	−	−
	Procedures (number)	−	−	−	−	−	−	−	−

(Continued)

Table 2—continued

			Firm size (number of workers w)			*Firm in capital city?*		*Legal structure: limited liability?*	
		All firms	*Small: $w \leq 20$*	*Medium: $20 < w \leq 100$*	*Large: $w > 100$*	*Yes*	*No*	*Yes*	*No*
Enterprise Survey Question: *Biggest obstacle for firm is customs and trade regulations?* [no = 0 , yes = 1]									
Doing Business topic: Trading across borders	Documents to export (number)	+	−	+	+	−	+	−	+
	Time to export (days)	+	+	+	+	+	+	+	+
	Cost to export (US$ per container)	−	−(*)	+	+	−	−	−	−(*)
	Documents to import (number)	−	−	−	−(*)	+	−	−	−
	Time to import (days)	−	−	+	−	−	+	+	−
	Cost to import (US$ per container)	+	+(*)	+	+	+	+	+	+(*)
Enterprise Survey Question: *How much of an obstacle are customs and trade regulations?* [High: More severe obstacle]									
Doing Business topic: Trading across borders	Documents to export (number)	+	+	+	+ (*)	+	+	+	+
	Time to export (days)	−	−	+	−	+	−	−	+
	Cost to export (US$ per container)	−	−	−	−	−	−	−	−
	Documents to import (number)	−(*)	−(*)	−	−(*)	−	−(*)	−(*)	−(*)
	Time to import (days)	+	+	+	+	+	+	+	+
	Cost to import (US$ per container)	−(*)	+(*)	+	+	+	+	+(*)	+(*)

Notes: Each panel of reported + or − is from a cross-country regression of the average answer to an Enterprise Survey on a selected set of *Doing Business* indicators. The sample-selection determines the group of firms within each country used to construct the Enterprise Survey average. All regressions include year fixed effects and the logarithm of per capita GDP in purchasing power parity (PPP) 2005 constant dollars. The Doing Business historical dataset for years 2004 to 2013 was accessed at http://www.doingbusiness.org/custom-query, retrieved on April 26, 2013. All economies across regions and income-groups were included and all topics except Getting Electricity and Paying Taxes were chosen. In the dataset, all observations which recorded ".." and "no practice" were coded as missing observations. Upon retrieval, the dataset contained 1,850 country-year observations. The World Bank Enterprise Surveys Standardized Data 2006–2911 was accessed on March 3, 2013 at the Full Survey data Portal of the Enterprise Survey website; the latest version of the Enterprise Surveys can be found here: http://www.enterprisesurveys.org/data/survey-datasets. Upon retrieval, the raw dataset contained 70,624 firm-country-year observations. The data series of GDP per capita, PPP at constant 2005 international dollars was retrieved from the World Development Indicators database on April 26, 2013, at http://databank.worldbank.org/data. For details, see online Appendix available with this paper at http://e-jep.org.
* Denotes 5% significance of the regression coefficient, robust standard errors.

sample who more closely resemble the firms towards which the Doing Business are targeted by varying the sample by firm size, whether the firm is a limited liability businesses, and whether the firm is located in the capital city.

The picture shown in Table 2 is quite mixed; a number of the correlations go in the expected direction, but are frequently not significant. And some puzzles also emerge. For example, the number of procedures to start a business is positively correlated with firms saying that regulation is the biggest obstacle to doing business, whereas the time taken to start a business is mostly negatively correlated. In this issue, Hallward-Driemeier and Pritchett also find only weak correlations between changes over time found in the firm-level Enterprise Surveys and the Doing Business indicators.

What to make of these patterns is largely a moot question. Enterprise-level surveys are a better way of exploring heterogeneity of the business climate within a country than the Doing Business Survey. However, data from the Enterprise Surveys do not necessarily give a better sense than the Doing Business project of country averages, given that firms will face different experiences with some dimensions of the regulatory and legal structures, which make the responses of professional lawyers in some ways more reliable. Moreover, firm-level data seek to be representative of the industrial structure as it stands, and cannot easily be used to explore how it would be different in the absence of barriers to formality or regulations. By definition, firms that do not exist because of adverse business conditions cannot be surveyed. Another potential explanation is that there is a distance between de jure and de facto regulatory processes. While it is impossible to say which of these factors could be at work, the poor correlation between Doing Business report indicators and firm-level surveys that ask similar questions is less surprising than it might seem at first blush.

Lawyers who are immersed in the complexities of law have been known to criticize the Doing Business indicators as a crude measurement that may fail to capture relevant complexities. This critique has some literal truth in it. By design, the Doing Business data is incapable of capturing the complexities of the legal system. As mentioned earlier, the answers to the Doing Business survey questions are in some cases a mixture of de jure legal requirements and what happens de facto to businesses. Moreover, there is some in-house World Bank processing of the received data, in a way that is not especially transparent. That said, the lawyers' criticism also reflects a culture clash between economists and lawyers. Economists can sometimes rightly be accused of being too willing to accept stylized and simplified characterizations of reality for the purposes of analysis. But the need for economic variables that can be expressed in a compact manner means that some short-cuts are inevitable.

The names of indicators or their content may in some cases be misleading, which emphasizes the need to look carefully behind the labels at the content of the indicators. For example, one of the indicators is Paying Taxes, which in the Doing Business data is proxied by the number of separate taxes to be paid, hours per year that the reference company would spend preparing tax returns, and the total tax payable by that reference firm as a share of gross profit. From the standpoint of the broad

questions in public finance—that is, how a tax system can be designed to raise revenues and to ensure compliance while avoiding unnecessary burdens on a country's citizens—the Doing Business measure is narrowly focused and not well-connected to the broader concerns. For example, two of the biggest challenges in public finance for many countries in recent years have been in rolling out a value-added tax and in having firms assist employees in their compliance with the personal income tax. The Doing Business ranking on paying taxes is actually on average higher (a "worse" performance) in countries which have introduced a value-added tax: for example, in the 2006 Doing Business data, the average rank of countries that had a value-added tax was 91 compared to an average rank of 68 among countries that did not have a value-added tax. While having a value-added tax may not cause this lower ranking, it emphasizes the point that there is more to having a good tax system than ease of compliance. The Doing Business indicators would measure the burdens imposed on firms, but from a social perspective, any gains in revenue-raising efficiency would need to be balanced against such costs. Of course, the shortcomings of this measure should be apparent to anyone who actually looks.

The Getting Credit indicator seems named in a potentially misleading way, too. The unwary consumer of this data might presume that it refers to actual measures of credit. Instead, it is based on a Legal Rights index based on collateral and bankruptcy laws and a Credit Information index based on the degree of sharing of credit information. This indicator is correlated with a range of credit outcomes, but it is not a measure of the quantity of credit nor how credit is allocated—or particularly the extent to which credit flows to the highest-return activities. Moreover, many aspects of the environment in which credit is obtained are not included, like the competitiveness of the financial system or measures of financial regulation such as capital requirements.

It is also important to realize that the Doing Business indicators do not capture anything close to a complete picture of the business environment. Indeed, they are not in any meaningful sense a "first best" set of indicators. The project is not based on a grand design that begins from the question: "What would it take to capture the complete business environment across countries and over time?" Instead, the project has tended to proceed as a bottom-up entrepreneurial exercise, taking opportunities to add indicators of particular interest, and using a lot of a priori judgment. As the Doing Business project matures (which will require that additional resources be found), it could develop a more comprehensive view of the business environment. At present, some notable gaps in looking at the bigger picture of the business environment include a wider view of infrastructure, competition policy, trade policy, and many dimensions of regulation such as workplace health and safety.

None of these issues should trouble an educated user of the data, who would be aware of such issues and could use the data accordingly. But journalistic accounts of the Doing Business rankings and the indicators on which they are based typically paint these data as a complete and representative picture.

If the Doing Business rank were strongly correlated with broad contours of the business environment, these limitations might not matter so much. Indeed, one

would expect strong correlations across the indicators, reflecting an overall level of state competence. For some purposes, this level of generality suffices. The fact that Singapore is top in the ranking in the Doing Business *2015* report is essentially sending this message. That Singapore does a little better than Sweden is probably not telling us very much of quantitative interest—they both have effective states that have helped to foster prosperity in a wide variety of ways. Equally, the fact that Chad or Libya have incompetent states probably does not hinge very much on the mix of indicators that are chosen to establish this.

For the basic task of getting regulatory and/or legal reform on the public policy agenda, just knowing where an aggregate ranking for a country is can be a useful start. But when the debate turns to specific dimensions of policy reform, the details matter. It is important not to follow any specific indicators slavishly, because it is always an open question whether the specific policy priorities are well reflected in the disaggregated Doing Business rankings. It's also important to remember that certain policy reforms are likely to have complementarities across several policy dimensions—economic and noneconomic—like steps to speed up court decisions and to train more competent lawyers.

The Validity of the Indicators

The Doing Business indicators have been a boon to research. Many economic studies (including some of my own) use the Doing Business indicators as either right- or left-hand-side variables in regressions. To give a flavor of the uses to which these data have been put, Table 3 lists some studies that use the Doing Business indicators on either the left- or right-hand side of regressions. There is no scientific basis for this selection of studies except that all of the studies are well-cited, suggesting that subsequent work has paid attention to them, for better or worse. In each case, the authors report patterns that emerge in comparisons across countries. In most cases, the concern is with whether a particular indicator is correlated with aggregate or firm-level outcomes, or whether, if used as a left-hand side variable, the indicators are correlated with country characteristics, history, or institutions.

Each study tells a story that can be elaborated in detail. To give a flavor of this, consider the highly influential paper by Djankov, McLiesh, and Shleifer (2007) listed in Table 3. It examined correlations between the Getting Credit indicators and a variety of credit market outcomes. They find that formal creditor protection, along with the existence of institutions that share information, are associated in cross-country data with a higher ratio of private credit to GDP. These correlations seem relatively more important in higher-income countries. They also find strong correlations between the legal origins of a country—in particular, whether it evolved as a common law country relying more on private contracting, or as a civil law country relying more on government regulation and ownership—and the Getting Credit indicators. This finding has fuelled debates about the importance of credit market reforms and their value. It has also raised the issue of whether some kinds of legal systems are more conducive to creditor protection. The findings are clearly intriguing. It would be naïve to think that strong policy conclusions can be

Table 3
Selected Studies that Make Use of the Doing Business Indicators

Author/Year/Title	*Core finding(s)*	*Doing Business indicator as left- or right-hand side variable?*
Djankov, La Porta, Lopez-de-Silanes, Shleifer, 2002, "The Regulation of Entry."	More burdensome regulation is associated with higher levels of corruption; greater size of informal economy; lower executive constraints; and less political rights	Right-hand side Left-hand side
Djankov, La Porta, Lopez-de-Silanes, Shleifer, 2008, "The Law and Economics of Self-Dealing."	Index measure of legal protection of minority shareholders against expropriation is positively correlated with financial development	Right-hand side Left-hand side
Djankov, McLiesh, Shleifer, 2007, "Private Credit in 129 Countries."	Common law is associated with higher creditor rights, while French civil law is associated with higher incidence of public credit registries. Increase in creditor rights and public registry incidence is associated with a higher ratio of private credit to GDP	Right-hand side Left-hand side
Djankov, La Porta, Lopez-de-Silanes, Shleifer, 2003, "Courts."	Procedural formalism, the extent to which dispute resolution is regulated, is associated with longer duration of dispute; lower enforceability; and, higher corruption	Right-hand side Left-hand side
Nunn, 2007, "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade."	Countries with better contract enforcement and judicial systems are also more specialized in production of goods for which relationship-specific investments are the most important	Right-hand side Left-hand side
Klapper, Laeven, Rajan, 2006, "Entry Regulation as a Barrier to Entrepreneurship."	Rate of firm-incorporation in 'naturally high-entry' industries is lower in countries where regulatory costs are higher	Right-hand side Left-hand side
Djankov, Hart, McLiesh, Shleifer, 2008, "Debt Enforcement around the World."	Index measure of efficiency of debt enforcement is correlated with per capita income and legal origins	Right-hand side Left-hand side

(Continued)

drawn from a single study of this kind. However, it may influence the climate of opinion around what matters in improving the legal framework for credit.

The fact that so much academic research has taken the Doing Business indicators seriously could be thought of as a prima facie case for using them in policy discussions. In general, the absence of a robust correlation between an indicator used as a right-hand side variable and an outcome of interest is taken to cast suspicion on whether policymakers should use the Doing Business indicators in policy dialogues. On the other side, finding a robust correlation is often taken to constitute a *prima facie* case for using a certain policy tool. Both claims should be treated with caution. All of the standard concerns and caveats of using regression methods to inform policy apply when using the Doing Business indicators. Is the measurement

Table 3—continued

Author/Year/Title	*Core finding(s)*	*Doing Business indicator as left- or right-hand side variable?*
Besley, Persson, 2009, "The Origins of State Capacity: Property Rights, Taxation and Politics."	Countries with greater historical incidence of external conflict also have stronger creditor rights and sharing of credit information, but there is no significant association with investor protection	Left-hand side
Lerner, Schoar, 2005, "Does Legal Enforcement Affect Financial Transactions? The Contractual Channel in Private Equity."	Countries that take longer time in contract disputes are less likely to rely on preferred stock and more likely to use debt for investment structure	Right-hand side
Cooley, Marimon, Quadrini, 2004, "Aggregate Consequences of Limited Contract Enforceability."	Cross-country lower contract enforceability is associated with larger economic growth volatility	Right-hand side
Bae, Goyal, 2009, "Creditor Rights, Enforcement and Bank Loans."	Stronger legal rights of creditors against defaulting debtors is associated with larger loan size, longer loan maturity, and a reduction in loan-spreads	Right-hand side
Berglof, Pajuste, 2005, "What Do Firms Disclose and Why? Enforcing Corporate Governance and Transparency in Central and Eastern Europe."	Voluntary public disclosure by firms of financial accounts is more prevalent in countries with better functioning legal systems	Right-hand side
Ciccone, Papaioannou, 2007, "Red Tape and Delayed Entry."	Countries with less burdensome regulation are associated with more entry in industries that benefited under expansionary global demand and technology shifts	Right-hand side Left-hand side
Freund, Bolaky, 2008, "Trade, Regulations, and Income."	The positive impact of openness on per capita income is reduced when there is more regulation	Right-hand side Left-hand side

precise enough? Are the independent variables exogenous? Correlations uncovered in this way are open to a number of common criticisms. For example, controlling for a range of omitted variables in cross-country regressions is always problematic. If government competence has benefits, then it could well be positively correlated with both the Doing Business indicators and the outcomes of interest.

In addition, given the frame of reference for the Doing Business project (a particular kind of firm in the largest business city), we would expect considerable measurement error to arise because this particular measure of the business climate may not apply to firms that do not fit the standardized profile. The usual assumption is that measurement error should tend to bias correlations—in this case between doing business indicators and measures of firm performance—towards

zero. However, if the Doing Business indicators are related to the composition of business enterprises, perhaps because there are fewer formal firms due to the business climate being poor, then the measurement error is no longer randomly distributed and the direction of bias is not clear a priori. This insight may further explain the inconsistent correlations reported on in Table 2.

Well-identified causal effects are frequently regarded as the "gold-standard" for much policy evaluation. But even if problems of endogenity and omitted variable bias could be overcome for the Doing Business project, well-identified causal effects are in many contexts of limited value for concrete policy advice. First, there is tendency to focus on the average effects whereas in practice there are important sources of heterogeneity that affect the impact of a policy reform in a specific context. Second, a persuasively established causal connection between a given indicator and an outcome of interest need not imply that reform in this policy dimension should be the highest priority for that country when there is limited capacity for reform. Third, there is no guarantee that reforms will be implemented effectively if they were attempted.

Finally, one should remember that the Doing Business indicators measure only formal processes. Countries vary enormously in the institutional structures outside the formal legal system. In China, for example, formal structures are weak but other mechanisms are conducive to private investment. Regression-based evidence struggles to capture this kind of contextual information. Moreover, the quality of formal legal institutions and informal responses may well be co-determined, making it hard to condition on such factors.

Taking stock of these concerns, it is tempting to conclude that using regression-based evidence as a basis for validating the Doing Business indicators is of limited value. But that conclusion would follow only from a narrow perspective on how evidence is accumulated and used in shaping opinion. The regression evidence matters as part of a narrative which blends theory and data without necessarily appealing to rigorous arguments about causation. Even well-identified causal effects, where they exist, mostly offer a one-dimensional perspective on what matters in policy debates, with external validity requiring a host of strong and untested assumptions. Observed correlations and their interpretation are frequently important in debates about public policy reform.

Deregulation Bias?

A frequently heard criticism is that the Doing Business indicators have a deregulation bias—that is, lower levels of regulation lead to better scores and rankings. In an ideal world, it should be possible to separate the measurement of regulations from their normative status; for example, the data could show whether hiring and firing is more difficult in one country or another, without necessarily implying a policy conclusion. However, if a measure is put into a ranking, then a line has to be taken on what constitutes a better performance.

The claim that there is a bias towards deregulation is difficult to assess. Some of the specific items in the Doing Business rankings are more about government

efficiency than about the merits of regulation. For example, it is difficult to argue that it is an important regulatory goal to impose especially long time delays or high costs for those who want to start a new firm, or register commercial property, or engage in international trade, or get a construction permit. In other cases, like measures of financial disclosure by firms, functional bankruptcy laws, and sharing credit information, the Doing Business indicators seem to have more to do with building useful institutions than with a bias toward deregulation.

However, the Doing Business project as a whole is clearly motivated by a belief that frictions due to poor regulation and an ineffective legal system inhibit the performance of firms and can therefore lower job creation and impede poverty reduction. This view does permeate the choice of indicators and the way that data are collected. But to argue that this creates a wholesale deregulation bias is an overstatement.

The greatest controversy over a potential deregulation bias arose regarding the Employing Workers measure, which looks at how easy it is to hire or fire workers, and the rigidity of working hours. These sorts of labor market rules are hotly contested political territory, enough so that the Employing Workers indicators were removed from the country ranking. (The International Labor Organization (2007) was among the voices arguing for this change.) However, even without taking a stand on the appropriate degree of labor market regulation, there is scope for nuanced work in this area. For example, labor regulation success could be judged on the basis of indicators that achieve more employment generation on one dimension, and more or less worker protection on a separate dimension, without trying to be specific about any tradeoffs between these two goals. Indeed, the *World Development Report* (World Bank 2012) took a rather broader approach to labor market issues to that implicit in the Doing Business indicators for employing workers.

The Political Economy of Data Collection

It is not difficult to understand why the Doing Business project has proven to be both influential and controversial. A range of academics have used the data, leading to publications in top journals. This gives the project academic credibility. Policymakers refer to the rankings and even target them as policy objectives. This gives the project salience. Concerned citizens and civil society organizations also use them. This puts the project at the center of global debates. Finally, the data deal with issues that go to the heart of debates about the role of the state in economic development. This makes the project politically charged.

The Doing Business report and the debates which it has provoked underline the important role that the collection and dissemination of data play in policymaking. It is easy to forget that national income accounts were created primarily as a tool for economic management. The collection of poverty statistics is intimately linked to the desire to assess social and economic progress as well as to monitor the success of policies. Doing Business follows in this tradition, purposefully collecting data as

the basis for scrutinizing policy and monitoring progress. That said, when the report is publishing rankings and distance to the frontier, it would be naive to believe that data collection and presentation is a purely technocratic process. The Doing Business indicators are part of the policy debate and hence have the potential to influence policymaking. They are used by governments who set their own goals and internally by the World Bank, as well as by other multilateral organizations and foreign aid agencies.

The interaction between data and politics can be tricky. One common concern when measures become salient is that policymakers may seek to "game" the indicators, rather than facing up to some of the more important challenges. This phenomenon is well-understood in other contexts and travels under a variety of names. For example, Campbell's (1976, p. 49) law states: "The more any quantitative social indicator (or even some qualitative indicator) is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." An alternative label is Goodhart's (1981, p. 116) law: "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes." This risk becomes particularly important when specific policy indicators can become politicized and hence salient in the eyes of policymakers. It is then important that there is not too much reliance on any particular single indicator, such as a ranking.

In the case of the Doing Business indicators, the accusation is that policymakers who desire to improve their Doing Business ranking may make pro forma changes in laws which have limited substantive value. The case of Rwanda, with 2015 Doing Business ranking of 47 (better than Italy) despite having per capita national income under $1,000 and more than 40 percent of its population in poverty, surfaces regularly in such discussions. For people to be taken in by gaming, it must be that the indicators are being used in a naive way, as ends in themselves, when the value of changes are being appraised. Any robust policy dialogue must rely on a much broader assessment of reform than on the Doing Business rankings. Those who have a strategic purpose will either play up or overlook data limitations, depending on their political goals. One role of professional economists is to help maintain standards of balanced analysis.

A more fundamental concern about the political economy is whether the World Bank has a legitimate role in collecting and publicizing this kind of data. As the Doing Business report has become influential, the question of accountability for the data and processes of internal scrutiny has become pressing. In the democratic world, at least, the datasets collected and their purposes come under the general aegis of democratic accountability, and derive their legitimacy in this way. Although statistical agencies in many countries enjoy a degree of independence, which enhances their effectiveness, they are ultimately accountable for the work that they do.

Accountability for data not collected by a democratic nation state is more open to question. For example, nongovernment organizations such as Transparency International, Freedom House, Amnesty International, and the Heritage Foundation collect data with a purpose and seek to influence policy through their efforts.

In these cases, accountability for the methods used and the validity of the measurement is frequently unclear. In addition, the fact that the data is being collected by an organization with a policy agenda may raise suspicions in the eyes of users about the independence of the data collection process.

Of course, the World Bank is not a free-standing organization, but instead is responsible to its member governments through its board. It is also monitored by and frequently criticized by civil society organizations. This reality may make it increasingly difficult to have a controversial exercise like the Doing Business project housed at the World Bank. Although providing capital for development projects remains a core part of the World Bank's activities, it is not the institution's exclusive focus. It also plays an important role in the development of data and ideas, and the World Bank would be a much less relevant organization if it chose always to back away from controversy. The World Bank has sponsored a variety of data-based initiatives: for example, since the 1980s, it has been involved in collecting the Living Standard Measurement Surveys, gathering rich data on households in countries around the world, and since the 1990s, it has been carrying out the Enterprise Surveys of firms.

The Doing Business project and the annual reports that it produces are a legitimate and useful part of the mission of the World Bank as a development organization. Moreover, there are sound reasons to collect data in a systematic way that allows comparisons across countries and over time. The exercise will never be free from controversy and the data must be used with appropriate caveats. It is also important to have independent oversight by those who are not viewed as having any particular policy agenda. The fact that data can be abused or that it may on occasion upset certain political interests would be poor excuses for not trying to collect it.

Data is an important compliment to democratic accountability in countries where democratic pressure can be applied. This insight also implies that the Doing Business report is destined to be most effective as a tool for inspiring debate over policy change in countries that already have an interest in making policy reforms. Indeed, the Doing Business rankings are likely to be much less influential in countries that already have the means of holding their governments to account and have evolved data relevant to doing this. Many developing countries around the world labor under repressive governments that limit freedom of debate, where the data cannot easily be incorporated into an accountability process. It is often in the weakly institutionalized parts of the world where governments are less attentive to their citizens' wishes that the loudest objections to data collection and dissemination are heard. However, access to the Internet and social media has made it much harder for governments to restrict access to data where they exist and the debates that they provoke.

As with any kind of economic data, nobody should use the measures from the Doing Business project without first understanding the details of how they are collected and what they do and do not measure. But even with its inevitable imperfections and growing pains, the Doing Business project seeks to measure issues

where economists knew almost nothing before the project began. Controversies over what the data can and cannot show should not be viewed as a distraction. Nor do they invalidate the work itself. They should instead be viewed as a useful and vital part of the public policy discussion. Such controversies illustrate how the measurement of key variables can fuel important policy debates and help to propel the impact of economic ideas.

■ *The author was a member of the World Bank panel that assessed the Doing Business project in 2013. He received no remuneration from any party for his input into the panel's work. He is a now a member of the Doing Business Advisory Panel for which he also receives no remuneration. The author is grateful for helpful input from the board of editors, Chang-Tai Hsieh, and Gillian Paull. Anders Jensen provided excellent research assistance.*

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369–1401.

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* New York: Crown.

Adelaja, Tai. 2012. "Vladimir Putin Aims to Boost Russia's Investment Climate." *Telegraph.* June 15, 2012. Supplement produced and published by *Rossiyskaya Gazeta.* http://www.telegraph.co.uk/sponsored/rbth/business/9333604/vladimir-putin-russia-investment.html.

Bae, Kee-Hong, and Vidhan K. Goyal. 2009. "Creditor Rights, Enforcement, and Bank Loans." *Journal of Finance* 64(2): 823–60.

Berglöf, Erik, and Anete Pajuste. 2005. "What Do Firms Disclose and Why? Enforcing Corporate Governance and Transparency in Central and Eastern Europe." *Oxford Review of Economic Policy* 21(2): 178–97.

Besley, Timothy, and Anne Case. 1995. "Incumbent Behavior: Vote-Seeking, Tax-setting, and Yardstick Competition." *American Economic Review* 85(1): 25–45.

Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99(4): 1218–44.

Besley, Timothy, and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters.* Princeton University Press.

Besley, Timothy, and Torsten Persson. 2014. "The Causes and Consequences of Development Clusters: State Capacity, Peace and Income." *Annual Reviews of Economics* 6: 927–49.

Buerkle, Tom. 2015. "Prime Minister Narendra Modi Has High Hopes for India." *Institutional Investor.* January 28. http://www.institutionalinvestor.com/article/3421742/banking-and-capital-markets-emerging-markets/prime-minister-narendra-modi-has-high-hopes-for-india.

Campbell, Donald T. 1976. "Assessing the Impact of Planned Social Change." The Public Affairs Center, Dartmouth College, Hanover New Hampshire, USA. http://portals.wi.wur.nl/

files/docs/ppme/Assessing_impact_of_planned_social_change.pdf.

Ciccone, Antonio, and Elias Papaioannou. 2007. "Red Tape and Delayed Entry." *Journal of the European Economic Association* 5(2–3): 444–58.

Cooley, Thomas, Ramon Marimon, and Vincenzo Quadrini. 2004. "Aggregate Consequences of Limited Contract Enforceability." *Journal of Political Economy* 112(4): 817–47.

Djankov, Simeon, Oliver Hart, Caralee McLiesh, and Andrei Shleifer. 2008. "Debt Enforcement around the World." *Journal of Political Economy* 116(6): 1105–1149.

Djankov, Simeon, Rafael La Porta, Florencio Lopes-de-Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *Quarterly Journal of Economics* 117(1): 1–37.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "Courts." *Quarterly Journal of Economics* 118(2): 453–517.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Law and Economics of Self-Dealing." *Journal of Financial Economics* 88(3): 430–65.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. "Private Credit in 129 Countries." *Journal of Financial Economics* 84(2): 299–329.

Easterly, William. 2013. *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor.* Basic Books.

Freund, Caroline, and Bineswaree Bolaky. 2008. "Trade, Regulations, and Income." *Journal of Development Economics* 87(2): 309–321.

Glaeser, Edward L., and Andrei Shleifer. 2003. "The Rise of the Regulatory State." *Journal of Economic Literature* 41(2): 401–425.

Goodhart, Charles. 1981. "Problems of Monetary Management: The U.K. Experience." *Inflation, Depression, and Economic Policy in the West,* edited by Anthony S. Courakis, pp. 111–46. Rowman & Littlefield.

Hallward-Driemeier, Mary, and Lant Pritchett. 2010. "How Business is Done and the 'Doing Business' Indicators: The Investment Climate when Firms have Climate Control." Unpublished paper.

International Labor Organization. 2007. "The World Bank *Doing Business* Report: The Employing Workers Indicator." For Debate and Guidance: Fourth Item on the Agenda, 300th Session of the Governing Body, Geneva, November 7. http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---relconf/documents/meetingdocument/wcms_085125.pdf.

Klapper, Leora, Luc Laeven, and Raghuram Rajan. 2006. "Entry Regulation as a Barrier to Entrepreneurship." *Journal of Financial Economics* 82: 591–629.

Lerner, Josh, and Antoinette Schoar. 2005. "Does Legal Enforcement Affect Financial Transactions? The Contractual Channel in Private Equity." *Quarterly Journal of Economics* 120(1): 223–46.

Nunn, Nathan. 2007. "Relationship-Specificity, Incomplete Contracts and the Pattern of Trade." *Quarterly Journal of Economics* 122(2): 569–600.

Salmon, Pierre. 1987. "Decentralization as an Incentive Scheme." *Oxford Review of Economic Policy* 3(2): 24–43.

Schneider, Friedrich. 2002. "Size and Measurement of the Informal Economy in 110 Countries Around the World." Unpublished paper.

UK Trade & Investment. 2014. "Guidance: Exporting to Peru." March 26. At Gov.uk: https://www.gov.uk/government/publications/exporting-to-peru/exporting-to-peru.

World Bank. 2012. *World Development Report 2013: Jobs.* Washington, DC: World Bank.

World Bank. 2014. *Doing Business 2015: Going Beyond Efficiency.* Washington, DC: World Bank.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 121–140

How Business is Done in the Developing World: Deals versus Rules[†]

Mary Hallward-Driemeier and Lant Pritchett

Over the last decade, two broad sets of facts about regulation of firms in the developing world have been established through both cross-national and individual country research. The first set of facts, established through sources including the "Doing Business" indicators from the World Bank, is that firms that attempt full regulatory compliance will face an extremely costly and time-consuming process. The second set of facts from a variety of sources shows that in practice, firms in developing countries are often able to sidestep the *de jure* legal rules, which makes intuitive sense because many developing countries have low rankings by international standards in categories like "rule of law," "bureaucratic quality," "government effectiveness," and "control of corruption." What happens when stringent formal rules that characterize the *de jure* investment climate in developing countries meet weak government willingness or capability to enforce those rules?

Evidence on the *de jure* legal and regulatory requirements facing firms often draw on evidence from the World Bank's Doing Business project, which in turn is built on the pioneering work of De Soto (1989) in Peru and others (for example, Djankov, La Porta, Lopez-De-Silanes, and Shleifer 2002, 2003; Botero, Djankov, La Porta, Lopez de Silanes, and Shleifer 2004; Djankov, McLiesh, and Shleifer 2007). The Doing Business project surveys experts concerning the legally required time and costs of regulatory compliance for various aspects of private enterprise—starting a

■ *Mary Hallward-Driemeier is Senior Principal Specialist, Jobs, World Bank, Washington, DC. Lant Pritchett is Professor of the Practice of International Development, Harvard Kennedy School, Cambridge, Massachusetts. Their email addresses are mhallward@worldbank.org and Lant_Pritchett@harvard.edu.*

[†] To access the Data Appendix and disclosure statements, visit
http://dx.doi.org/10.1257/jep.29.3.121

doi=10.1257/jep.29.3.121

firm, dealing with construction permits, trading across borders, paying taxes, getting credit, enforcing contracts, and so on—around the world. Each year, the Doing Business reports document large numbers of routine business procedures that would take many days. According to the most recent data, while it takes 9.2 days to start a business in the high-income OECD countries, compliance with these procedures would take on average 27 days in countries of sub-Saharan Africa, 30.1 days in Latin America, and 34.4 days in the East Asia and Pacific region. Similarly, following the required procedures to get a construction permit would take an average of 155 days in countries of sub-Saharan Africa, 178 days in Latin America, and 199 days in South Asia. Clearly, most developing countries of the world do not make it fast or easy to comply with the regulations that govern business.

A different body of research has examined the state's capability and willingness to implement policy and has established, in both the specifics and in overall rankings, which developing countries have especially weak governance. Chong, La Porta, Lopez-de-Silanes, and Shleifer (2014) document that even in relatively straight-forward services like handling misaddressed international mail, developing-country governments generally show little (or no) compliance with their own policies. Widespread dysfunction in the provision of government services also shows up in detailed studies of simple regulatory functions like giving driver's licenses (Betrand, Djankov, Hanna, and Mullainathan 2007), pollution reporting (Duflo, Greenstone, Pande, and Ryan 2013), and service delivery, including problems inducing health clinic staff to show up for work (Dhaliwal and Hanna 2014; Banerjee, Duflo, and Glennerster 2008) or using simple disease treatment protocols (Das, Holla, Das, Mohanan, Tabak, and Chan 2012). The Worldwide Governance Indicators report each year on "Government Effectiveness." On a scale in which Finland is 2.2, the United States is 1.5, and a high-income country with weak governance like Portugal is 1.03, a country like India is near the developing country middle of the pack at −.2, and even relatively effective developing country governments like Colombia or China score near zero.

Given these two well-established facts about developing countries—a *de jure* legal environment that creates complex and burdensome regulatory procedures, together with weak governance in implementation—the question arises: how is business actually done? In India, compliance with the formal rules to get a construction permit would take 25 procedures and 186 days. On the other hand, one could get a driver's license in Delhi without taking the legally required driving exam (or knowing how to drive) by hiring an agent to facilitate the process. In Cambodia, the Doing Business data reports it would take 20 procedures and 652 days to get a construction permit, but in 2012, its "government effectiveness" was −.8 and "control of corruption" −1.0. If government enforcement is not very effective and bribery is fairly common, how long does it really take?

We have long known from country-level studies of tax compliance or regulatory burdens that the *de jure* and the *de facto* can diverge enormously. There are entire literatures on tax evasion of all kinds in developing countries: for a few examples, see Pritchett and Sethi (1994) on tariffs; Fisman and Wei (2004) on misdeclaration

of import classification; and Gauthier and Gersovitz (1997) and Carillo, Pomeranz, and Singhal (2014) on misreporting costs. Stone, Levy, and Paredes (1996) showed that although Chile was far superior to Brazil on the formal measures of business procedures, informality meant that in actual practice there was little difference on several indicators.

In this paper, we focus on providing some cross-national evidence about how business is actually done. The World Bank, in addition to its Doing Business and Worldwide Governance Indicators, has also helped carry out firm-level Enterprise Surveys around the world that ask questions of managers at a wide array of firms about their business. Managers from firms were asked questions about how long it took to go through various processes like obtaining an operating license or a construction permit, or bringing in imports. The main task of this paper is to compare the results of three broadly comparable indicators from the Doing Business and Enterprise Surveys, drawing on results from two previous papers (Hallward-Driemeier, Khun-Jush, and Pritchett 2010; Hallward-Driemeier and Pritchett 2011).

Overall, we find that the single numerical estimate of legally required time for firms to complete certain legal and regulatory processes provided by the Doing Business survey does not summarize even modestly well the experience of firms as reported by the Enterprise Surveys. The Doing Business estimates are problematic in four ways: First, there is huge variance reported by firms within the same country. Second, the average times reported *de facto* in the Enterprise Surveys are much, much less than *de jure* times reported by Doing Business. Third, there is almost zero correlation across countries between the single Doing Business survey number and the Enterprise Survey responses of firms. Fourth, for those countries with repeated Enterprise Surveys data, changes in the reported Doing Business times are not strongly associated with changes in actual times as reported in the Enterprise Surveys, and if anything, reductions in Doing Business times are associated with *higher* reported actual times in the Enterprise Surveys.

When strict *de jure* regulation and high rates of taxation meet weak governmental capabilities for implementation and enforcement, we argue that researchers and policymakers should stop thinking about regulations as creating "rules" to be followed, but rather as creating a space in which "deals" of various kinds are possible. Within this space there will be winners and losers—among firms and among officials. The structure of these deals can in some cases reduce times to get permits, but the very uncertainty itself can favor officials (Hallward-Driemeier, Khun-Jush, and Pritchett 2010; Freund, Hallward-Driemeier, and Rijkers 2015). In the conclusion, we reflect on how the "deals versus rules" distinction helps us analyze policy formulation and the effects of reform in many low- and middle-income countries.

The Anatomy of Three Comparisons

The Doing Business Reports from the World Bank (data and detailed description of the methods are available at http://www.doingbusiness.org/) assess eleven

elements of a country's policies that relate to private firms, including starting a business, trading across borders, dealing with construction permits, enforcing a contract, and paying taxes. The measures for each element are created by experts (one to four lawyers or accountants) in each country who are asked to estimate the "typical" time and cost that it would take a hypothetical "standard" firm (privately and domestically owned, limited liability company with 10–50 employees, operating in the country's largest city) to comply based on their assessment of formal regulations as they exist on the books. To promote comparability, guidance is provided such as use the case of "building a warehouse" for assessing construction permits, or the case of "a contract worth twice the country's income per capita" for assessing the enforcement of contracts. The survey explicitly assumes full compliance by the business, that no third parties are used to "facilitate" completing procedures, and that no payments or other activities are undertaken to influence policy outcomes. Doing Business provides the raw data for each country as well as rankings across countries for each of these 11 regulatory areas and an overall "Ease of Doing Business" ranking that combines ten of the categories (leaving out the Hiring Workers category).

In contrast, the Enterprise Surveys do not try to measure what *should* happen as a result of formal policies and regulations, but rather what *did* happen in practice. They gather responses from large, random samples of firms in the main urban centers in a country, using both questionnaires and interviews. Registered firms above a minimum size, generally 10 employees, are included, as well as foreign and state-owned enterprises. The questionnaire includes modules that ask about the firm and its operations. The face-to-face interviews ask what the firms perceive as obstacles to their own business, but also quantitative information from the firm owners and senior managers about how long it takes to get various regulatory procedures done and how much they cost—including actions to influence policy implementation like bribes or gifts or meeting with government officials.

The Enterprise Surveys have covered over 130,000 firms in 135 countries, some with repeat surveys. The Enterprise Surveys build on a number of prior efforts with different names, including the Regional Program on Enterprise Development (RPED) in Africa, Business Environment and Enterprise Performance (BEEPS) in Eastern Europe and Central Asia, and Investment Climate Surveys, that were conducted prior to 2005. Those surveys used generally the same questionnaire but with samples restricted to manufacturing firms (the Enterprise Surveys also include services). Since 2005, the Enterprise Surveys have been standardized and all use the same questionnaire and methodology, while maintaining their earlier-established names in their respective regions.

A substantial body of academic empirical work has drawn on these data. The Doing Business data has been used in a wide array of applications: some prominent examples include studying the effect of labor regulations (Micco and Pagés 2006), regulation of entry (Bruhn 2011; Klapper, Laeven, and Rajan 2006), or a wide set of regulations on cross-country outcomes of interest (Loayza, Oviedo, and Servén 2006; Freund and Bolaky 2008; Eifert 2009). The data from the Enterprise Surveys and its precursors under different names have also been widely used (for surveys, see

Bigsten and Söderbom 2006; Xu 2010). Some prominent examples include studying finance, corruption, and property rights (Beck, Demirgüç-Kunt, and Maksimovic 2005; Ayyagari, Demirgüç-Kunt, and Maksimovic 2008; Fisman and Gatti 2006; Cai et. al 2006; Fisman and Svensson 2007); the relationship between the business environment and firm growth, (Dollar, Hallward-Driemeier, and Mengistae 2005, 2006; Haltiwanger and Schweiger 2005; Hallward-Driemeier, Wallsten, and Xu 2006; Aterido, Hallward-Driemeier, and Pages 2011; Eifert, Gelb, and Ramachandran 2008; Fernandes and Pakes 2008); firm innovation (Almeida and Fernandes 2008; Ayyagari, Demirgüç-Kunt, and Maksimovic 2010); informality (La Porta and Shleifer 2008); and labor protections (Pierre and Scarpetta 2006; Almeida and Carneiro 2009).

The most interesting difference between these two approaches to assessing the business or investment climate is that Doing Business focuses on *de jure* processes and Enterprise Surveys on *de facto* practice. The Doing Business survey requests that the respondents—who are local experts but not firms—answer the questions under the assumption that firms fully comply with the rules and that no direct or third-party facilitation (monetary or otherwise) is involved. The Enterprise Surveys provide the reported experiences of firms. To put it another way, Doing Business estimates the time and monetary costs for dealing with government regulations on the assumption of compliance and no deals, whereas the Enterprise Surveys ask explicit questions about how the firms make deals (for example payment of bribes and time spent with officials). Indeed, many firms do report engaging in a wide variety of influence activities. (Perhaps Doing Business experts answer some questions based on the experiences of the firms for which they work and hence are influenced by the *de facto*, but this would mean the huge gaps we find between the two sources of data are an *under*estimate of the "true" gap.)

We identify three particular questions where the conceptual correspondence between what the two approaches are attempting to measure appears high and where the answers are expressed in the same metric of days.

The first set of questions involves construction permits. Because the Doing Business estimate is a single number, it has to be about a specified type of building, which is a warehouse to be used "for general storage activities, such as storage of books or stationery" and "not be used for any goods requiring special conditions, such as food, chemicals or pharmaceuticals." It will have two above-ground stories and 14,000 square feet, and each floor will be three meters high. It will be on the fringe of the city, with road access, on a plot of land that is 10,000 square feet. The warehouse will be "valued at 50 times income per capita." In contrast, a typical Enterprise Surveys question on the same topic asks: "Over the last two years, did this establishment submit an application to obtain a construction-related permit?" And if the answer is yes, then: "In reference to that application for a construction-related permit, approximately how many days did it take to obtain it from the day of the application to the day the permit was granted?"

A second set of questions is about starting a business or getting an operating license. In Doing Business, the expert is asked to count the number of required procedures needed before actual operation of the standardized business, the time

needed, and any minimum capital requirements. The Enterprise Surveys ask, "Over the last two years, did this establishment submit an application to obtain an operating license?" and if yes, then, "Approximately how many days did it take to obtain this operating license from the day of the application to the day it was granted?"

The third set of questions involves the time delay in importing goods. Doing Business spells out a standardized import: "traded product travels in a dry-cargo, 20-foot, full container load. It weighs 10 tons and is valued at $20,000," is not a hazardous material, is not for the military, does not require refrigeration, and so on. The Enterprise Surveys ask, "Were any of the material inputs or supplies purchased in fiscal year [insert last complete fiscal year], imported directly?" and if yes, then, "In fiscal year [insert last complete fiscal year], when this establishment imported material inputs or supplies, how many days did it take on average from the time these goods arrived to their point of entry (e.g. port, airport) until the time these goods could be claimed from customs?" The Doing Business overall estimate of import time includes an estimate of inland transport time—which is explicitly not included in the Enterprise Surveys, so we subtract that from the Doing Business total time of the estimated inland transport.

While both surveys are eliciting length of delays to private firms in getting construction permits, starting businesses, and importing goods, the approaches are not identical—and cannot be made so. Doing Business is attempting to estimate a single "typical" number for a necessarily standardized firm and transaction, and the Enterprise Surveys are eliciting responses from a wide range of firms. In this paper, we report results using the full Enterprise Surveys sample, but in our previous papers we did the analysis with only those firms in the Enterprise Surveys that match the "standard" Doing Business firm (10–50 employees, domestically owned, in capital city) and found the same basic results.

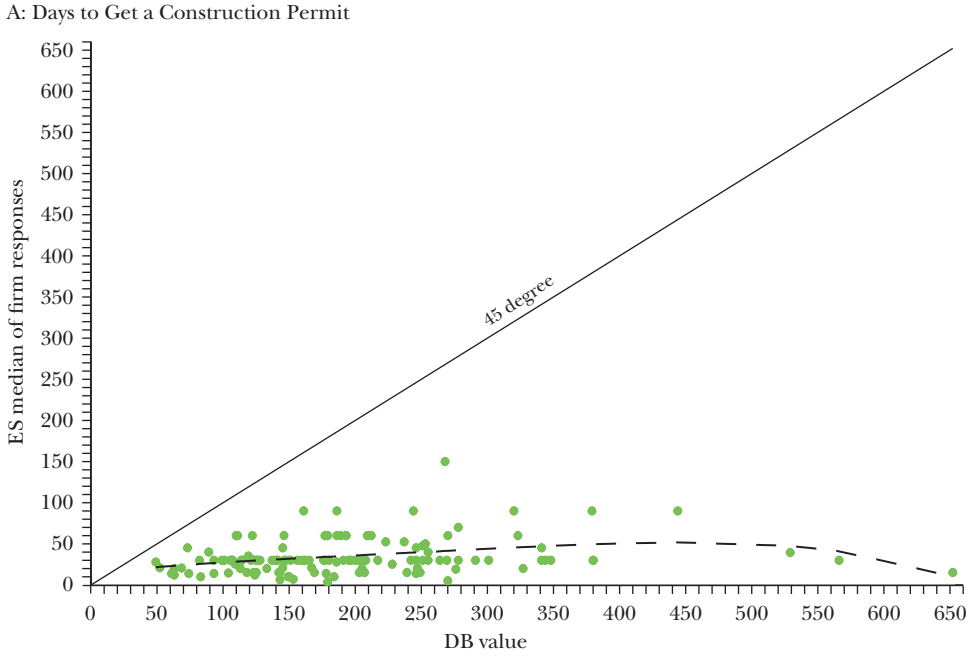
Respondents to the Enterprise Surveys are unlikely to have precise records as to the length of time procedures took and so almost certainly cluster responses around "focal" times like 7 or 15 or 30 or 60 days. This creates noise, but we don't know for sure whether it creates downward or upward bias in estimates of the gaps between Doing Business and Enterprise Surveys.

Comparing Country Averages: Levels and Changes

Before looking at a scatter plot between the Doing Business (DB) and the Enterprise Surveys (ES) results across countries, it is worth thinking about what one would expect. If firms are actually complying with regulations as measured by the Doing Business assessment then one might expect to see a one-to-one, linear relationship between the DB days and the central tendency of the firms' reported days. That is, if one country had DB days 100 days higher than another country then the central tendency of the ES days in that country would also be about 100 days higher.

All three sets of questions show the opposite pattern. First, for most of the range of the data, the median (typical) firm reports much quicker times completing

Figure 1
Comparing Doing Business (DB) and Enterprise Survey (ES) Results



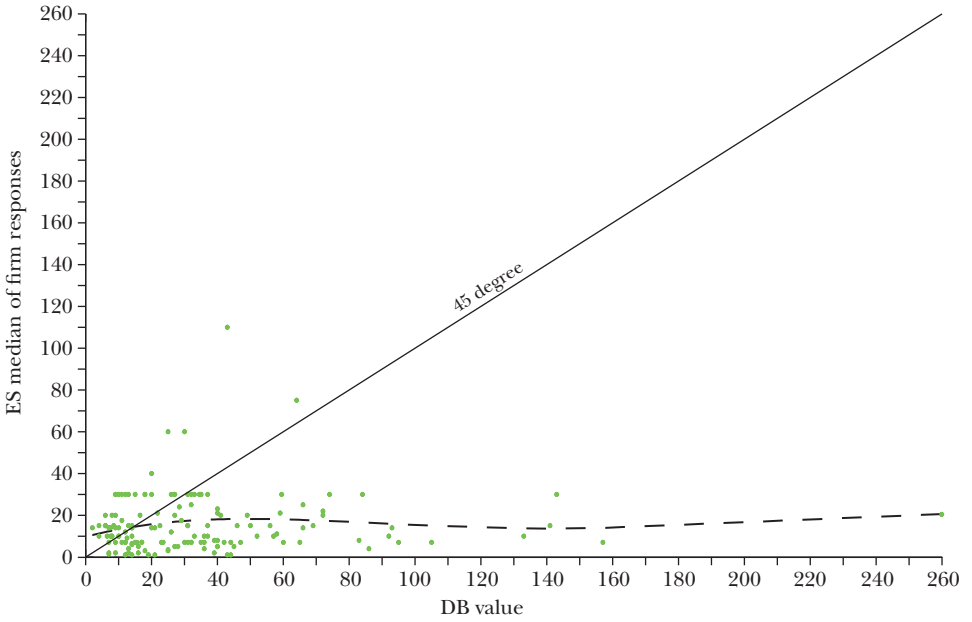
the three business processes than the Doing Business suggests. Second, the power of the country Doing Business measure of *de jure* compliance times for explaining the median of the firm Enterprise Surveys reported times is very weak.

Construction Permits

Figure 1A shows the scatter-plot between the Doing Business reported time and the median of the times reported in the Enterprise Surveys by firms who obtained permits for all 137 available country/year observations between 2006 and 2014 in non-OECD countries. The 45-degree line depicts a one-to-one linear relationship. The median Doing Business reported time is 177 days. The median of the Enterprise Surveys firm medians is 30 days. So the typical gap between *de jure* and median *de facto* times is about five months. A striking feature of the data is that "30 days" or "one month" is clearly a focal response (visible as the horizontal band of dots), and it is the median firm response over the entire range of Doing Business values. Tunisia at 93 DB days, Mongolia at 197 DB days, Paraguay at 291 DB days, Venezuela at 380 DB days, and Zimbabwe at 566 DB days—a range across these countries of a year and a half in the *de jure* estimated times—all have their typical ES firm report 30 days. The Enterprise Survey values do not increase quickly as we move between from lower to higher responses; therefore, the larger the Doing Business value, the larger is the gap between Doing Business and Enterprise Surveys results. There is

Figure 1 (continued)

B: Days to Start a Business/Get Operating License



clearly little association between the two variables: specifically, the R^2 of regressing the Enterprise Surveys median on a quartic (to allow a nonlinear and flexible functional form) in Doing Business values is only .11.

Start a Business/Operating License

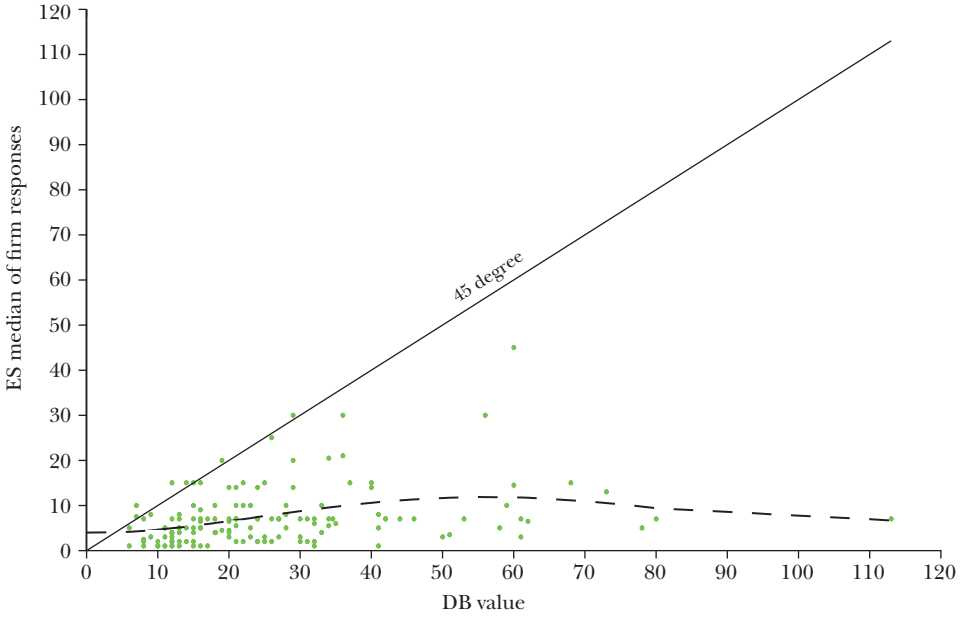
Figure 1B shows the data for starting a business/getting an operating license. Again there is little or no association between the two variables: the R^2 of a quartic in Doing Business values in explaining variation in the median of Enterprise Surveys values is only .023. In this case, at low values of the Doing Business time (estimated assuming full regulatory compliance) there are some countries in which the median of Enterprise Surveys responses (reflecting actual experiences) is above the Doing Business value. But at higher levels (above 30 days), the reported medians from the Enterprise Surveys are nearly uniformly below the *de jure* Doing Business values (the four observations with Enterprise Survey medians at 60 days and above are two from Argentina and two from Uruguay). Again, given the clumping of the Enterprise Surveys medians in 24 countries, the median is 30 days—which span a range of Doing Business reports from 9 days in Panama to 74 days in Peru to 143 days in Venezuela.

Importing Goods

Figure 1C, dealing with times for importing goods, shows the same two striking features as the previous graphs: little or no association between the two variables and

Figure 1 (continued)

C: Days to Import/Clear Customs



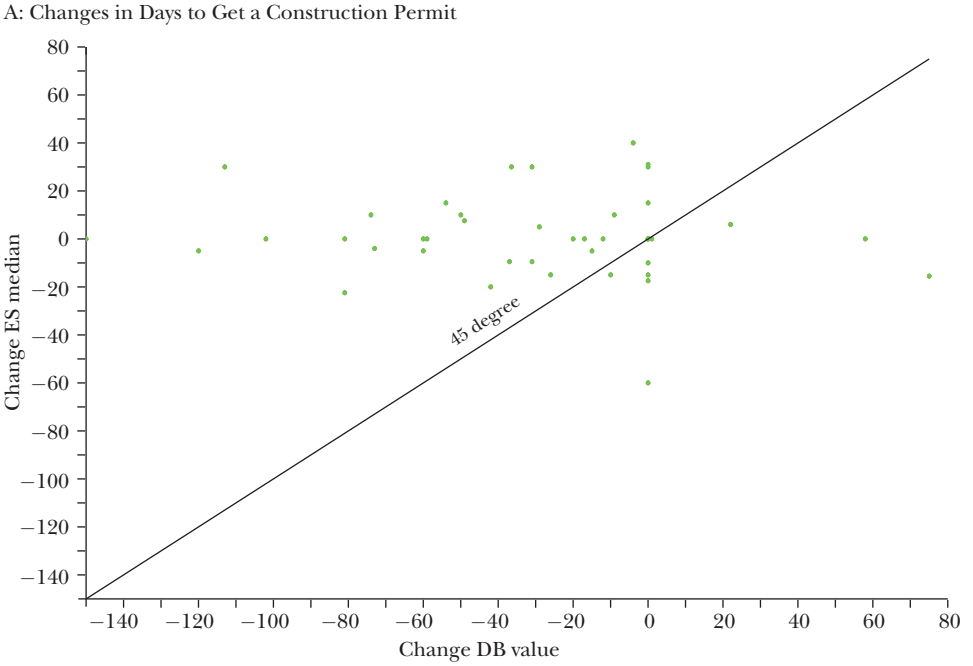
Source: Authors' calculations using Doing Business and Enterprise Surveys data.
Note: Figure 1 shows scatter-plots between Doing Business reported times (reflecting the *de jure* business environment) and the median of the times that firms reported experiencing in the Enterprise Surveys for all 137 available country/year observations between 2006 and 2014 in non-OECD countries with responses from more than 20 firms for the relevant indicator.

clumping of Enterprise Survey values around focal responses. The median time to import goods (net of inland transport) in the Doing Business data is 21 days and the median of the Enterprise Surveys medians is 6.25 days. While some few observations are near the 45-degree line (the Doing Business value and Enterprise Surveys median nearly equal), nearly all are below the line. The quartic regression R^2 is .125.

In one of our previous papers, we experimented with a variety of functional forms (linear, quartic, and different splines) between the Enterprise Surveys and Doing Business data for all three pairs of variables. None produced evidence that Doing Business values and Enterprise Surveys medians were strongly associated (much less in a one-to-one relationship) over any range of the data. The data pretty much laughed at hypothesis tests that the functional form was linear and the coefficient on Doing Business variables was 1 (that is, the *t*-statistics were often above 30).

While it might be the case that across nations the Doing Business and Enterprise Surveys are uncorrelated, it could be the case that *changes* in the Doing Business times (reflecting changes in the regulations) are related to changes in firms' reported experience in the Enterprise Surveys, so that reductions in the Doing Business days are associated with reductions in days as reported by firms. We

Figure 2
Changes in Doing Business and Enterprise Survey Results

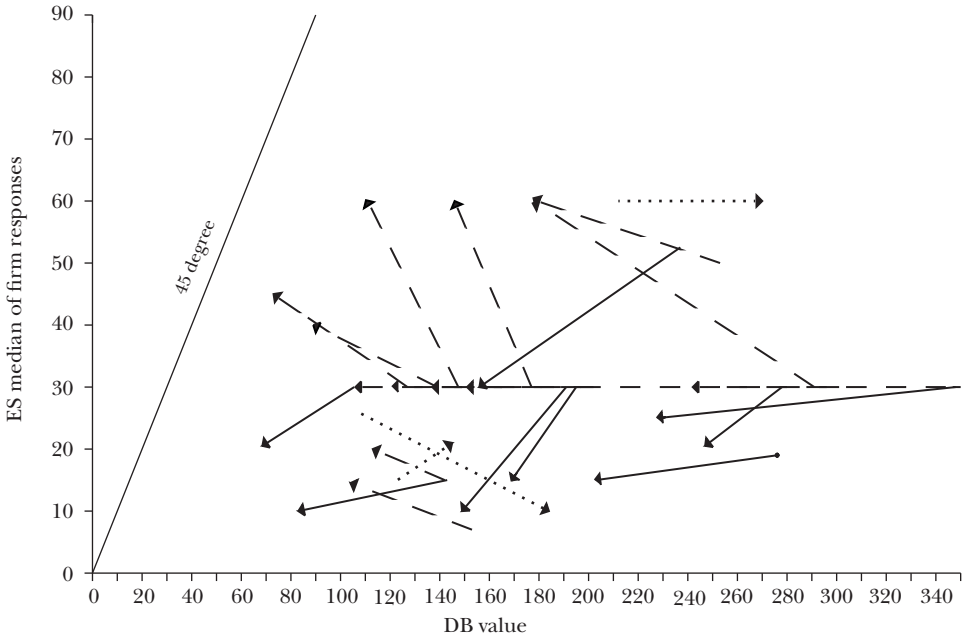


can examine this for fewer countries because, while the Doing Business data are updated every year, firm Enterprise Surveys are conducted only occasionally, and formal regulatory changes occur only periodically. Figure 2A focuses on changes in the time to get a construction permit and shows the changes in the Doing Business value and changes in the Enterprise Surveys median value, for each of the 45 countries for which there was a repeated ES survey between 2006 and 2014; we used the longest possible gap if there were more than two Enterprise Surveys available. Changes show the same lack of correlation as levels.

Figure 2B shows the evolution of Doing Business and Enterprise Surveys values for the 24 countries that had a change in their Doing Business reported value of at least 20 days for getting a construction permit along with a repeated ES survey that straddles this reform. This graph shows arrows connecting the before and after DB/ES-median coordinates for the countries, with the arrowhead at the most recent values. All but three arrows point left, indicating a reduction in the Doing Business value. If the Enterprise Surveys median also declined, the arrow points southwest (and is solid). This is 8 of the 21 experiences of reduction in Doing Business values greater than 20 days. Note that in this graph, the axes are not equal, so the 45 degree line of "one for one" change in Doing Business and Enterprise Surveys is much steeper, and even countries with large Doing Business reductions had modest Enterprise Surveys changes. For instance, in Angola, the Doing Business value

Figure 2 (continued)

B: Evolution of Days to Get a Construction Permit



Source: Authors' calculations using Doing Business and Enterprise Surveys data.
Notes: Figure 2A focuses on changes in the time to get a construction permit and shows the changes in the Doing Business value and Enterprise Surveys median, for each of the 45 countries for which there was a repeated Enterprise Surveys survey between 2006 and 2014, using the longest possible gap if there were more than two Enterprise Surveys available. Figure 2B shows the evolution of Doing Business and Enterprise Surveys values for the 24 countries that had a change in their Doing Business reported value of at least 20 days for getting a construction permit along with a repeated Enterprise Surveys survey that straddles this reform. Croatia was excluded because its high Doing Business values make the graphs hard to read. This graph shows arrows connecting the before and after DB and the ES median coordinates for the countries, with the arrowhead at the most recent values.

decreased from 276 to 203 from 2006 to 2010, but the median Enterprise Surveys was only 19 days in 2006, and the median days declined by 4 days to 15. The other 13 episodes of large Doing Business reduction either show no change or an *increase* in the time reported by firms (point northwest).[1]

The results of changes in days for starting a business/operating license or clearing customs are similar, with a weak (not statistically significant) *negative* association between reductions in Doing Business times and reductions in the Enterprise Surveys medians for both indicators (whereas a one-for-one relationship would produce a coefficient of *positive 1*).

[1] It is possible that in some countries there was only a short lag between the Doing Business reform and the launch of the second Enterprise Surveys such that some respondents' answers could still refer to pre-reform conditions. If this is true, it would likely cause the arrows to be horizontal pointing left.

The "Fast" and "Slow" Firms within Countries

The previous section focused on comparing the *central tendency* or median of firm responses in the Enterprise Surveys. However, for any given country or question, the Enterprise Surveys also reveal enormous reported differences *across* firms within each country. When firms are asked about time to obtain an operating license or a construction permit, or to clear goods through customs, the responses not only do not cluster tightly around the Doing Business value, they also do not cluster around the Enterprise Surveys central tendency. A significant fraction of firms report that regulatory compliance (or their version of compliance) takes essentially no time at all, while others report significant delays.

This pattern leads to three additional facts. First, the variance across firm responses in the *same* country is typically much larger than the variance in actual median compliance times *across* countries. Second, the firm-reported times are highly right-skewed, such that the firms for which compliance takes a long time takes a very long time—the 90th percentile is much further from the median than the 10th percentile. The implication of these two facts is that while the "fast" (10th percentile) firms report times that appear completely impervious to the regulations considered by Doing Business, the "slow" (90th percentile) firms report delays often significantly longer than the Doing Business times. Here, we consider these three facts in more detail.

Is It Where You Are or Who You Are?

In thinking about regulatory compliance, there is a lot of discussion of the investment "climate," a term which suggests that firms face similar conditions. But what if each firm, through a variety of ways of influencing regulatory outcomes, has its own particular temperature? The Enterprise Surveys data show that there is typically far more variation across firms in the same country than in the typical firm across countries. The 90th percentile of the median time to get a construction permit is 60 days (shared by Peru, Bolivia, Afghanistan, and Brazil) and the 10th percentile is 14 days (shared by Philippines, Liberia, and others). So the cross-national difference between typical "slow" and "fast" experiences is 46 days. Still looking at construction permits, only 9 of 136 country/year observations have a within-country 90th–10th spread lower than 46 days, and the median 90th–10th gap is 116 days. For getting an operating license and clearing customs the same is true: the gap in reported times between "fast" and "slow" countries is only about half the typical spread between "fast" and "slow" firms in the same country, as shown in Table 1.

Short is Always Short, Long is Very Long

There is large variation across countries in the Doing Business reports of regulatory times for compliance: for example, the difference between the 10th percentile country (Burundi at 99 days) and the 90th percentile country (Albania at 320 days) on construction permits is 221 days, suggesting this takes seven months longer in slow than fast countries. However, among the "fast" firms in each country there is

Table 1

The Variance in Firm Reported Times for Dealing with Regulations is Much Larger within Firms in the Same Country than for the Typical Firm across Countries

	Within-country 90th–10th percentile gap in days, median across countries	*Median of 90th percentile country minus median of 10th percentile country (in days)*
Construction permit	116	46
Operating license	55	26.5
Clear customs	27	14

Source: Authors' calculations using Doing Business and Enterprise Surveys data.
Note: The gap in reported times between "fast" and "slow" countries is only about half the typical spread between "fast" and "slow" firms in the same country.

very little variation at all as even in the slowest countries, the fast firms report very little delay. The difference between the "fast" firms in the slowest country and the "fast" firms in the fastest country is only 13 days for construction permits, 5 days for operating licenses, and 3 days for clearing customs. In contrast, the variation across countries appears predominantly at the upper tail. But the difference between the "slow" firms in slow countries and "slow" firms in fast countries is very large—the 90th to 10th percentile gap across countries in the 90th percentile of the firm reported times in the Enterprise Surveys is 305 days in construction permits (larger than the 221 days across the DB times), over five months for the operating license, and almost two months for clearing customs , as shown in Table 2.

Fast and Slow Firms and the Doing Business Measures

Figure 3 shows the scatter plots of the fast firms for each country (circles) and slow firms for each country (triangles) in the Enterprise Survey for each country measured against the Doing Business days reported for construction permits. The quadratic fit is shown for fast firms (circles) DB and slow firms (triangles). The "fast" firms report very little delay in any of the three indicators—clearing customs, getting construction permits, or operating licenses. For example, the median Doing Business time to get a construction permit is 177 days, about six *months*. The data suggest that the "fast" firms are not delayed at all by increases in the Doing Business indicator—though of course they may have to undertake other influence activities to achieve these rapid times.

For "slow" firms, those firms where regulations seem to be an obstacle, the extent of delay for those firms does vary widely across countries. The "slow" firms—those in the 90th percentile—show huge differences across countries; the standard deviation across countries in getting a construction permit for the slow firms (90th percentile) is 130 days (Table 2). And in a slow country (90th percentile), the slow firms (90th percentile) take a year to get a permit (compared to only 60 days in the fast 10th percentile country). However, as we see in Figure 3, even

Table 2
Variation across Countries in the Doing Business Days, Variation across Countries in Reported Regulatory Times of the "Fast" (10th percentile) Firms, and Variation across Countries in the Reported Regulatory Times of the "Slow" (90th percentile) Firms
(time in days)

	Gap between 90th and 10th percentiles across countries for DB, and for ES percentiles			*Standard deviation across countries for DB and ES percentiles*		
	Gap between 90th and 10th percentile countries in Doing Business	*Fast firms (10th percentile of Enterprise Survey responses)*	*Slow firms (90th percentile of Enterprise Survey responses)*	*Doing Business*	*Fast firms (10th percentile of Enterprise Surveys responses)*	*Slow firms (90th percentile of Enterprise Survey responses)*
Construction permit	221	13	305	98	5	130
Operating license	64	5	160	34	3	80
Clear customs	41	3	53	17	2	19

Note: There is large variation across countries in the Doing Business days, very little variation across countries in reported actual times for the "fast" firms, and massive variation across countries in the reported actual regulatory times of the "slow" firms. See text for details.

these large differences are not well explained by an association with the Doing Business indicators. There is some nonlinear association between the 90th percentile of ES firm responses and the DB values, but even there, the fit is far from tight.

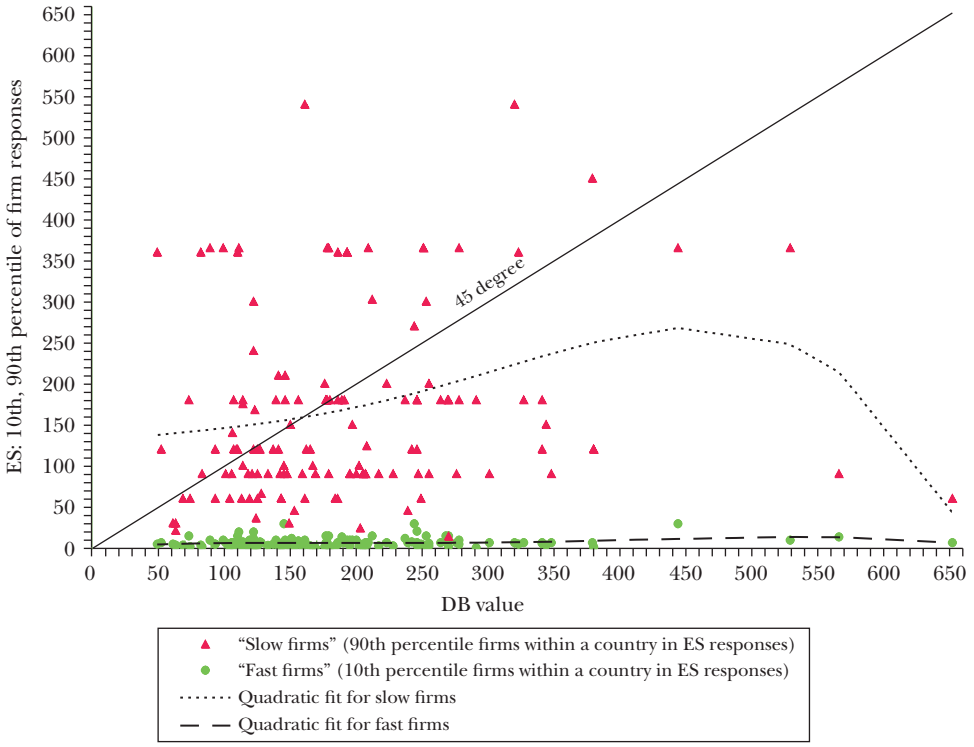
Figure 3 could be taken to represent a saying attributed to various South American presidents: "For my friends, anything; for my enemies, the law." The interpretation of the comparison of Doing Business and Enterprise Surveys cannot be: "the *de jure* regulation doesn't matter because it is uniformly ignored" as countries with similar Doing Business values around 200 days have 90th percentiles of Enterprise Surveys responses between one month and one year. The issue seems much more to be selective enforcement that potentially creates a hugely tilted playing field if the "fast" firms are "favored" firms and conversely the "slow" are "disfavored" based on identifiable characteristics (like political connections, family ties, influence activities, or corruption).

Conclusion

In *Dragnet*, the old television show from the 1950s and 1960s, Sergeant Joe Friday of the Los Angeles Police Department always used to ask for "just the facts." So far, this paper is one that Joe Friday could love. That said, Joe Friday, like the rest of us, was ultimately interested in a "theory of the case"—piecing the facts into a coherent

Figure 3

Days to Get a Construction Permit, Comparing Doing Business (DB) Values with the 10th and 90th percentiles of the Enterprises Surveys (ES) for Each Country



Notes: Figure 3 shows the 10th percentile "fast" firms (circles) and the 90th percentile "slow" firms (triangles) in terms of ES values for each country against DB values. So each country has two points that are vertically aligned (as each country/year pair has the same DB value for both the 10th and 90th percentiles of ES).

narrative that explains events. The facts we have presented fit into four broad literatures related to policy and its impact on firm and economic performance.

First, although ours is the first attempt to examine the differences between the Doing Business estimates of compliance times with the actual distribution of experienced times reported by firms in Enterprise Surveys, our findings build on earlier literatures about the heterogeneity of regulatory compliance in specific countries and sectors. It is commonly observed that policy implementation often deviates from the stated policy in firm-specific ways, but this hypothesis has not been easy to document. It appears that when strict rules meet weak state capability—or, more broadly, "institutions"—the rules bend and become more like individuated "deals" where outcomes are not the result of a neutral application of policy to the facts but rather have to be negotiated case by case.

Second, our findings are also related to the large literature on corruption and its relationship to firm profitability, regulatory compliance, and regulations themselves—and hence to the literature on economic performance (for example, Shleifer and Vishny 1993). Fisman (2001) uses the connection between variations in the stock market value of firms in Indonesia, their connections to President Suharto, and information about Suharto's health to show that often a significant portion of connected firms' market valuation was due to their political connections. Research on the transition in post-communist countries at a time when "institutions" were in flux showed that firms used their profits to block future reform (Hellmann 1998), to create "crony bias" through an "inequality of influence" that deterred other firms from relying on or engaging with state institutions (Hellmann and Kaufmann 2002), and even to "capture" the state to change the laws and regulations themselves to disadvantage rivals (Hellman, Jones, and Kaufmann 2000). Similarly in the aftermath of the Arab Spring, the extent of crony capitalism in Egypt (Chekir and Diwan 2012) and Tunisia (Rijkers, Freund, and Nucifora 2014) has been documented. While our paper provides no direct evidence at all of corruption, the very large and firm-specific deviations in reported compliance times and the deviation from the Doing Business estimated compliance times are at least consistent with environments where policy implementation has considerable flexibility and hence generates large potential rents. Freund, Hallward-Driemeier, and Rijkers (2015), however, show that a simple view of deals or bribes as "grease in the wheels" does not hold; officials can also use delays to extract payments, including payments from larger and more-productive firms.

Third, economists have faced a difficult puzzle in explaining the differential response of economic growth to policy reforms of the 1980s and 1990s. Some countries boomed after modest reforms, while others stagnated even after massive reforms (World Bank 2005). More broadly, economists have struggled to explain the observed pattern of growth accelerations and decelerations across countries (Hausmann, Pritchett, and Rodrik 2005; Jones and Olken 2008). At least one hypothesis for explaining the differential responses to "policy reform" may be that when *de jure* and *de facto* policy diverge, the impact of *de jure* reform might have wildly different effects (Pritchett 2005). Estimates of the onset of large growth accelerations suggest that improvements in the "closed ordered deal" environment—not the improvement in "rules" or "reform," or the improvements in "institutions"—may be responsible for growth accelerations (Kar, Pritchett, Raihan, and Sen 2013).

Given our evidence, it is a completely open question how reforms that altered the Doing Business indicators will actually affect the investment climate that most firms actually experience. In Hallward-Driemeier, Khun-Jush, and Pritchett (2010), we showed that firm performance was affected by measures of the *variability* of the policy implementation they faced, more so than the *level*. From this perspective, one can imagine that initiatives that have minimal impact on *de jure* policy but which signal a decisive shift in policy implementation might have substantial impacts on investor expectations and initiate an acceleration of growth. Many argue that something like this happened in China in 1978 after Deng's announcements, or in India

either with a shift to "pro-business" attitudes in the 1980s (Rodrik and Subramanian 2004) or in the 1990s with decisively announced but gradualist implementation.

Finally, our evidence speaks to the emerging debates about "industrial policy" and its role in spurring "structural transformation" (Lin 2009). However, our data speak to this debate in two very different ways.

On the one hand, there is a notion that "industrial policy" is a danger because governments should seek to maintain a "level playing field" and enforce a neutral set of rules that treats all firms, activities, and sectors alike. Whatever merits this argument might have in principle, our evidence strongly suggests that almost nowhere do firms have a level playing field now—both in that there appear to be massive deviations between rules and compliance on average and in the sense that those deviations vary widely across firms. If so, industrial policy would be just another way in which the field is not level rather than tilting a now-level field. Given that much of the economy may be in the hands of "power brokers" who dominate regulated sectors, and given that the incentives of the "private sector" to lobby for better rules may be very weak when they can instead pitch their own deals, some type of preferential action for new entrants and industries may be necessary (Pritchett and Werker 2013).

On the other hand, our evidence could be taken easily to show just how hard it is for governments to pursue discretionary industrial policy. With weak institutions, the risks of available discretion being abused for rent-seeking and directly unproductive activities are very real, as in the classic critique of Krueger (1974). There is nothing inherently contradictory about believing that industrial policy, if one could implement it well, would accelerate growth but also believing that most countries, and especially those that most need growth, lack the wherewithal for policy implementation. This possibility then raises the question of whether or under what conditions it is possible to devise "institutionally robust" industrial policies that can be implemented even when overall institutions are weak (Hausmann and Rodrik 2006; Rodrik 2008, 2009).

References

Almeida, Rita, and Pedro Carneiro. 2009. "Enforcement of Labor Regulation and Firm Size." *Journal of Comparative Economics* 37(1): 28–46.

Almeida, Rita, and Ana Margarida Fernandes. 2008. "Openness and Technological Innovations in Developing Countries: Evidence from Firm-Level Surveys." *Journal of Development Studies* 44(5): 701–27.

Aterido, Reyes, Mary Hallward-Driemeier, and Carmen Pagés. 2011. "Big Constraints to Small Firms' Growth? Business Environment and Employment Growth across Firms." *Economic Development and Cultural Change* 59(3): 609–647.

Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic. 2008. "How Important Are Financing Constraints? The Role of Finance in

the Business Environment." *World Bank Economic Review* 22(3): 483–516.

Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic. 2010. "Are Innovating Firms Victims or Perpetrators? Tax Evasion, Bribe Payments, and the Role of External Finance in Developing Countries." Policy Research Working Paper 5389, World Bank.

Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster. 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6(2–3): 487–500.

Beck, Thorsten, Asli Demirgüç-Kunt, and Vojislav Maksimovic. 2005. "Financial and Legal Constraints to Growth: Does Firm Size Matter?" *Journal of Finance* 60(1): 131–77.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. 2007. "Obtaining a Driving License in India: An Experimental Approach to Studying Corruption." *Quarterly Journal of Economics* 122(4): 1639–76.

Bigsten, Arne, and Måns Söderbom. 2006. "What Have We Learned from a Decade of Manufacturing Enterprise Surveys in Africa?" *World Bank Research Observer* 21(2): 241–65.

Botero, Juan C., Simeon Djankov, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "The Regulation of Labor." *Quarterly Journal of Economics* 119(4): 1339–82.

Bruhn, Miriam. 2011. "License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico." *Review of Economics and Statistics* 93(1): 382–86.

Cai, Hongbin, Hanming Fang, and Lixin Colin Xu. 2011. "Eat, Drink, Firms, Government: An Investigation of Corruption from the Entertainment and Travel Costs of Chinese Firms." *Journal of Law and Economics* 54(1): 55–78.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal. 2014. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." NBER Working Paper 20624.

Chekir, Hamoudi, and Ishac Diwan. 2012. "Crony Capitalism in Egypt." CID Working Paper 250, Harvard Kennedy School, Harvard University.

Chong, Alberto, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2014. "Letter Grading Government Efficiency." *Journal of the European Economic Association* 12(2): 277–99.

Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak, and Brian Chan. 2012. "In Urban and Rural India, a Standardized Patient Study Showed Low Levels of Provider Training and Huge Quality Gaps." *Health Affairs* 31(12): 2774–84.

De Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World*. London: I.B. Tauris.

Dhaliwal, Iqbal, and Rema Hanna. 2014. "Dealing with the Devil: The Successes and Limitations of Bureaucratic Reform in India." NBER Working Paper 20482.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *Quarterly Journal of Economics* 117(1): 1–37.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "Courts." *Quarterly Journal of Economics* 118(2): 453–517.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. "Private Credit in 129 Countries." *Journal of Financial Economics* 84(2): 299–329.

Doing Business. No date. A database. Launched in 2002, the Doing Business project "provides objective measures of business regulations and their enforcement across 189 economies and selected cities at the subnational and regional level." World Bank. www.doingbusiness.org.

Dollar, David, Mary Hallward-Driemeier, and Taye Mengistae. 2005. "Investment Climate and Firm Performance in Developing Countries." *Economic Development and Cultural Change* 54(1): 1–31.

Dollar, David, Mary Hallward-Driemeier, and Taye Mengistae. 2006. "Investment Climate and International Integration." *World Development* 34(9): 1498–1516.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicolas Ryan. 2013. "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128(4): 1499–1545.

Eifert, Benn. 2009. "Do Regulatory Reforms Stimulate Investment and Growth? Evidence from the Doing Business Data, 2003–07." Working Paper 159, Center for Global Development.

Eifert, Benn, Alan Gelb, and Vijaya Ramachandran. 2008. "The Cost of Doing Business in Africa: Evidence from Enterprise Survey Data." *World Development* 36(9): 1531–46.

Enterprise Surveys. No date. A database. World Bank. www.enterprisesurveys.org.

Fernandes, Ana M., and Ariel Pakes. 2008. "Factor Utilization in Indian Manufacturing: A Look at the World Bank Investment Climate Surveys Data." NBER Working Paper 14178.

Fisman, Raymond. 2001. "Estimating the Value of Political Connections." *American Economic Review* 91(4): 1095–1102.

Fisman, Raymond, and Roberta Gatti. 2006. "Bargaining for Bribes: The Role of Institutions." CEPR Working Paper 5712.

Fisman, Raymond, and Jakob Svensson. 2007. "Are Corruption and Taxation Really Harmful to Growth? Firm Level Evidence." *Journal of Development Economics* 83(1): 63–75.

Fisman, Raymond, and Shang-Jin Wei. 2004. "Tax Rates and Tax Evasion: Evidence from 'Missing Imports' in China." *Journal of Political Economy* 112(2): 417–500.

Freund, Caroline, and Bineswaree Bolaky. 2008. "Trade, Regulations, and Income." *Journal of Development Economics* 87(2): 309–321.

Freund, Caroline, Mary Hallward-Driemeier, and Bob Rijkers. 2015. "Deals and Delays: Firm-Level Evidence on Corruption and Policy Implementation Times." *World Bank Economic Review*, published online May 4, 2015, http://wber.oxfordjournals.org/content/early/2015/05/04/wber.lhv001.

Gauthier, Bernard, and Mark Gersovitz. 1997. "Revenue Erosion through Exemption and Evasion in Cameroon, 1993." *Journal of Public Economics* 64(3): 407–24.

Hallward-Driemeier, Mary, Gita Khun-Jush, and Lant Pritchett. 2010. "Deals versus Rules: Policy Implementation Uncertainty and Why Firms Hate it?" NBER Working Paper 16001.

Hallward-Driemeier, Mary, and Lant Pritchett. 2011. "How Business is Done and The 'Doing Business' Indicators: The Investment Climate When Firms Have Climate Control." Policy Research Working Paper 5563, World Bank.

Hallward-Driemeier, Mary, Scott Wallsten, and Lixin Colin Xu. 2006. "Ownership, Investment Climate and Firm Performance: Evidence from Chinese Firms." *Economics of Transition* 14(4): 629–47.

Haltiwanger, John, and Helena Schweiger. 2005. "Allocative Efficiency and the Business Climate." Unpublished paper, University of Maryland.

Hausmann, Ricardo, Lant Pritchett, and Dani Rodrik. 2005. "Growth Accelerations." *Journal of Economic Growth* 10(4): 303–329.

Hausmann, Ricardo, and Dani Rodrik. 2006. "Doomed to Choose: Industrial Policy as Predicament." September 2. https://www.sss.ias.edu/files/pdfs/Rodrik/Research/doomed-to-choose.pdf.

Hellman, Joel S. 1998. "Winners Take All: The Politics of Partial Reform in Post-Communist Transitions." *World Politics* 50(2): 203–34.

Hellman, Joel S., Geraint Jones, and Daniel Kaufmann. 2000. "Seize the State, Seize the Day: State Capture, Corruption, and Influence in Transition." Policy Research Working Paper 2444, World Bank.

Hellman, Joel S., and Daniel Kaufmann. 2002. "The Inequality of Influence." Available at SSRN: http://dx.doi.org/10.2139/ssrn.386901.

Jones, Benjamin F., and Ben A. Olken. 2008. "The Anatomy of Stop-Start Growth." *Review of Economics and Statistics* 90(3): 582–587.

Kar, Sabayachi, Lant Pritchett, Salim Raihan, and Kunal Sen. 2013. "Looking for a Break: Identifying Transitions in Growth Regimes." *Journal of Macroeconomics* 38(Part B): 151–66.

Klapper, Leora F., Luc Laeven, and Raghuram G. Rajan. 2006. "Entry Regulations as a Barrier to Entrepreneurship." *Journal of Financial Economics* 82(3): 591–629.

La Porta, Rafael, and Andrei Shleifer. 2008. "The Unofficial Economy and Economic Development." *Brookings Papers on Economic Activity,* no. 2, pp. 275–352.

Lin, Justin Yifu. 2009. *Economic Development and Transition: Thought, Strategy and Viability.* Cambridge University Press.

Loayza, Norman V., Ana María Oviedo, and Luis Servén. 2006. "The Impact of Regulation on Growth and Informality: Cross-Country Evidence." In *Linking the Formal and Informal Economy: Concepts and Policies*, edited by Guha-Khasnobis, Basuded, Ravi Kanbur, and Elinor Ostrom. Oxford University Press.

Micco, Alejandro, and Carmen Pagés. 2006. "The Economic Effects of Employment Protection: Evidence from International Industry-Level Data." Discussion Paper 2433. Institute for the Study of Labor (IGA), Bonn, Germany.

Pierre, Gaelle, and Stefano Scarpetta. 2006. "Employment Protection: Do Firms' Perceptions Match with Legislation?" *Economics Letters* 90(3): 328–34.

Pritchett, Lant. 2005. "Reform is Like a Box of Chocolates: An Interpretive Essay on the Pleasant and Unpleasant Surprises from Policy Reform." http://www.hks.harvard.edu/fs/lpritch/Economic%20Growth%20-%20docs/Economic%20Growth%20NEW/reform%20is%20like%20a%20box%20of%20chocolates.pdf.

Pritchett, Lant, and Geeta Sethi. 1994. "Tariff Rates, Tariff Revenue, and Tariff Reform: Some New Facts." *World Bank Economic Review* 8(1): 1–16.

Pritchett, Lant, and Eric D. Werker. 2012. "Developing the Guts of a GUT (Grand Unified Theory): Elite Commitment and Inclusive Growth." ESID Working Paper 16/12.

Rijkers, Bob, Caroline Freund, and Antonio Nucifora. 2014. "All in the Family: State Capture in Tunisia." Policy Research Working Paper 6810, World Bank.

Rodrik, Dani. 2008. "Second-Best Institutions." NBER Working Paper 14050.

Rodrik, Dani. 2009. "The New Development Economics: We Shall Experiment, But How Shall We Learn?" In *Thinking Big and Thinking Small,*

edited by Jessica Cohen and William Easterly. Brookings Institution.

Rodrik, Dani, and Arvind Subramanian. 2004. "From 'Hindu Growth' to Productivity Surge: The Mystery of the Indian Growth Transition." NBER Working Paper 10376.

Shleifer, Andrei, and Robert W. Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108(3): 599–617.

Stone, Andrew, Brian Levy, and Ricardo Paredes. 1996. "Public Institutions and Private Transactions: A Comparative Analysis of the Legal and Regulatory Environment for Business Transactions in Brazil and Chile." In *Empirical Studies in Institutional Change*, edited by Lee J. Alston, Thráinn Eggertsson and Douglass C. North. New York: Cambridge University Press, 95–128.

World Bank. 2005. *Economic Growth in the 1990s: Learning from a Decade of Reform*. Washington DC: World Bank.

Xu, Lixin Colin. 2010. "The Effects of Business Environments on Development: Surveying New Firm-Level Evidence." Policy Research Working Paper 5402, World Bank.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 141–164

The Microeconomic Dimensions of the Eurozone Crisis and Why European Politics Cannot Solve Them[†]

Christian Thimann

The academic and policy debate about the crisis in Europe's single currency area is usually dominated by macroeconomic and public sector considerations. It mostly focuses on such matters as the degree of fiscal retrenchment or "austerity" in the eurozone, the monetary policy decisions of the European Central Bank (ECB), the appropriate treatment of public debt, or the merits of public investment schemes. The microeconomic dimensions of the crisis and the private-sector issues typically get much less attention. However, the crisis will only be over when unemployment falls to socially sustainable levels in the most highly stressed economies (including Cyprus, Greece, Ireland, Portugal, and Spain) and in the larger and virtually stagnant economies of France and Italy. Programs of fiscal and monetary loosening or an array of bilateral support loans will not in and of themselves create jobs in the stressed countries on a sustainable basis or avoid Europe's possible "secular stagnation." It is the private sector hiring choices of domestic and foreign firms that will ultimately be decisive.

So what is currently discouraging domestic firms in the stressed countries from investing more and creating jobs? What is preventing firms abroad from increasing foreign direct investment in these markets?

This paper argues there are two main problems holding back private sector employment creation in the stressed eurozone countries. First, there is a persistent

■ *Christian Thimann is a Member of the Executive Committee of the AXA Group, and affiliated professor at the Paris School of Economics, both in Paris, France. He is an external member of the Council of Economic Advisers to the French Prime Minister. He was Director General and Adviser to the President at the European Central Bank (ECB) from 2008 to 2013. His email address is christian.thimann@psemail.eu.*

[†] To access the disclosure statement, visit
http://dx.doi.org/10.1257/jep.29.3.141

doi=10.1257/jep.29.3.141

competitiveness problem due to high labor costs relative to underlying productivity. Over the first ten years of the euro, wage developments relative to productivity diverged strongly across the eurozone.[1] Between 1999 and 2008, Austria, Belgium, Finland, Germany, Luxembourg, and the Netherlands lifted their real productivity on average by 12 percent and nominal wages by 22 percent. France, Greece, Ireland, Italy, Portugal, and Spain collectively lifted their real productivity by only 7 percent, but boosted nominal wages by 40 percent. Hence, the latter group of countries had become both less productive *and* more expensive compared with the former group. This cost-productivity disadvantage of 20–25 percent has been corrected only partially, and a substantial competitiveness problem remains. This problem is aggravated by high tax and social insurance burdens for companies, especially in France and Italy. These developments have led to a considerable erosion of firms' margins, which discourages investment.

Second, widespread structural barriers make job creation in these countries far more arduous than in many other advanced economies, and even more arduous than in some key emerging economies and formerly planned economies. Structural barriers to private sector development are particularly widespread in the areas of labor market functioning, goods market functioning, and government regulation. Evidence from the World Economic Forum's Global Competitiveness Index and the World Bank's "Doing Business" dataset confirms the immense size and persistence of these barriers, despite improvements in some countries in recent years.

Many prominent economists have worked on these topics,[2] but has this research fed into the eurozone policy debates? The answer may seem to be "yes," but substantively it is "no." Yes, because references to structural reforms abound: the European Commission has devoted countless workstreams, committees, and reports to the topic; the monthly monetary policy decisions of the European Central Bank have for more than a decade been accompanied by a call for structural reform; and there is hardly a speech by a leading European policymaker that does not include a reference to structural reforms. But despite abundant declarations, actual reforms fall far short of what is needed.

[1] The 19 eurozone countries, ranked by economic weight in descending order, are: Germany, France, Italy, Spain, the Netherlands, Belgium, Austria, Finland, Ireland, Greece, Portugal, Slovakia, Luxembourg, Slovenia, Lithuania, Latvia, Estonia, Cyprus, and Malta. The European Union includes 28 countries: those of the eurozone plus United Kingdom, Sweden, Poland, Denmark, Czech Republic, Romania, Hungary, Croatia, and Bulgaria.

[2] For example, Acemoglu, Johnson, Robinson, and Thaicharoen (2002) and Rodrik, Subramanian, and Trebbi (2004) investigate the fundamental role of institutional quality for economic outcomes; Benassy-Quéré, Coupet, and Mayer (2007) show how structural impediments hinder foreign direct investment; Botero, Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2004) and Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2008) focus on driving forces and implications of labor regulation; Djankov, McLiesh, and Shleifer (2007) focus on credit availability; Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2002) focus on the regulation of market entry; and Djankov, Ganser, McLiesh, Ramalho, and Shleifer (2010) focus on the effect of taxation on investment.

There are standard economic arguments as to why structural economic reforms affecting labor and product markets can be difficult: in particular, organized special interests often seek to block such changes. However, in Europe, where structural reforms are even more crucial given the absence of an exchange rate adjustment, they are actually *more difficult*. This is because countries joining the European monetary union were promised that their membership would not affect their social models and because the well-intended movement of economic policy discussions to the highest political level in Europe (observed increasingly in recent years) creates a "European political overlay" that actually distracts from national reform.

More specifically, the main focus of European economic integration has been the single currency and a mixture of rules about the "four freedoms" of movement across European borders: of people, goods, services, and capital. The European project is explicitly based on respect for national diversity, and in particular the notion that each country retains authority over its own labor and social policies. Any suggestions about European processes to reform these policies would be seen as a challenge to the national identities of economic models that have always been excluded from the goal of European harmonization.

Equally important, and related to the financial crisis as well as institutional changes, the level of economic policy discussions in Europe has increasingly moved from technocrats to elected politicians, up to the level of heads of state. Regular meetings of heads of state have been established some years ago, and the EU/IMF-sponsored adjustment programs are discussed at the level of finance ministers in the eurozone, compared to a more technocratic level at the IMF. In a context where the main economic and social policy issues are in the national domain, the increasing shift of policy discussions to the European political level is actually counterproductive. International policy discussions at the European political level are appropriate for dealing with common macroeconomic policy issues, but they are not appropriate for dealing with the microeconomic, structural issues of individual countries. Topics such as labor market regulations, judicial reform, or administrative modernization involve highly complex legal and institutional issues about which politicians from other countries will have insufficient information; and foreign politicians, even in the European context, do not have legitimacy to deliberate on reforms that deeply interfere with the economic and social fabric of another country. Whenever they do so, they expose themselves and the European project to political backlash.

This paper presents a range of evidence on these problems, including wage and productivity developments across countries and over time, examples of tax and social insurance wedges, and a wide array of structural indicators to establish the hypothesis that cost disadvantages, public charges, and structural barriers hold back investment, job creation, and growth. The paper also presents a novel explanation for the difficulty of structural reforms in the eurozone by tracing the challenge to the current trend to "Europeanize" and "politicize" economic reform discussions in national policy fields where "Europe" is not a legitimate actor and the European political level is not effective.

Competitiveness and Productivity

Competitiveness is often seen as a vague concept. For example, Krugman (1994) offered a powerful reminder that competitiveness can be "a dangerous obsession" when misused as a basis for national mercantilism, and that it is firms and not nations that compete internationally. But competitiveness is well-defined when understood as the ability of firms to sell goods and services profitably in an open economy and to sustain market shares, domestically or abroad. Krugman emphasized that one was "fully justified to speak about competitiveness" when asking "whether a country's exports and import-competing industries have low enough costs to sell products and services in competition with rivals in other countries."

Competitiveness is often used interchangeably in policy debates with productivity, although the two concepts are distinct. For example, the exposition of a European Commission (2014) report on competitiveness jumps straight to productivity: "Promoting productivity growth is crucial to improving competitiveness. However, in recent years EU productivity growth has been low compared to the US. To enhance competitiveness, the EU must use more resources and in better ways." This exposition is at best incomplete and at worst misleading, because it overlooks the price and cost dimension that is at the heart of competitiveness. High-productivity economies can be uncompetitive when wages and other costs are too high, as in Germany in 2000. And low-productivity economies can be highly competitive when wages and/or the exchange rate are sufficiently low and therefore boost exports, as is the case for some developing countries.

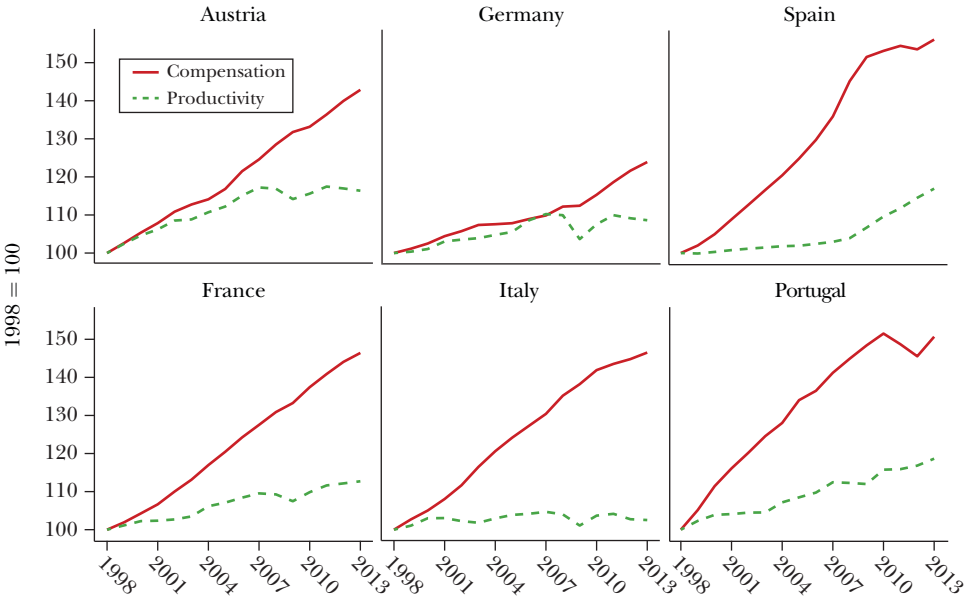
The essence of competitiveness is a comparison of relative costs and relative productivity. Improving competitiveness means improving the price or cost structure of a firm or an economy relative to trading partners; improving productivity means to augment the level of output for a given level of inputs.

The two concepts are related in that competitiveness improves if, for a given level of productivity, costs (and hence prices) can be lowered, or if productivity rises while costs remain unchanged. However, the latter phenomenon is rare because rising (labor) productivity generally triggers demands for higher wages. Competitiveness comparisons usually focus on developments in wages and labor productivity, as cross-country differences in the cost of capital are relatively muted across developed countries.

Competitiveness Problems within the Eurozone

How have eurozone countries fared in terms of competitiveness? Wage and productivity developments in the main eurozone countries are shown in Figure 1. Differences in growth of labor productivity are relevant, but the more striking differences are in wage developments. Austria and Germany are examples of "surplus countries," meaning they have consistently run current account surpluses during this time period. Spain, France, Italy, and Portugal are prominent examples of "deficit countries"—they have consistently run current account trade deficits.

Figure 1
Wage and Productivity Developments, 1998–2013: Selected Countries



Source: European Commission AMECO database.
Notes: 1998 = 100. Productivity is defined as real GDP per employee. Wage growth is in nominal terms, which together with productivity gives nominal unit labor costs, the most widely used competitiveness indicator (see also Draghi 2013). Greece is not shown in the chart because, while the productivity increase is broadly comparable to that of Portugal, the wage growth was even steeper, rising by 2008 to 180 percent of the 1998 value, hence exceeding the scale of the countries shown; wages have declined by about 20 percent since the crisis to 160 percent.

Why did wages surge in the deficit countries? Three reasons seem important: First, wage growth was fueled by a certain growth illusion, stemming from rapid employment growth that had different causes across countries: for example, real estate bubbles as in Ireland or Spain; a rise in hidden government expenditures such as in Greece; or in Spain a surge in immigration between 2000 and 2010, which boosted employment by 10 percent.

Second, public sector wages rushed forward at an astounding pace in a number of these countries. Over the first ten years of the euro, public wages grew by 40 percent in the eurozone as a whole and by 30 percent in Germany. But public sector wages rose by 50 percent in France, 60 percent in Italy, 80 percent in Spain, 110 percent in Greece, and 120 percent in Ireland. Holm-Hadulla, Kamath, Lamo, Pérez, and Schuknecht (2010) provide an overview of public sector compensation trends in the first ten years of the euro, while Lamo Pérez, and Schuknecht (2008) offer a theoretical and empirical analysis of spillovers between public and private sector wages in Europe. This boost in public sector compensation went largely unnoticed at the time, because sharply falling interest rates upon entry into monetary union lowered

the debt-carrying burden and made room in national budgets, but it put pressure on private sector wages to rise as well.

Third, higher domestic inflation rates in some countries made nominal wage increases of 6–7 percent, as seen in many deficit countries, seem justified, because the resulting real wage growth was thought to be in line with productivity growth. However, using domestic prices as the deflator was a grave error, because within a monetary union, real wages deflated with domestic prices are not the relevant benchmark. International trade denominated in euros takes place at a single price level that is largely independent of price developments in individual countries. What matters for competitiveness are nominal wages relative to that internationally determined price level. At the same time, business associations and unions in Germany established a period of wage restraint to correct the overvaluation of its exchange rate from the post-unification boom, when it entered the euro. Not even ten years into the euro, wage growth in some European countries was so strong that even *absolute* wage levels exceeded those in Germany: for example, labor costs per hour in 2008 were higher in Belgium, France, the Netherlands, and Ireland than in Germany, according to Eurostat data.

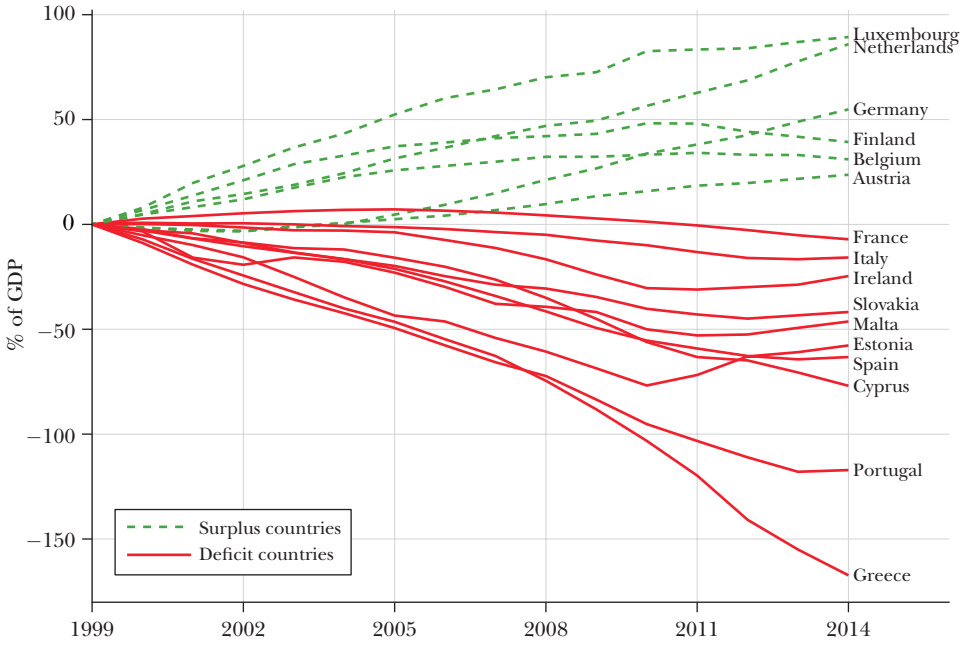
The divergence between relative wages and relative productivity was noted by some observers at the time, but European leaders later admitted that they had not fully realized the importance of the competitive divergences. For example, German chancellor Angela Merkel (2012) noted: "The differences in the competitiveness of the member states of the eurozone have increased, not decreased. We need only look at the development of unit labor costs. Jean-Claude Trichet as ECB president has often pointed this out to us. But all too often, it fell on deaf ears" (see also Trichet 2004).

This shift in competitiveness has had various consequences. For example, some countries began to run systematic current account surpluses, while others began to run systematic current account deficits, as shown in Figure 2. In the financial crisis, a large part of the deficits ended up being financed through central bank money (by the ECB and the national central banks) such that the deficits became a public-sector issue (Cour-Thimann 2013).

These macroeconomic divergences were also reflected at the microeconomic level in an erosion of firms' profitability in deficit countries. Back in 2000, the ratio of gross operating surplus (revenues net of labor costs) to value-added was about 38 percent across the eurozone, and the average was the same for the group of countries that would become deficit countries as for the group that would become surplus countries. But as Figure 3 shows, the gross operating ratio then diverged sharply between these two groups of countries, boosting the incentive for firms to create jobs in one set of countries and reducing it in the other.

Since the economic crisis began in late 2007, nominal wage growth flattened in Spain, while productivity per employee rose, mainly through labor shedding in which less-productive jobs are typically cut first. Hence, Spain apparently regained competitiveness, with the improvement in the current account balance being based almost exclusively in export growth. Yet if one looks at the competitiveness gains

Figure 2
Cumulated Current Account Balances in the Eurozone since the Start of the Euro
(in percent of GDP)



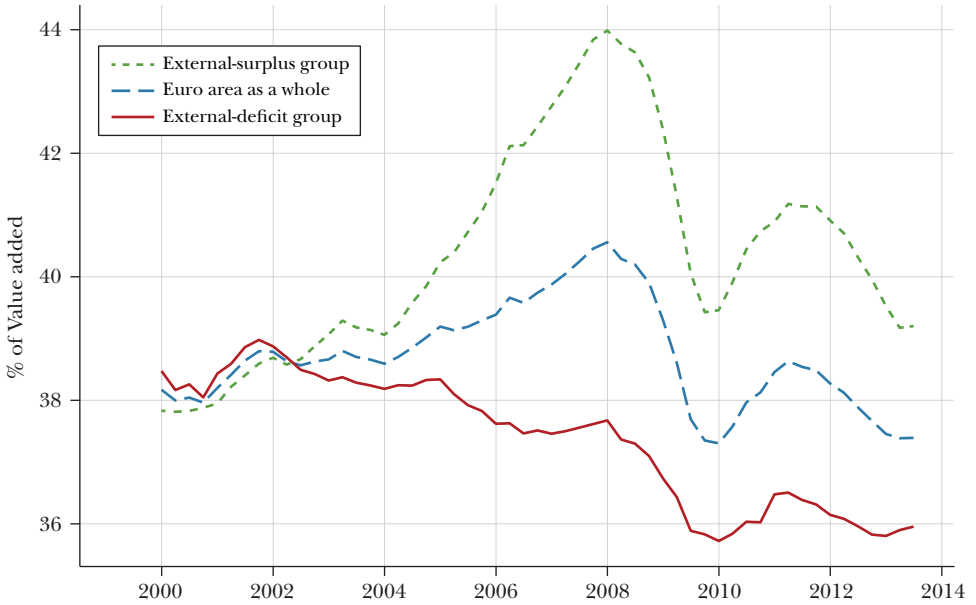
Source: European Commission AMECO database.
Note: Annual data.

based on the total labor force rather than on remaining employment, the improvement in Spain's competitiveness is no longer material. Indeed, measured in terms of real GDP per potential worker (that is, employment plus unemployed), productivity would show a decline in the stressed countries (Sinn 2014). Portugal witnessed a similar development but with more volatility due to earlier nominal wage cuts that were later reversed. Greece has also seen some reduction and ultimate elimination of its current account deficit since 2008, but about three-quarters of the change has been due to import contraction. Perhaps most importantly, in the two largest stagnant countries, France and Italy, the path of nominal wage growth has remained virtually unchanged and the competitiveness problem remains. In France's wage development, neither the global financial crisis nor the eurozone crisis appear to have had an impact, and the current account deficit persists.

The Cost of Creating Jobs and Attracting Talent

It is not just the wage/productivity disadvantages embodied in competitive concerns that hold back economic growth and private sector job creation in the stressed countries. The costs imposed on employers and workers by social insurance contributions and taxation regimes add to the disadvantage of several countries.

Figure 3
Gross Operating Surplus after Labor Costs in the Corporate Sector of Different Country Groups
(in percent of value added)



Source: European Commission AMECO database.
Note: Gross operating surplus is defined as revenues net of labor costs, in percent of value-added.

The variation in employer costs and net compensation can be substantial for a sample of eurozone countries consisting of Belgium, France, Germany, Italy, and Spain, as well as some relevant outside references, such as Switzerland and the United Kingdom. For simplicity and purpose of comparison, Table 1 provides information on five benchmark types of positions, with annual gross salaries from €35,000 to €1,000,000, which for illustrative purposes shall be labeled "clerk," "professional," "manager," "executive," and "chief executive officer (CEO)." Such a comparison is directly relevant for the many multinational corporations that operate across many European countries. It is equally relevant for domestic firms in these countries because it affects the price and cost competitiveness for their goods and services in Europe's highly integrated single market.

For example, for a job with a local annual gross salary of €50,000, corresponding broadly to the median compensation in Western Europe and many professional functions, the costs for the employer range from €59,000 in Switzerland and €65,000 in Germany to €74,000 in Italy and Belgium and €77,000–82,000 in France. The costs are €77,000 in France if the position is in the nonfinancial sector and €82,000 if it is in the financial sector because France maintains a special levy on firms in this sector that is a function of the wage level. As a result, for a single position with the

Table 1

Labor Costs across Eurozone Countries for a Given Level of Salary Category (with Switzerland and the United Kingdom for Comparison)

(annual costs for the firm in bold; net wages for the employee in brackets below; in €)

	"Clerk"	*"Professional"*	*"Manager"*	*"Executive"*	*"CEO"*
Gross salary:	**35,000**	**50,000**	**100,000**	**300,000**	**1,000,000**
Belgium	**53,000**	**74,000**	**143,000**	**321,000**	**1,059,000**
	(25,000)	(32,000)	(52,000)	(141,000)	(458,000)
France					
Financial sector	**57,000**	**82,000**	**166,000**	**502,000**	**1,559,000**
Other sectors	**53,000**	**77,000**	**154,000**	**450,000**	**1,375,000**
	(26,000)	(36,000)	(66,000)	(169,000)	(497,000)
Germany	**46,000**	**65,000**	**120,000**	**330,000**	**1,065,000**
	(25,000)	(32,000)	(62,000)	(172,000)	(548,000)
Italy	**53,000**	**74,000**	**143,000**	**419,000**	**1,385,000**
	(24,000)	(32,000)	(57,000)	(155,000)	(501,000)
Spain	**53,000**	**68,000**	**120,000**	**322,000**	**1,022,000**
	(25,000)	(34,000)	(59,500)	(163,500)	(534,500)
Switzerland	**41,000**	**59,000**	**118,000**	**352,000**	**1,137,000**
	(25,000)	(36,000)	(68,000)	(183,000)	(552,000)
United Kingdom	**48,000**	**67,000**	**127,000**	**366,000**	**1,176,000**
	(26,000)	(36,000)	(63,000)	(162,000)	(526,000)

Sources: HR department, AXA Group, and author's compilation.

Notes: The labor costs for the employer (in bold) are calculated by adding to the given gross salary the local social security contributions that the employer has to pay (including in France a specific tax ("taxe sur les salaires") for the financial sector). The net salary for the employee (in brackets) is calculated by subtracting from the given gross salary the local social security that the employee has to pay and the income tax, which is calculated for a married employee without dependents.

median salary in the country, a firm in France has to pay €12,000 more per year, or almost €1,000 each month, than a company in neighboring Germany. If this firm belongs to the financial sector, the difference rises to more than €1,400 per month. Overall, Switzerland and the United Kingdom stand out as attractive; within the eurozone, Germany and Spain are relatively attractive, whereas Belgium, but even more so Italy and especially France, must be seen as relatively unattractive as far as cost of labor is concerned.

Importantly, the differences in net wage income are far less relevant, ranging between €32,000 and €36,000 for all countries (see the column for the "professionals," who receive the median compensation for the region). Hence, when speaking about lowering labor costs to foster job creation, the focus should be on taxation and social insurance charges at the firm level, rather than on lowering compensation net of employer costs.

For higher levels of remuneration, the differences in labor costs across European countries are even more striking. The labor costs are increasing most for

France, as important social insurance contributions such as health insurance do not level off, but remain a fixed proportion of the notional salary. For a gross compensation of €1 million, the employer faces a total cost of pay of about €1.1 million in most countries but almost €1.6 million in France. The difference amounts to the gross salary of almost ten regular employees and might help explain the exodus of many executives and firms that is observed in France.

Some warnings are in order. Part of the pay differences may be explained through differences in national price levels, even though in a European context the difference is likely to be significant only for Switzerland; the price level in Germany is about the same as in Italy and actually somewhat lower than in France. Other parts of the differences in social insurance contributions may stem from differences in deferred compensations or pensions and different qualities of social insurance services. But the possible differences in the quality of social support systems are likely to be small compared to these cost differences and may not necessarily favor France and Italy compared to other European countries that also have highly developed social systems.

Structural Impediments to Growth

Business operations can be arduous and expensive in some of the eurozone economies. Indeed, there is a remarkable correspondence between the eurozone countries that have significant structural impediments to business growth and the countries that find themselves under economic stress—and hence a correspondence between microeconomic conditions and macroeconomic performance. The strongest impediments to growth appear again in the labor market, exactly where the cost competitiveness problems lie.

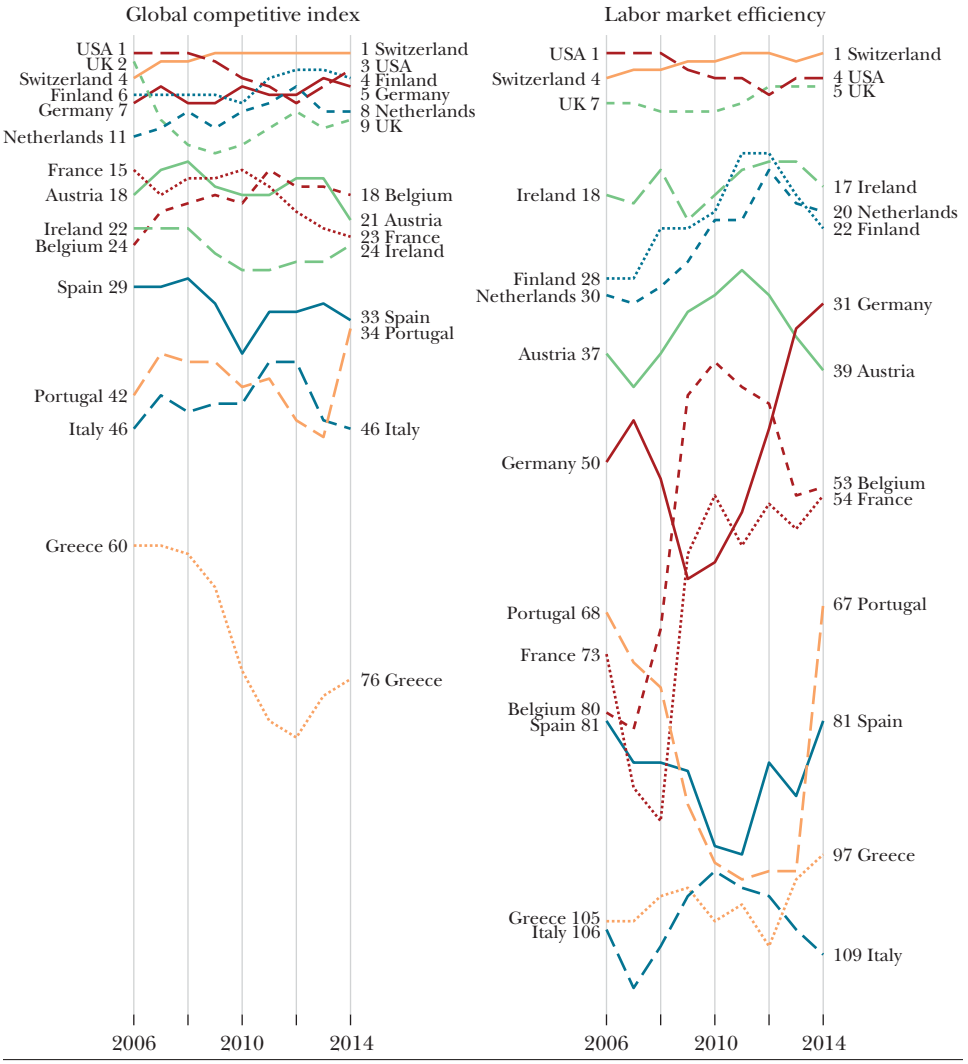
Consider the largest available set of structural competitiveness measures, the Global Competitiveness Indicators that the World Economic Forum has been compiling for over 30 years.[3] The average of eurozone countries (not weighted by GDP) would be ranked 33 in the world, which is not impressive for a single currency area composed of advanced economies founded on trade and free mobility of goods, labor, capital, and services. The European countries in the global top ten for competitiveness indicators—Switzerland, Finland, Germany, Sweden, the Netherlands, and the United Kingdom, in that order—do include some eurozone countries, as shown in the left-hand panel of Figure 4. Some other eurozone countries—Austria, Belgium, France, Ireland, and Luxembourg—are found among the top 30. But Spain ranks significantly lower, Italy and Portugal are placed around rank 40, and Greece ranks 76 out of 148 countries in this worldwide series, with virtually no advanced economy ranked lower. The overall pattern is that eurozone countries usually perform rather well by world standards on the more

[3] These indicators started with a limited number of countries and have been continuously adjusted to take account of newly available data and the latest understanding of factors driving economic performance.

Figure 4

The World Rankings of Eurozone Countries in Structural Competitiveness

(Switzerland, the United Kingdom, and the United States are included for comparison)



Sources: World Economic Forum, Global Competitiveness Indicators, and author's compilation.
Note: This graph shows the ranking over time of eurozone countries within the global sample in terms of overall competitiveness (left panel) and labor market functioning (right panel).

macroeconomic indicators—stability of the macroeconomic environment, and so on—but are weaker than one would expect based on their per capita GDP in the microeconomic dimensions.

"Simple level accounting," also called "development accounting," has been used as a tool to explore the sources of income and productivity differences across

countries. The literature has distinguished between differences in factor endowments and differences in the efficiency with which these factors are used. The overwhelming evidence points to differences in efficiency as the chief explanatory source (Caselli 2005). Structural barriers can be seen mostly as standing in the way of efficient use of factor endowments. These structural barriers include the complexity, cost burden, and uncertainty generated by the regulatory, institutional, and bureaucratic framework in each county, which are important variables not usually captured by economic models.

Among efficiency indicators, eurozone countries rank unfavorably in goods market efficiency and very poorly in labor market efficiency as shown in the right-hand panel of Figure 4. The reasons for the low labor market ranking are *difficult labor-employer relations* (France, Greece, and Italy rank about 130 in this category, compared with Ireland, which ranks 13, Austria 10, and the Netherlands 5), *adverse effects of taxation on incentives to work* (Italy ranks 148, the world's lowest among the countries covered; Belgium and Greece rank about 140), *hiring and dismissal practices* (Belgium, France, and Italy again around rank 140), the country's capacity to *retain talent* (Italy, Portugal, and Spain are particularly weak within the eurozone and rank about 110), and *female participation* in the labor force, where Italy and Greece rank about 100.

In the areas of goods market efficiency, the results are particularly striking for the effect of *taxation on incentives to invest*: here, France, Greece, Italy, and Portugal—which together account for more than one-third of eurozone GDP—rank about 140 in the world, and they rank almost as poorly as far as the *overall tax burden* and the business impact of *rules on foreign direct investment* is concerned.

Many formerly planned economies in central and eastern Europe rank higher than several western European eurozone countries; the Czech Republic ranks 37 and Poland 43, which is well above some longstanding European Union member states. It is remarkable that formerly planned economies have overtaken many western European countries in terms of important aspects of economic environment, and especially in goods and labor market functioning.

These rankings can change relatively quickly if effective structural reforms are undertaken. For example, France's labor market reform of 2008 allowed the country to jump 30 ranks higher in the market efficiency index over the next two years, while Portugal jumped 30 ranks in one year (between 2013 and 2014) due to its important labor market reforms. Greece has improved its ranking throughout, although it remains by these measures the weakest-rated eurozone country.

For another perspective on corporate decision-making on investment and operations, one can consider the data on the "Ease of Doing Business" index established by the World Bank. These data are divided into ten main areas: starting a business, dealing with construction permits, getting electricity, registering property, getting credit, protecting minority investors, paying taxes, trading across borders, enforcing contracts, and resolving insolvency. Each of these main headings is then evaluated in a number of subcategories. There is also a separate section on labor market regulation, with subcategories including difficulty of hiring, rigidity of

Table 2
Ease of Doing Business: The Global Ranking of Selected Eurozone Countries

Rank	*Eurozone country*	*Weaknesses (In italics: selected competitor countries)*
		(1–8 ranks include: Singapore, New Zealand, Denmark, Korea, Norway, United States, United Kingdom)
9	Finland	Protecting minority interests
13	Ireland	Dealing with construction permits
14	Germany	Starting a business, Registering property
		(15–20 ranks include: Canada, Malaysia, Taiwan, Switzerland)
21	Austria	Starting a business, Dealing with construction permits
25	Portugal	Getting credit
27	Netherlands	Dealing with construction permits, Getting electricity, Protecting minority interests
		(29 rank: Japan)
31	France	Dealing with construction permits, Registering property, Paying taxes
33	Spain	Dealing with construction permits, Getting electricity, Paying taxes
42	Belgium	Dealing with construction permits, Getting electricity, Registering property, Getting credit, Paying taxes
56	Italy	Dealing with construction permits, Getting electricity, Getting credit, Paying taxes, Enforcing contracts
61	Greece	Dealing with construction permits, Getting electricity, Getting credit, Enforcing contracts
65	Cyprus	Dealing with construction permits, Getting electricity, Registering property, Getting credit, Enforcing contracts

Source: World Bank, Ease of Doing Business, 2015.
Notes: "Weaknesses" lists those sub-categories in which the country concerned is below world average. Those ranked below 65 are mostly developing countries.

hours, difficulty of redundancy, redundancy costs, and social protection and labor disputes. The components of the index have been subject to in-depth academic research (for example, Botero, Djankov, La Porta, Lopez-de-Silanes, and Shleifer 2004; Djankov, La Porta, Lopez-de-Silanes, and Shleifer 2002, 2008; Djankov, McLiesh, and Shleifer 2007; Djankov, Ganser, McLiesh, Ramalho, and Shleifer 2010). This research has investigated implications of these "doing business" factors on labor regulation, credit availability, market entry regulation, and the effect of taxation on investment. It has found significant impact from such indicators on investment and job creation.

On Table 2, we see that eurozone countries are above the world average as regards trading across borders, starting businesses, and obtaining permits, which is not really a surprise for high-income countries within the European Union. But eurozone countries are below average—sometimes significantly so—as regards other corporate activity, in particular, dealing with administrations (construction permits, paying taxes), utilities (getting electricity), or courts (enforcing contracts). While all eurozone countries have weak areas in which they rank below world average, the number of weaknesses rises significantly for stressed eurozone countries, making

corporate activity also structurally more difficult and providing another reason for lower investment and job creation.

The Political Economy of Structural Reforms in Europe's Monetary Union

Calls for structural reforms are ubiquitous in European policy documents and declarations at both the country and European Union level, but there is a distinct and longstanding lack of achievement. One might think that an EU-style arrangement should be a fertile setting for a program of structural reforms, even if the process is a slow and evolving one. Yet progress toward structural reform in the most deeply afflicted countries, with a few notable exceptions including Ireland and Spain, has been so limited relative to what is at stake that there may well be higher barriers to action deeply rooted in the political economy of implementing Europe's construction.

Why the Eurozone Should Facilitate Structural Reforms

In principle, eurozone participation should facilitate structural reforms because countries have strong economic incentives to engage in them, and because there are multiple formal and informal schemes for devising and coordinating such reforms. The main incentive for structural reforms, especially to increase the flexibility of product and labor markets, comes from the absence of the exchange rate as an adjustment mechanism (Duval and Elmeskov 2005; Bean 1998). Moreover, enhanced competition should make the costs of rigidity higher; the costs of protection for insider firms and workers should become more visible to consumers and voters; firms facing greater cost pressures from competition should exert greater political pressure for deregulation of the market for services, energy, and transport; and countries should have incentives to improve labor market functioning to foster wage adjustments and labor mobility (for an overview of these arguments, see Alesina, Ardagna, and Galasso 2011).

The European project has built vast and far-ranging formal and informal arrangements to "coordinate" economic policies and foster reforms. Formal coordination mechanisms involve precommitments through multi-annual policy plans and public declarations, thus seeking to overcome time inconsistency problems. A range of informal coordination mechanisms seeks to foster an early and politically acceptable process for launching structural reforms (Begg, Hodson, and Maher 2003). The instruments available are plentiful, consisting of guidelines, recommendations, and warnings, based on peer reviews, benchmarking, and a large number of studies and reports provided by all European institutions (Cini 2001).

The question is therefore why, despite these built-in incentives and institutional arrangements, structural reforms have not been up to the challenge.

Why Structural Reform is Slow to Work: Conventional Explanations

The blatant shortcomings of structural reforms have given rise to a large literature explaining why such reforms are slow in general, and within the European

Union and the eurozone in particular (for an overview, see Leiner-Killinger, López Pérez, Stiegert, and Vitale 2007). Several arguments are commonly presented.

One argument is that the case for major structural reforms in Europe is overblown. Bentolila and Saint-Paul (2000) argue that the optimal size of reform may be smaller in Europe's monetary union if, as suggested by Elmeskov, Martin, and Scarpetta (1998) and Orszag and Snower (1998), there are political or economic complementarities across individual reforms. Others argue to the contrary that the case for structural reforms was underestimated for a long time such that now a large amount of catch-up is required with regard to reforms. In this journal, Fernández, Garicano, and Santos (2013) explained how the boom in the run-up to the financial and economic crisis beginning in 2007 relaxed the budget constraint for countries and made reform incentives vanish during that period.

A standard argument, applicable to any country context, is that reform is slow because the long-term benefits of reform are often less visible and more diffuse than the possible short-term costs. Special interest groups focus intensely on these short-term costs, and the schemes often proposed by economists to compensate special interests can be difficult to implement (Rajan 2004; Duval and Elmeskov 2005). For the eurozone, in particular, the upfront costs of reforms may be larger with a single currency because there is no softening certain costs of reform through an exchange rate adjustment (Bean 1998).

Given the recession and persistent joblessness in some European countries over the last seven years, the short-term costs of structural reforms have often become entangled with arguments about contractionary macroeconomic policy. In general, eurozone countries can find it more difficult to conduct expansionary macroeconomic policies given that monetary policy for the currency area is determined with an area-wide focus, and rules across the eurozone limit fiscal policy (Eggertsson, Ferrero, and Raffo 2014; Calmfors 2001; Duval and Elmeskov 2005). However, the lack of reforms means that labor-cost competitiveness problems go unaddressed, making it harder for eurozone countries to participate in the growing global economy.

Finally, the sequencing of reforms is believed by some to pose particular challenges in Europe. Alesina, Ardagna, and Galasso (2011), Blanchard and Giavazzi (2003), and Fiori, Nicoletti, Scarpetta, and Schiantarelli (2007) have pointed to an ideal sequencing in which goods market reforms precede labor market reforms. But within the European Union, some product market regulation is linked to the single market, which can make it harder for individual countries to take actions that would sequence reforms in this way.

Deeper Reasons Why Structural Reform Has Not Occurred

There are deeper reasons for the lack of sufficient structural modernization and reform, which can be understood from the eurozone's specific political economy context. Three issues in particular have to be considered: 1) there are different economic models across eurozone countries, reflecting different social preferences and institutional frameworks, which monetary union was not supposed to change;

2) recognizing this, the European institutional set-up is such that European institutions and fora have little leverage over national economic and social policies; and 3) the European "political overlay" created by taking economic policy discussions up to a higher and cross-country political level can actually interfere with efficient national reform.

1) Diversity of economic, social, and institutional arrangements across European nations. It is well-known that structural reforms can tear at a nation's social fabric. Almost by definition, labor market and social policies relate to deep-rooted preferences for protection, redistribution, flexibility, mobility, and short-term/long-term tradeoffs that often differ from country to country. As a result, there is no "one-size-fits-all" model for social welfare policy, labor market functioning, or the national pension system. Proposals for structural reforms would change institutional arrangements that Shepsle (1986) describes as representing the "frozen preferences" of those who created them. Using the example of Bismarckian welfare systems and their reform over the twentieth century, Palier and Martin (2007) illustrate how difficult and drawn-out such reforms are and what time span, crises, and other disruptions are necessary to alter such systems.

In the context of monetary union, national identities and national models are seen as excluded from the European integration agenda. The social protection of the employee in France, the constitutionally guaranteed tax exemption of ship owners in Greece, the highly federal structure of legal frameworks in Spain or Germany—each reflect not only economic choices, but democratic choices and often historic outcomes. When countries joined monetary union, this was seen as a decision about their national monetary and exchange rate policies in the first place, and about the imposition of constraints on fiscal policy in the second place. Possible implications for economic or even social policies were not discussed rigorously and would have been vociferously refused (Marsh 2011, p. 99f). The public was told that national economic and social identities would be maintained, and diversity was stressed as an essential European characteristic. Fear of social and economic harmonization was one of the reasons for the rejection of a European constitution in 2005 by voters in France and the Netherlands (Jérôme and Vaillant 2005).

The presence and strength of this social fabric also means that possible changes can only come from within the system—that is, they need to come from national governments and other national economic policy actors. Changes can be accelerated in periods of economic disarray or loss of financial market access, but they cannot be imposed by other countries, nor by European institutions.

2) The European institutional set-up. The institutional design of European monetary integration can be viewed as a pyramid with four layers. At the top sits the single currency, with a single institution and a single monetary policy. Just below are fiscal policies that remain national but are closely scrutinized at the European level with formal enforcement mechanisms and the possibility of financial sanctions. Further below are general economic policies, which also remain national but are coordinated much more loosely. Finally, at the bottom of the pyramid are the social policies that are fully national and have hardly any European framework at all, with

European activity being restricted to information, consultation, and "encouragement." The two layers where the microeconomic dimensions of the eurozone crisis lie are therefore overwhelmingly national, with little European leverage.[4]

Because this issue is so fundamental for understanding the difficulties of structural reform in the eurozone, it is worthwhile recalling the exact EU Treaty formulations for economic and social policies. The EU Treaty chapter on "economic policy coordination" starts with a solemn declaration: "The Member States shall regard their economic policies as a matter of common concern and coordinate them."[5] But how is coordination enforced? It turns out that the ultimate "threat" that the European level can exercise is that the EU Council of Ministers adopts a "recommendation" for the country and "may decide to make the recommendation public."

What does a Council recommendation—the ultimate position on economic policies and structural reforms in Europe—look like? It is a short document of usually six to ten pages, the bulk of which is taken up by recitals recalling the legal basis and policy context. The actual recommendations are extremely short and general in nature.

For example, the June 2014 recommendations for Spain from the EU Council take two pages. The recommendations include "promote real wage developments consistent with the objective of creating jobs"; "strengthen the job search requirement in unemployment benefits"; "enhance the effectiveness of active labor market policies"; and "reinforce the coordination between labor market policies and education and training policies." The June 2014 EU Council recommendations for Italy are one-and-a-half pages long.[6] They include generalities such as "work towards a comprehensive social protection for the unemployed, while limiting the use of wage supplementation schemes"; "strengthen the link between active and passive labor market policies"; "improve school outcomes and reduce rates of early school leaving"; and "increase the use of work-based learning and upper secondary vocational education." These recommendations represent at best general policy orientations; they are by no means actionable reform proposals.

By comparison, Germany's own Hartz reforms, which triggered what is widely seen as the most comprehensive labor market reform in Europe in recent decades, are based on a report of 350 pages (Bundesministerium für Arbeit 2002). The Hartz commission was launched in February 2002 by the German government and presented its report in August 2002; the legal transposition was conceptually prepared so that the first laws could be submitted to the German Parliament shortly

[4] There are two policy fields of a more microeconomic nature that are fully coordinated at the European level of the EU (not the eurozone): trade policies and competition policies. These two fields are explicitly identified as EU mandates in the Treaty.

[5] The quotation is from Article 121, available at http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:12012E/TXT.

[6] The recommendations for Spain are at http://ec.europa.eu/europe2020/pdf/csr2014/csr2014_council_spain_en.pdf. The recommendations for Italy are at http://ec.europa.eu/europe2020/pdf/csr2014/csr2014_council_italy_en.pdf.

thereafter and adopted in December 2002, entering into force from January 2003. Such a format is not unusual for national reform commissions and points to the gap between European and national reform discussions. European program countries are faced with much more detailed structural reform agendas, but their implementation is often blurred due to the "political overlay" discussed in more detail below.

On social policies, the European Union legal framework is even looser and essentially focused on information sharing, facilitation, and encouragement: "The Commission shall *encourage* cooperation between the Member States and *facilitate* the coordination of their action in all social policy fields. To this end, the Commission shall *make studies, deliver opinions* and *arrange consultations.*"[7] In an environment that involves difficult political decision-making, the weight of these "encouragements" is virtually nil.

3) The unintended consequences of a political overlay in European economic and monetary union. It has become customary in recent years to elevate negotiations about economic policies and adjustment programs, such as the one for Greece, to the level of elected politicians operating on a European level. Uncountable bilateral visits by politicians across the eurozone have taken place in recent years for this purpose. In many cases, representatives from countries that are receiving adjustment loans have visited "creditor" countries, often giving public speeches and interviews in the home countries of creditors, and even had public meetings with politicians of the opposition.

Escalation from technocratic and national levels to political and European levels was further supported when the Lisbon Treaty in 2009 recognized the European Council of heads of state or government as an official institution of the EU. During the financial crisis, the European Council, sometimes in Eurozone composition, met as frequently as every two months and dealt directly with the adjustment in stressed countries, fiscal targets, and the conditions of financial support including the amount of support loans, interest rates charged, and repayment periods.

While escalation to political and European levels may seem appealing given the broader ambitions of the European project (Habermas 2012), it entails unintended consequences for the efficiency of reform for several reasons. With an escalation to the political level, information asymmetries grow. Finance ministers and heads of state are not familiar with intricate structural, regulatory, administrative, or social insurance issues in other countries. Also, the risk of "negative politization" grows. Well-intended policy proposals can easily be perceived as interference in the domestic issues of other countries. Finally, as the discussion rises to higher political levels, there is a tendency to move from microeconomic issues to macroeconomic issues and from allocation to distribution, as notions of fairness play a particularly important role in politics. Hence, rather than discussing the nitty-gritty of structural reforms, there is a temptation to focus on the "bigger

[7] The quotation is from EU Treaty Article 156, available at http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:12012E/TXT, with italics by the author.

picture" and the "European dimension" of the crisis. For all these reasons, a movement of the debate to higher political levels up to heads of state has gone hand in hand with a move away from national, structural microeconomic issues towards European-level macroeconomic issues.

For example, the documents outlining the Greek adjustment program, as prepared by a team of staff experts from the European Commission, the ECB, and the IMF, are actually highly detailed on structural reforms. They cover measures related to privatization, revenue administration, tackling tax evasion, tax reform, and expenditure control; moreover, the program includes far-reaching measures regarding labor market functioning, regulated professions, energy and transport, product and services markets, and judicial reform (European Commission 2012). However, whenever these documents were brought to the European political level for discussion and approval, the debates surrounding them have almost exclusively focused on the macro issues, in particular the fiscal policy stance and the sustainability of Greek government debt. As these political discussions have become very thorny and drawn out over time, they have frequently led to delays in implementation of the urgent microeconomic reforms. Therefore, the "political overlay" has created a distraction, both at the European level and at the national level, from important microeconomic reforms that are essential to getting the economies of the crisis countries on their feet again.

More generally, in comparison with the IMF, Europe has taken a different approach to the governance of adjustment programs. At the IMF, country programs are discussed and approved by an Executive Board of technocrats sent by national administrations; this is a de-politicized process. In Europe, the country programs for Greece and other stressed countries are approved and discussed by the Eurogroup, which is the body of finance ministers that in virtually all cases are elected politicians, who are of course sensitive to political issues, including their reputation and chances of reelection; this is a political process. This difference has far-reaching implications for the contents of the debate, which in the case of Europe has shifted to macroeconomic issues and become heavily politicized; and this has not been conducive to a smooth implementation and has actually led to political backlash.

The rising politicization has been accompanied by a new version of the European "blame game": no longer "against Brussels," but instead among countries. Monetary union, which was created to avoid "beggar-thy-neighbor" policies, has recently witnessed a lot of "blame-thy-neighbor" politics. The rising national accusations in the press of individual countries that invoke national stereotypes and historical images are ample testimony of this issue.

For example, there is a widespread public perception in countries such as France, Italy, and Spain, nurtured by many political declarations, that Germany's fiscal policy plays an important role in explaining those countries' economic malaise. The fact that this argument rests on weak economic grounds has not obviated its political attractiveness. A French think-tank considered a substantial fiscal impulse in Germany amounting to one percentage point of GDP for three consecutive years,

and estimated that a German fiscal expansion of this size would "unambiguously stimulate German growth, but the spillovers to Spain would be negligible" (Blot, Cochard, Ducoudré, Schweisguth, Timbeau, and Creel 2014, p. 25).

Finally, the "political overlay" has led to a blurring of responsibilities. As it is very difficult for politicians to tell other countries what they should do, the debate has shifted from "country-specific" recommendations to recommendations focused on Europe as a whole. Politically, it is much easier to say what an abstract "Europe" should do than to say what a concrete other country should do.

Conclusion

At the core of the economic crisis in the eurozone is the problem of unemployment in several countries. Roughly 18.2 million people are unemployed in early 2015. In about half the eurozone countries, the unemployment rate is below 10 percent, and in Germany it is actually below 5 percent (Eurostat data, February 2015), but in France, 10.7 percent of the labor force are unemployed; in Italy, 12.7 percent; in Portugal, 14.1 percent; in Spain, 23.2 percent; and in Greece, 26.0 percent. In the latter group of countries, the overall unemployment problem is compounded by large-scale youth unemployment.

It is legitimate to speak about this as a problem *for* the eurozone in the sense that economic policies in a single currency area are truly a matter of common concern, and also because high unemployment interferes with the smooth functioning of the eurozone, challenging its economic and political cohesion. But it is not accurate to attribute responsibility for the problem, or the solution, to the eurozone as a whole, to European institutions, or to other countries.

Jobs fail to be created in a number of these countries not because of a "lack of demand" as often claimed, but mainly because wage costs are high relative to productivity, social insurance and tax burdens are heavy, and the business environment is excessively burdensome. All of this should be viewed not in absolute terms, but in relative terms, compared with other economies in Europe and countries around the world where labor costs and productivity are more advantageous, and the business environment is friendlier.

"Europe" is not an all-powerful actor in the field of national economic policies, but only a potentially useful facilitator. Only the country concerned is the legitimate and able party to improve its own economic functioning in line with its social preferences and economic setup. This is why European politics cannot solve the microeconomic dimensions of the eurozone crisis. Within individual countries, it is the governments, administrative authorities, social partners, and all other economic stakeholders that are the legitimate actors in the field of economic and social policies.

The founders of the European Union always had in mind the overarching goal of lasting peace in Europe. For this reason, Europe's political construction allows for a broad range of different economic models and social preferences, and

leaves responsibility for the consequences of different models with the countries concerned. It is unlikely that a "deepening of integration" to include social or economic policy harmonization is the answer desired by the democracies in Europe. Such a proposal fails to take account of the historic diversity on the European continent. Moreover, voters showed a lack of support for "deepening integration," when they rejected a European constitution in 2005.

For the eurozone countries, their economic and unemployment problems are not primarily a question about some countries versus other countries within the monetary union, but about finding their place in an open global economy—that is, about competing and cooperating successfully with advanced, emerging, and developing economies across the globe. An inward-looking European debate on the distribution of the relative adjustment burden for structural reforms would dramatically overlook the much broader challenges of integration into the global economy.

If Europe wants to focus on microeconomic adjustment and reform, a de-escalation from the political level seems essential. Rather than discussing structural reform among elected politicians, Europe could provide technical advice by subsuming the extensive expertise on structural reform issues existing across European institutions into a "fiscal institute" or "competitiveness institute," perhaps under the umbrella of the European Commission but at arm's length from the political level. In such a context, experts could analyze experiences across countries and provide insights on best practices. It would benefit country governments to take account of any such analysis, which would be provided on a technical rather than a political level.

In addition, academic research could usefully turn more to the microeconomic dimensions of the eurozone crisis: wage determination and labor market reforms; labor costs and productivity developments; regulatory and tax burdens and business operations; and the efficiency of judiciary and administrative systems. Policy advice could focus on promising options for lowering costs for employing labor while not reducing disposable incomes to workers; raising labor force participation rates and part-time work arrangements; and, ultimately, raising long-term productivity growth. It may be more glamorous to focus on European monetary policy, the "European architecture," or the "bigger macro picture." But the real issue of—and solution to—the crisis in the eurozone lies in the mostly *micro*economic trenches of national economic, social, and structural policies.

■ *The author is indebted to Andrei Shleifer for the encouragement to write this paper; to Timothy Taylor, Ulrike Malmendier, and Chang-Tai Hsieh for editorial excellence in commenting and review; to Mario Draghi and Jean-Claude Trichet for numerous discussions of the issues raised in this paper; to Henri de Castries, Frédéric Clément, Philippine Cour-Thimann, Denis Duverne, Amélie de Montchalin, Ann Norman, Philippe de Rougemont, Frank Schorkopf, Romesh Vaitilingam, and Alfons Weichenrieder for valuable comments; and to Eric Persson for outstanding research assistance. Views expressed are those of the author.*

References

Acemoglu, Daron, Simon Johnson, James Robinson, and Yunyong Thaicharoen. 2002. "Institutional Causes, Macroeconomic Symptoms: Volatility, Crises and Growth." NBER Working Paper 9124.

Alesina, Alberto, Silvia Ardagna, and Vincenzo Galasso. 2011. "The Euro and Structural Reforms." *Review of Economics and Institutions*, Universita di Perugia, Dipartimento Economia, Finanza e Statistica, 2(1).

Bean, Charles R. 1998. "The Interaction of Aggregate Demand Policies and Labor Market Reform." *Swedish Economic Policy Review* 5(2): 353–82.

Begg, Iain, Dermot Hodson, and Imelda Maher. 2003. "Economic Policy Coordination in the European Union." *National Institute Economic Review* 183(1): 66–77.

Benassy-Quéré, Agnès, Maylis Coupet, and Thierry Mayer. 2007. "Institutional Determinants of Foreign Direct Investment." *The World Economy* 30(5): 764–782.

Bentolila, Samuel, and Gilles Saint-Paul. 2000. "Will EMU Increase Eurosclerosis?" CEPR Discussion Paper 2423.

Blanchard, Olivier, and Francesco Giavazzi. 2003. "Macroeconomic Effects of Regulation and Deregulation in Goods and Labor Markets." *Quarterly Journal of Economics* 118(3): 879–907.

Blot, Christophe, Marion Cochard, Bruno Ducoudré, Danielle Schweisguth, Xavier Timbeau, and Jérôme Creel. 2014. "Fiscal Consolidation, Public Debt and Output Dynamics in the Euro Area: Lessons from a Simple Model with Time-Varying Fiscal Multipliers." OFCE Working Paper 2014-14.

Botero, Juan C., Simeon Djankov, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "The Regulation of Labor." *Quarterly Journal of Economics* 119(4): 1339–82.

Bundesministerium für Arbeit (Federal Ministry of Labor and Social Affairs). 2002. "Moderne Dienstleistungen am Arbeitsmarkt." Bericht der Hartz-Kommission ("Modern Services on the Labor Market." Report of the Hartz Commission), Berlin, October 2002. http://www.bmas.de/DE/Service/Publikationen/moderne-dienstleistungen-am-arbeitsmarkt.html.

Calmfors, Lars. 2001. "Unemployment, Labor Market Reform, and Monetary Union." *Journal of Labor Economics* 19(2): 265–89.

Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." Chap. 9 in *Handbook of Economic Growth*, edition 1, vol. 1, edited by Philippe Aghion and Steven Durlauf, pp. 679–741. Elsevier.

Cini, M. M. C. 2001. *From Cooperation to Integration: Policy Co-ordination in the European Union*. Kogan Page.

Cour-Thimann, Philippine. 2013. "Target Balances and the Crisis in the Euro Area." *CESifo Forum*, Special issue, April 2013, pp. 5–50.

Djankov, Simeon, Tim Ganser, Caralee McLiesh, Rita Ramalho, and Andrei Shleifer. 2010. "The Effect of Corporate Taxes on Investment and Entrepreneurship." *American Economic Journal: Macroeconomics* 2(3): 31–64.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *Quarterly Journal of Economics* 117(1): 1–37.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Law and Economics of Self-dealing." *Journal of Financial Economics* 88(3): 430–65.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. "Private Credit in 129 Countries." *Journal of Financial Economics* 84(2): 299–329.

Draghi, Mario. 2013. "Euro Area Economic Situation and the Foundations for Growth." Slide presentation at the Euro area Summit, Brussels, March, 14, 2013.

Duval, Romain, and Jørgen Elmeskov. 2005. "The Effects of EMU on Structural Reforms in Labour and Product Markets." OECD Economics Department Working Paper 438.

Eggertsson, Gauti, Andrea Ferrero, and Andrea Raffo. 2014. "Can Structural Reforms Help Europe?" *Journal of Monetary Economics* 61(C): 2–22.

Elmeskov, Jörgen, John P. Martin, and Stefano Scarpetta. 1998. "Key Lessons for Labour Market Reforms: Evidence from OECD Countries' Experience." *Swedish Economic Policy Review* 5(2): 205–252.

European Commission. 2012. *The Second Economic Adjustment Program for Greece*. Occasional Paper no. 94, March.

European Commission. 2014. "Productivity and Competitiveness Developments." Website of Economic and Financial Affairs, European Commission.

Felipe, Jesus, and Utsav Kumar. 2011. "Unit Labor Costs in the Eurozone: The Competitiveness Debate Again." Working Paper Archive 651, Levy Economics Institute.

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.

Fiori, Giuseppe, Giuseppe Nicoletti, Stefano Scarpetta, and Fabio Schiantarelli. 2007. "Employment Outcomes and the Interaction Between Product and Labor Market Deregulation: Are They Substitutes or Complements?" IZA Discussion Paper 2770, Institute for the Study of Labor (IZA).

Habermas, Jürgen. 2012. "The Crisis of the European Union in the Light of a Constitutionalization of International Law." *European Journal of International Law* 23(2): 335-48.

Holm-Hadulla, Fédéric, Kishore Kamath, Ana Lamo, Javier J. Pérez, and Ludger Schuknecht. 2010. "Public Wages in the Euro Area—Towards Securing Stability and Competitiveness." Occasional Paper 112, European Central Bank.

Jérôme, Bruno, and Nicolas G. Vaillant. 2005. "The French Rejection of the European Constitution: An Empirical Analysis." *European Journal of Political Economy* 21(4): 1085–92.

Krugman, Paul. 1994. "Competitiveness: A Dangerous Obsession." *Foreign Affairs* 73(2): 28–44.

Lamo, Ana, Javier Pérez, and Ludger Schuknecht. 2008. "Public and Private Sector Wages: Co-movement and Causality." Working Paper 0963, European Central Bank. Available at SSRN: http://ssrn.com/abstract=1291172.

Leiner-Killinger, Nadine, Víctor López Pérez, Roger Stiegert, and Giovanni Vitale. 2007. "Structural Reforms in EMU and the Role of Monetary Policy: A Survey of the Literature." Occasional Paper no. 66, European Central Bank, July.

Marsh, David. 2011. The Euro: The Battle for the New Global Currency. Yale University Press.

Merkel, Angela. 2012. Speech by Federal Chancellor Angela Merkel in the European Parliament in Brussels, November 7.

Orszag, Mike, and Dennis J. Snower. 1998. "Anatomy of Policy Complementarities." *Swedish Economic Policy Review* 5(2): 303–343.

Palier, Bruno, and Claude Martin. 2007. "From 'A Frozen Landscape' to Structural Reforms: The Sequential Transformation of Bismarckian Welfare Systems." Editorial Introduction, *Social Policy and Administration* 41(6): 535–54.

Rajan, Raghuram. 2004. "Straight Talk: Why Are Structural Reforms So Difficult? The Benefits of Structural Reforms Aren't Always as Obvious as the Smaller, Shorter-Term Costs." *Finance & Development*, IMF, June, pp. 56–57.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9(2): 131–65.

Shepsle, Kenneth. 1986. "Institutional Equilibrium and Equilibrium Institutions." In *The Science of Politics*, edited by H. Weisberg, pp. 51–82. New York: Agathon.

Sinn, Hans-Werner. 2014. *The Euro Trap: On the Bursting of Budgets, Bubbles and Beliefs*. Oxford University Press.

Trichet, Jean-Claude. 2004. "Structural Reforms and Growth, as Highlighted by the Irish Case." Whitaker Lecture, Dublin, May 31.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 165–184

E-books: A Tale of Digital Disruption[†]

Richard J. Gilbert

Electronic books are not new. They have existed at least since the launch of Project Gutenberg in 1971, a volunteer effort to create a digital library of books and cultural media in the public domain (OECD 2012), but they occupied a small niche in the book industry prior to the launch of the Amazon Kindle at the end of 2007. The Kindle offered a comfortable reading experience, wireless connectivity, and access to Amazon's inventory of books. In addition, Amazon's well-established business selling printed books online gave it ready access to a large set of book buyers, and the company promoted the Kindle e-reader and e-books aggressively.

The rapid adoption of e-books since the launch of the Kindle has become a focus of several controversies in the book industry. One issue is the extent to which e-books are substitutes for printed books, and thus whether they are driving brick-and-mortar sellers of printed books out of business. A second issue is the competitive relationship between e-books and e-readers and whether Amazon as an e-retailer has monopoly power as a seller or monopsony power as a buyer, or both.

A third issue involves the conflict between Amazon and publishers, and in particular their preferences for different industry pricing models. Amazon supported a continuation of the traditional "wholesale" pricing model, in which publishers sold a book to retailers at a wholesale price and retailers set the retail price. Publishers objected to Amazon's low retail prices and instead supported the "agency" model, in which the publisher specifies the retail price with a commission for the retailer. Publishers were concerned that low e-book prices erode prices for printed books,

■ *Richard J. Gilbert is Emeritus Professor and Professor of the Graduate School, University of California, Berkeley, California. His email address is rjgilbert@berkeley.edu.*

[†] To access the Data Appendix and disclosure statement, visit
http://dx.doi.org/10.1257/jep.29.3.165

doi=10.1257/jep.29.3.165

threaten traditional brick-and-mortar book distribution, and strengthen Amazon's influence in the book industry. They backed agency pricing to raise the prices of e-books, and also to facilitate the entry of Apple as a new e-retailer and to support existing e-retailers, such as Barnes & Noble. By doing so, publishers sought to undermine Amazon's dominance of online book sales. The conflict between these pricing models has led to several high-profile legal and contractual disputes, including an antitrust complaint brought by the US Department of Justice and 33 states and US territories against Apple and five major publishers, and lengthy contract disputes between Amazon and publishers, in particular Amazon's dispute with the publisher Hachette that was resolved in November 2014.

E-book Sales

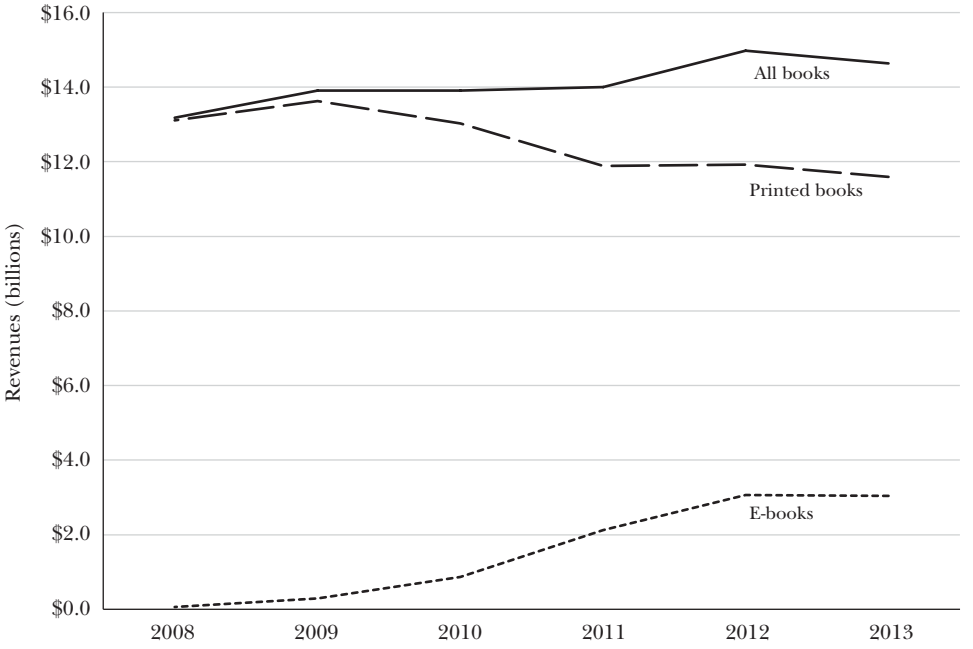
Sales of e-books in the United States increased at triple-digit rates after Amazon introduced the Kindle e-reader, reaching 20.8 percent of revenues and 23.8 percent of unit sales of trade books in both print and electronic formats in 2013. Figure 1 shows this trend. Trade books include adult and juvenile fiction and nonfiction and religious titles. These genres account for about one-half of all book sales. Non-trade books include educational materials and professional and scholarly books (Book Industry Study Group 2011, 2014).[1] Amazon is by far the largest e-book retailer, with about 70 percent of e-book sales in the first quarter of 2012. Barnes & Noble's share of e-book sales was about 20 percent, and Apple sold about another 10 percent through its Apple iBookstore (Gilbert 2013). A few other e-retailers (including Google) had very small shares of e-book sales. These statistics do not count a small amount of direct sales of e-books by publishers. Publishers have generally hesitated to launch and promote their own e-commerce operations because it would put them in competition with their retail partners.

The growth of e-book sales in the few years following the launch of the Kindle e-reader led to predictions that the days of the traditional printed book ("p-book") are numbered. But the growth of e-books has slowed and the share of e-books in total trade book sales was about flat from 2012 to 2013. There are other indications of saturation in the demand for e-books. The fraction of American adults who own either a dedicated e-reader or tablet increased from less than 10 percent in 2010 to about 50 percent by 2014. About 76 percent of adults read a book in any format in calendar year 2014, which implies that most potential e-book readers have already purchased an e-reading device (Zickuhr and Rainie 2014).

It is difficult to know whether the flattening of e-book sales is a pause in an upward trend of e-book sales or an indication that e-books have found their place and that the printed book will continue to be the most popular format. Currently, most people who read e-books also read printed books, and many readers prefer the

[1] Unless stated otherwise, all data references in this article are for US sales and are from this source.

Figure 1
Sales of Trade E-books and Printed Books



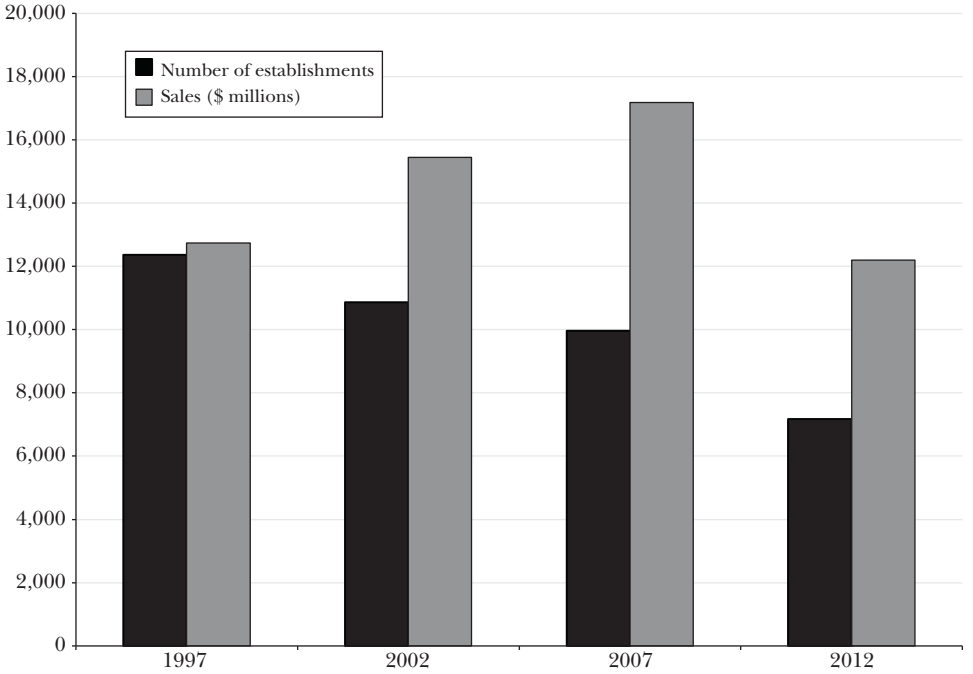
Source: Book Industry Study Group (2011, 2014).
Notes: Trade books include adult and juvenile fiction and nonfiction and religious titles. These genres account for about one-half of all book sales. Non-trade books include educational materials and professional and scholarly books.

look and feel of a printed book. Of course, preferences may change as consumers become more accustomed to electronic readers and new e-reading technologies improve the digital experience. Another factor that can affect the diffusion of e-books is the use of digital rights management (DRM) by publishers and e-vendors. DRM technologies currently limit the ability of readers to share their e-books, and changes to DRM that permit more flexible sharing could enhance future e-book usage (Bläsi and Rothlauf 2013).

The Impact of E-books on Traditional Booksellers

E-books are only one of many industry developments that have challenged the economics of the corner bookstore. Independent booksellers had already suffered fierce competition from chains, mass merchandisers such as Walmart and Costco, and online printed book sales before e-book sales achieved commercial significance. The number of booksellers declined from 12,363 in 1997 to 9,955 in 2007, a reduction of almost 20 percent (US Census, *Monthly and Annual Retail Trade Report*, 1997,

Figure 2
Number of Bookstores and Total Sales at Bookstore Establishments



Source: US Census, Economic Census of the United States and Monthly Retail Trade Survey, 1997–2012.

2007). From 1994 to 2007, out of total employment in brick-and-mortar bookstores, the share with nine or fewer employees fell from about 38 to 18 percent; conversely, the share of employment in bookstores with more than 50 employees increased from about 12 to 38 percent (Lieber and Syverson 2012). Sales of e-books were a tiny fraction of all book sales prior to 2007 and could not have had a significant impact on the profitability of independent booksellers in this period.

However, even if e-books could not possibly have contributed to the decline of independent bookstores before 2007, such an effect could have occurred since then. Figure 2 shows the number of book-selling establishments and total sales (of all items) at these establishments from 1997 to 2012. The number of booksellers fell to 7,177 by 2012, a 28 percent reduction from 2007. Sales at physical bookstores peaked in 2007, just before Amazon launched the Kindle. More recent establishment data are not available, but sales at physical bookstores have continued to decline, falling to $11.4 billion in 2013 and $10.9 billion in 2014. Online sales accounted for about 40 percent of all trade book sales in 2013, and on a unit basis roughly two-thirds of online sales in 2013 were e-books.

The data show that the number of brick-and-mortar booksellers and total sales at these establishments declined while online sales of all books rose, but further

analysis is needed to reveal whether e-books are the primary cause for the demise of physical booksellers and whether independent bookstores in particular have felt the brunt of online competition.

As part of such a fuller analysis, it is important to recognize that e-books may either substitute for sales of p-books or else may be incremental sales that would not otherwise have occurred. Hu and Smith (2013) explore this question by exploiting events in 2010 when several publishers delayed the release of a number of their e-book titles. The authors found that on average the withholding of e-books resulted in a permanent loss of cumulative e-book sales and a small (and statistically insignificant) increase in the sales of the corresponding hardcover book. There was a wide variation in consumer responses depending on awareness for individual titles as well as characteristics of the printed book, such as weight and page count. Withholding of e-books for titles with low awareness resulted in a loss of hardcover sales in addition to the loss in e-book sales, while for the most popular titles, withholding the e-book resulted in a significant substitution to sales of the printed book.

In a different study, Li (2015) uses data on the shopping behavior of Internet users to evaluate the substitution between purchases of e-books and paperback books. Estimating a demand model and counterfactually assuming the absence of e-books, she concludes that on average over the period 2008–2012, 42 percent of e-book sales occurred at the expense of sales of paperback books and the remaining 58 percent of e-book sales were incremental sales that would not have occurred if e-books were not available. These findings are consistent with the Hu and Smith (2013) evidence that, on average, withholding e-books tended to reduce total book sales.

These studies find that, on average, e-books expand book sales, but they do not identify the cause of the incremental demand. E-books cost less than their corresponding printed versions, and the incremental demand could be driven solely by price. Alternatively, consumers could have strong preferences for books in the digital format, or the increase in demand could be driven by a combination of price and format preferences.

As a rough estimate of the impact of e-books on brick-and-mortar booksellers, first note that 550 million trade e-books and 1.76 billion trade p-books were sold in 2013 (Book Industry Study Group 2014). Suppose one-half of e-book sales substituted for sales of printed books (a high estimate given the results of the cited empirical studies). Then, if e-books did not exist, sales of p-books would have increased by 275 million to about 2 billion units. Under these assumptions, the sale of e-books lowered total unit sales of printed books by all booksellers by about 14 percent in 2013.

It is likely that online sales of printed books absorbed much of this substitution because for many consumers online sales of p-books are close substitutes for e-books. It is also plausible that book chains and mass merchandisers were impacted more than independent booksellers because the large bookstores stores compete more directly on price with online sellers of printed books. Indeed, the surge in online book sales coincided with the demise of major chains such as

Borders and B. Dalton. Thus, it is likely that e-books lowered sales of p-books by independent booksellers by no more than about 14 percent in 2013 (relative to the hypothetical situation of no e-books).

E-books cannibalize the sales of some printed books, but the growth in sales of digital books is only one of many factors that have affected the profitability of independent booksellers. In a counterfactual world without e-books, independent booksellers would still have to contend with competition from chain bookstores, mass merchandisers, and online sales of printed books. It is also likely that many e-book sales are incremental to p-book sales, which should provide some comfort to authors and publishers.

Independent booksellers are enjoying a modest recovery. Membership in the American Bookseller Association reached a low 1,401 in 2009 but has rebounded to 2,094 members in 2013 (Karabell 2014). Some of the recent increase in the number of independent booksellers can be attributed to the Borders bankruptcy in 2011, which left a gap in retail bookselling establishments for independent booksellers to fill.

The future of the independent bookseller ultimately will depend on how consumer tastes evolve for books in digital formats and how new technologies may enhance the e-book reading experience. The recent flattening of e-book sales suggests that consumers will continue to enjoy printed books for years to come and the independent bookseller will not soon disappear from the retail landscape.

E-Readers and E-Books: A Platform Market?

E-books and e-readers are complementary products in a market that connects authors to readers. In some circumstances complementary products (or services) are components of a "two-sided" or "platform" market. I use these terms interchangeably, although the platform terminology applies more generally to many complements. An example would be a video game market, where a platform—say, the PlayStation—has to attract both video game developers and video game players. Complementary components in the PlayStation example are the PlayStation itself and the games. In the e-book example, complementary components are the e-reader and the e-books.

A key characteristic of a two-sided market is that purchases of one component affect consumers who purchase the other component (Rysman 2009). In a two-sided market, a firm that offers both components chooses different price-cost margins to manage demands for the components in order to increase transactions and profits (Rochet and Tirole 2006). Examples include negative prices ("rewards") for consumers to use some credit cards and positive price-cost margins for the merchants that accept the cards, and zero prices for Internet search along with positive price-cost margins for search-related advertising.

Reimers and Waldfogel (2014) examine whether Amazon treats its e-books and e-readers as components of a two-sided market. They compare Amazon's price-cost

margins for e-books to its margins for printed books. If the e-book and e-reader were components of a two-sided market for electronic books, the price-cost margins for e-books should differ from the margins for sales of p-books because Amazon would choose e-book prices taking into account the effects on consumers of e-readers. Two-sided pricing is irrelevant for printed books, which do not require the purchase of a complementary reading device.

The authors test this proposition by estimating the elasticity of demand for e-books and p-books sold by Amazon along with estimates of Amazon's wholesale marginal costs. If there are no platform effects, the price-cost margin for each product should satisfy what is called the "Lerner condition" for single-product profit maximization. If M is the price-cost margin, P is the price of the book, and e is the magnitude of the elasticity of demand,[2] the corresponding Lerner conditions for p-books and e-books are:

$$\frac{M_p}{P_p} = \frac{1}{e_p} \qquad \text{and} \qquad \frac{M_e}{P_e} = \frac{1}{e_e},$$

where the subscript "p" denotes printed books and the subscript "e" denotes the corresponding quantity for electronic books.

Reimers and Waldfogel find that Amazon prices e-books below the level that would maximize Amazon's profits solely from its e-book sales. However, they find a similar result for Amazon's sales of printed books: Amazon sells p-books at prices below the level that would maximize its profit solely from its printed book sales. Their results suggest that Amazon does not price e-books differently because they are components of a two-sided market for digital books.

Compatibility, Installed Base, and Capture for E-Books

Low prices for e-books or e-readers could be a prelude to much higher prices after customers have become habituated shoppers or the e-retailer has amassed a large installed base of users who are locked in to a proprietary e-reader format. This "capture" strategy could involve high prices for e-books (Gans 2012; Johnson 2014; Bittar 2014) or e-readers (Gaudin and White 2014). For example, an e-retailer could choose low e-reader prices and high prices for e-books to "meter" demand, much as manufacturers of printers choose high prices for their proprietary ink cartridges and low printer prices. Metering allows a firm to obtain greater revenues from consumers who buy more e-books.

A capture strategy requires that consumers be captive to an e-reading platform that is not compatible with other platforms. Otherwise, in response to higher prices for e-books or e-readers they would purchase e-books from another e-retailer and perhaps switch devices. To date neither Amazon nor Barnes & Noble—the two

[2] The elasticity e is the magnitude of the percentage change in Amazon's sales in response to a percentage in Amazon's price. It is not the overall industry demand elasticity.

largest retailers of e-books—has pursued a strategy to lock its readers into the use of its proprietary e-readers. Both have "apps" that allow their e-books to be accessed using Apple or Android devices. Among the major e-retailers, only Apple limits its customers to using its proprietary e-readers to access a book purchased from its iBookstore. However, Apple has not been particularly successful as an e-retailer, despite the popularity of the iPad as a device to read e-books.

Amazon, in particular, could raise prices and profit from a capture strategy if existing Kindle users are wedded to the Kindle platform and if higher prices do not deter consumers from purchasing Amazon's e-books and e-readers. To compete against such a strategy, existing e-book retailers would have to convince current or prospective Kindle users to purchase new e-books from their bookstores to be read on their e-readers or compatible devices. New entrants into e-retailing can source e-books from wholesalers such as Ingram and Baker and Taylor, but they would either have to offer their own e-readers or convince consumers to download apps to read their e-books on Apple or Android devices.

These impediments to competition for e-retailing are significant, but they are not so large as to sustain monopoly prices for e-books or, for that matter, monopoly prices for online sales of printed books. The retail market for e-books does not display the intensity of brand-specific network effects that allows suppliers in industries such as computer hardware and software to profit from a large installed base of users. Many consumers would abandon the Kindle platform if Amazon's e-book prices were not competitive with prices available from other e-retailers. Furthermore, it is likely that consumers would switch to a different e-book platform if they anticipate that the cost of upgrading to a newer Amazon e-reader will not be competitive with the cost of e-readers from other vendors, taking into account the expected cost of purchasing e-books from these alternatives.

The evidence is consistent with the view that Amazon sets low prices for both e-books and p-books to attract consumers to its website, where they can purchase any of the plethora of other items that Amazon sells itself or for which Amazon acts as an agent for merchants that sell in the Amazon Marketplace (Stone 2013). It is in this respect that Amazon operates a platform market, choosing low e-book prices to maximize total transactions on the Amazon.com platform.

Amazon's Monopsony Power as an E-Retailer

Amazon accounts for at least 30 percent of all trade book sales in print and digital formats, and some sources put the figure in excess of 40 percent (Milliot 2014). A large share of book sales and the absence of alternatives that can easily substitute for sales on Amazon.com could give Amazon *monopsony* power as a buyer of e-books and p-books.

A concern is that the exercise of monopsony power by Amazon will reduce the supply of books below socially optimal levels or have an adverse impact on their quality. The supply of books is socially optimal when the incremental benefit from another book equals the incremental cost of producing the book. This optimality

condition holds for both the sales of a particular book title (the intensive margin) and the number of titles (the extensive margin).

In the textbook description of monopsony, the supply of an input is an increasing function of its price. The cost of another unit to a firm with monopsony power is the price of the unit plus the additional amount the firm pays for all the other units it buys because an increase in demand raises the price. As a result, the firm exercises monopsony power by buying less than the socially efficient quantity to lower its costs, which in turn results in a reduction of output (Blair and Harrison 1991; Carlton and Perloff 2005).

The exercise of monopsony power for sales of a particular e-book title differs from the textbook model because the supply of the e-book is not an increasing function of its price when publishers license their e-books for sales by e-retailers at a uniform wholesale price. Consequently, holding the number of titles fixed, Amazon has no incentive to affect sales adversely on the intensive margin by reducing its demand for an e-book in order to lower its wholesale price.

Nonetheless, Amazon could seek to exploit its power as a large buyer to obtain low wholesale prices, rebates, or other concessions from its suppliers, and a credible concern is that Amazon will continue to press its suppliers for better terms. Publishers complain that at Amazon, today's wholesale price is the starting point for tomorrow's negotiations. Holding the number of e-book titles constant, a low wholesale price is socially desirable because the social marginal cost of reproducing and licensing another unit of an e-book for sale is close to zero. Consumers also benefit from low wholesale prices to the extent that they suppress retail prices.

In contrast to the supply of units of a particular e-book, the supply of book *titles* depends on prices. Retail prices are correlated with wholesale prices and affect revenues from the sale of printed and digital books. These revenues in turn affect the earnings of publishers, which fund author royalties and book promotions. A lower retail price reduces the revenues earned by a book if the price elasticity of demand for sales of the book is less than one, meaning that a percentage increase in the price of the book induces a less than proportional increase in demand. An Amazon press release claims, to the contrary, that publishers benefit from low e-book prices because the demand for e-books is highly price elastic.[3]

Amazon's claims about the price elasticity for e-books are unproven and do not account for the displacement of p-book sales by e-books. Authors and publishers are concerned that Amazon will use its clout to negotiate to pay publishers lower prices for e-books, leaving less for publishers to offer authors as royalty income, as well as to fund promotions and to retain as profit. However, low wholesale book prices are of no value to Amazon if they result in little content for Amazon to distribute. Furthermore, calculation of the welfare effects from low wholesale e-book prices should

[3] "It's also important to understand that e-books are highly price-elastic. This means that when the price goes up, customers buy much less." The Amazon Books team, Update re: Amazon/Hachette Business Interruption, available at http://www.amazon.com/forum/kindle?_encoding=UTF8&cdForum=Fx1D7SY3BVSESG&cdThread=Tx3J0JKSSUIRCMT, accessed May 24, 2015.

weigh the harm to consumers on the extensive margin from fewer book titles against the benefit to consumers on the intensive margin from greater sales of the books titles that are produced.

Disintermediation: Vertical Integration into Publishing

Publishers have the additional concern that they will become an antiquated and redundant component of the book industry as Amazon increasingly deals directly with authors to supply books. Publishers fear that Amazon will "disintermediate" the supply chain, replacing the traditional role of publishers to source and distribute content. Amazon Publishing already features more than a dozen imprints that specialize in publications such as literary fiction, nonfiction, and self-help books. The entry of Amazon and other e-retailers such as Apple into a business in which the e-retailer acts as an agent for authors who utilize the e-retailer's platform to self-publish their work is an example of backward vertical integration in the supply chain for books.

E-books provide a low-risk format for self-publishing because fixed costs are low, there are no inventory costs, and e-books are easily replicated to meet demand. By accessing authors directly, Amazon and other electronic publishers can increase the share of revenues available for author royalties and give authors easy access to potential readers, but typically without the promotion that traditional publishers might offer. Authors have embraced these opportunities and self-published books account for an increasingly large share of sales of e-books as well as online sales of printed books. In April 2013, five of the ten best-selling e-books were self-published (Greenfield 2013).

Publishers and some authors have raised concerns that Amazon's efforts to promote self-published books will discourage the production of creative works. This reflects a belief that Amazon's pursuit of profits as a publisher will fail to produce the quality of book content that readers desire and, by competing head-to-head with traditional publishers, Amazon will erode the margins that traditional publishers require to source and promote high-quality books.

Market forces do not necessarily produce the level of product quality that benefits consumers, as the pioneering work by Spence (1975) explained. When products such as books are sold at a uniform price, revenues are driven by the preferences of consumers who are on the margin between making and not making a purchase. However, the content of a book affects all of its readers, not just those on the purchase margin. When consumers who are not on the purchase margin have preferences that differ from the preferences of the marginal consumers, market incentives to supply the relevant nonprice attributes can differ from socially efficient incentives. Compared to socially optimal levels, the market can supply too little quality, or too much.

Amazon and traditional publishers have market incentives to produce creative content to sell books, but these market incentives may result in a poor selection of published books to the extent that the publishers focus on the choices made by

readers who are on the margin between buying and not buying their books and not on the welfare of all readers. The concerns expressed by publishers and authors imply that Amazon's pursuit of marginal sales will diminish the quality of the books it sells, although they have not articulated why these incentives should be different for Amazon than for traditional publishers.

Self-publishing is not a perfect substitute for the services of major publishers, which include quality control, editing, promotion, and, for printed books, access to brick-and-mortar bookstores. Demand for these services will continue and assure a role for major publishers, although that role is likely to become more confined to potential blockbuster titles that account for a significant share of book sales. Smaller publishers also can remain viable despite the surge of self-publishing, provided that they develop and maintain a brand reputation as a source of supply of titles that appeal to segments of the reading public.[4]

Publisher Strategies to Confront Amazon's Buyer Market Power

Publishers have sought ways to reduce Amazon's influence in book distribution, either by restricting e-books to a smaller market niche or by encouraging Amazon to set higher e-book prices. Several publishers delayed, or announced plans to delay, the release of their e-books in an attempt to limit what they perceived as cannibalization of their print sales (Rich 2009). Macmillan delayed the release of several titles in 2009, including a volume in the popular "Wheel of Time" fantasy series. Simon & Schuster delayed the release of the e-book editions of about 35 major titles by four months in the first half of 2010. Hachette announced plans to delay the release of many of its e-book titles during that time period, and was joined by HarperCollins and Macmillan announcing they would do the same for some of their titles.

This "windowing" strategy has a parallel in the way that publishers control printed book formats. Publishers initially perceived softcover books as a threat to their traditional business. They resolved the threat by delaying the release of paperbacks and in the process implemented an effective method of price discrimination. However, the analogy between e-books and softcover printed books is imperfect. The paperback book is essentially the same product as its hardcover sibling with a relatively small quality differential. The e-book is a different product than a printed book and a significant fraction of e-book sales are incremental to printed book sales. Moreover, the delayed release of e-books encourages piracy to fill the void left by titles that are withheld from the market. After a flurry of actual and threatened delays, publishers stopped windowing their e-book releases, reflecting a belief that the strategy risks permanently losing or significantly delaying sales that they would have made otherwise.

[4] Although I consider how monopsony power may affect the supply of e-books, this article does not address the public policy concerns that may arise from Amazon's commanding position as an e-retailer and its various practices.

An alternative strategy that publishers considered to address their concerns about low e-book retail prices is to increase their wholesale prices. E-books typically have a lower wholesale price than p-books to reflect their lower production costs. Some publishers experimented with raising the wholesale price of their e-books to the wholesale price of the p-book (Cote 2013b). However, this strategy failed to move Amazon from its prior commitment to a $9.99 price for many new releases and best sellers; moreover, Amazon could neutralize an attempted price increase by a single publisher by refusing to promote the publisher's titles. Furthermore, an effort to raise e-book prices by a single publisher would not address broader concerns about the threat of low e-book prices to the traditional book retailing industry and would not address Amazon's commanding position as an e-retailer.

Yet another alternative is for the publisher to control the retail prices of its e-books directly. In this pricing model, Amazon and other e-retailers would become agents of the publisher, selling e-books at prices set by the publisher and retaining a commission for their retailing services. By setting a sufficiently high retail commission, publishers sought to promote the entry of new e-retailers and assure the profitability of existing e-retailers such as Barnes & Noble. Greater retail competition would undermine Amazon's monopsony power for sales of e-books and online sales of printed books, and the retail commission would facilitate the publishers' goal of raising e-book retail prices.

This "agency" pricing model raises issues similar to the classic problem of "resale price maintenance," for which an upstream supplier determines or sets bounds on the retail price. Both involve maintaining a retail margin. A distinction is that agency pricing specifies both the retail price and the margin retained by the retailer for its services.

Neither resale price maintenance nor agency pricing necessarily offends the antitrust laws. In a recent case, *Leegin Creative Leather Products, Inc. v. PSKS, Inc.*, (551 US 877 [2007]), the US Supreme Court recognized the economic argument that resale price maintenance can have pro-competitive effects by promoting retail services and thereby allow the firms that use these services to be more effective competitors, while also noting that the practice has the potential to harm consumers. Similar arguments apply to agency pricing. Both resale price maintenance and agency pricing are examples of "vertical restraints," in which firms at one level of the supply chain determine prices or other conditions of sale at a different level of the chain. Vertical restraints differ from fixing prices among competitors at the same level of the supply chain, which is generally condemned as unlawful.

Many industries operate on a wholesale pricing model, while others use agency pricing, and some firms use both. Apple uses agency pricing for apps, allowing app developers to set their prices while providing Apple with a revenue share for placement on the iPad and iPhone. In contrast, Apple employs wholesale pricing for its iTunes store, paying music suppliers a wholesale price for their content and setting the retail price itself. Amazon employs agency pricing for its "Marketplace," which allows merchants to sell their products on the Amazon.com website.

Tradeoffs in Wholesale and Agency Pricing

Most publishers of printed books use a wholesale pricing model for sales to brick-and-mortar bookstores. They sell books to these retailers at a wholesale price equal to a discount off the suggested retail price, and retailers are free to set their own prices. Many retailers sell the book at its suggested retail price. Chain bookstores and mass merchandisers often price books more aggressively, particularly for new releases and selected titles that have the power to attract customers to their stores.

Publishers initially applied the wholesale pricing model to their e-books, too. Amazon priced many popular "front-list" titles and best sellers at $9.99, which was often below the wholesale price of the e-book. Publisher and author objections to Amazon's low prices may seem odd to economists who are sensitized to the potential costs imposed by a traditional wholesale pricing model. If a publisher sells a book to a retailer at a wholesale price, W, which is a mark-up over its production cost, economists expect the retailer to add its own mark-up, M, to arrive at the retail price, $W + M$. This "double-marginalization" reduces sales and lowers the profits available to the publisher and the author.

In the short run, a low retail margin is good for authors and publishers. Publishers sell more books; and authors, whose royalties are based on the wholesale price, thus earn more royalties. However, publishers complain that low e-book prices cannibalize their sales of printed books and jeopardize the brick-and-mortar bookstores that promote their bestsellers. Publishers and some authors are concerned that Amazon will use low retail prices as a bargaining lever to negotiate lower wholesale prices, squeezing publisher revenues and author royalties. Publishers sought agency pricing to exercise greater control over e-book pricing.

Economic theory does not offer a general conclusion as to whether firms in a vertical relationship, such as publishers and e-retailers, or their consumers are necessarily better or worse off with agency pricing compared to wholesale pricing. The comparison depends on numerous industry factors. When firms share a common objective to maximize the profits from a vertical relationship, they can choose the pricing arrangement that maximizes their joint profits and then bargain over their respective shares. In that case, the choice between wholesale and agency pricing is a decision about whether pricing authority should vest in the upstream firms that supply a good or the downstream firms that retail the good in order to maximize joint profits. If agency pricing yields more total profit, then each publisher, acting alone, would prefer the agency pricing model, which can be structured to provide a retail margin that generates at least as much profit as the retailer would have earned with wholesale pricing.

If the upstream industry is highly competitive, but competition at the downstream retail level is weak, firms can achieve higher industry profits by delegating pricing to the downstream firms (Foros, Kind, and Shaffer 2013). Delegating pricing authority to downstream retailers can be accomplished with a wholesale pricing structure. Allowing downstream firms to determine prices may incur double-marginalization, but that outcome can be avoided by sharing revenues with retailers after they set

prices or by providing nonlinear pricing incentives such as volume discounts to encourage retailers to set lower prices.

In contrast, if downstream retailing is highly competitive compared to competition upstream, industry profits can be higher if the upstream firms have the authority to set retail prices. Agency pricing or resale price maintenance allows the upstream firms to determine retail prices. Other factors also influence the profit-maximizing locus of pricing authority. For example, retailers may be better able to set profit-maximizing prices if they have better information about market demand.

Complicating the economic evaluation between wholesale and agency pricing for the e-book industry was the lack of a common objective to maximize profits from the sale of e-books. Amazon sought low e-book prices in part to attract more consumers to its website. Publishers sought higher e-book prices to protect their traditional brick-and-mortar distribution business and, more significantly, to create retail competition in order to parry Amazon's relentless drive to reduce its wholesale acquisition costs. These divergent preferences impeded attempts to move Amazon to agency pricing in bilateral negotiations with individual publishers.

The Apple Agency Contract and the Antitrust Case against Apple and Five Publishers

In the months preceding Apple's launch of the iPad in April 2010, Apple and five major publishers—Hachette Book Group, HarperCollins, Macmillan, Penguin, and Simon & Schuster—negotiated similar agency contracts that allowed each publisher to set retail prices for newly released e-books and the e-book versions of *New York Times* bestsellers, subject to price caps, and allowed Apple as the retailer to keep 30 percent of the retail price. The price caps were related to the retail price of the corresponding hardcover book. For example, the e-book price cap was $12.99 for a title with a hardcover retail price between $25.01 and $27.50 (Gilbert 2013).

On April 11, 2012, the US Department of Justice together with 33 states and US territories filed a complaint alleging that Apple and the five publishers had conspired to raise, fix, and stabilize the retail price for newly released and bestselling trade e-books. The case alleged that the five publishers, with the assistance of Apple, overcame Amazon's resistance to agency pricing by acting jointly to require Amazon to accept agency pricing or else do without their e-books (US Department of Justice 2012). Although Apple's relationship to the publishers is vertical as a retailer of their e-books, the allegation was that Apple facilitated collusion among the five publishers to adopt the agency model, with resulting higher prices for e-book consumers.

The Apple agency contracts with the five publishers included a "most favored nation" clause, which gave Apple the right to sell their e-books at the lowest retail price charged by any of the publishers' e-retailers, while retaining a commission for the sale. The most-favored-nation clause discouraged publishers from selling their e-books to e-retailers that would set retail prices below the levels specified in the agency contracts and, as a result, made it easier to convince Amazon to accept agency pricing.

The federal and state governments' case against Apple and the publishers did not challenge the agency pricing model itself, nor did the case allege that the most-favored-nation provisions in the agency contracts are inherently anticompetitive. In fact, contracts often include most-favored-nation provisions. Their effects depend on the circumstances in which they are deployed (Crocker and Lyon 1994), and courts have upheld most-favored-nation clauses in a variety of contexts (Steuer 2015; Dennis 1995).

A complication in the e-books case is that the collusion that was alleged and challenged was not collusion to raise agency prices, but instead was collusion to facilitate a transition from wholesale pricing to agency pricing, in a situation where either approach could be an equilibrium industry model for e-book pricing. A further complication is that, while agency pricing raised e-book prices to consumers, under the agency formula adopted by Apple and the publishers, the publishers earned less on many of their e-book sales than they earned with wholesale pricing, at least in the short run.

The publishers adopted agency pricing in part to sponsor the entry of Apple's iBookstore as a new e-retailer and to increase competition from other e-retailers to undermine Amazon's monopsony power.[5] In this respect, one might argue that the move to agency pricing was pro-competitive. However, the intent and effect of agency pricing was to increase the prices that consumers pay for e-books. Comparing prices of e-books in the window one week before the switch to agency pricing with prices one week later in Gilbert (2013), I found that on average, the publishers' retail prices increased by 18.6 percent at Amazon and 19.9 percent at Barnes & Noble. Although the agreements were limited to new releases and *New York Times* bestsellers, prices for the agency publishers' other e-books increased by similar amounts over this period.

Random House resisted the move to agency pricing when the five major publishers reached their agreements with Apple, but then moved to agency pricing more than a year later. Using e-books sold by Random House as a control and adding other explanatory variables and fixed effects, Ashenfelter (2013) compared e-book prices six months before the switch to agency to prices in the ensuing six-month period. He concluded that the switch to agency increased e-book retail prices by an average of 16.8 percent and lowered sales by 14.5 percent. Alternative specifications using different pre- and post-agency windows produced similar results.

Subsequent to the US Department of Justice complaint, the five publishers entered into settlement agreements that terminated their agency contracts and prevented the publishers from restricting the right of any e-retailer to set the retail price of their e-books for a period of about two years, in effect reinstating the wholesale pricing model in this two-year period (Cote 2012, 2013a, 2013c). De los Santos and Wildenbeest (2014) utilize these agreements to provide an additional difference-in-differences test of the effects of agency on the prices of e-books. They

[5] Apple sought agency pricing to improve the economics of the iBookstore, a feature of the iPad. The launch of Apple's iPad did not depend on the viability of the iBookstore (Cote 2013b).

compared publishers' prices in the period during which the settlement agreements compelled a return to the wholesale model with their prices when publishers employed agency pricing. They found that retail prices at Amazon decreased on average by 17 percent after the settlement agreements allowed Amazon to regain control of prices, while Barnes & Noble's prices decreased by 7 percent, compared to the period with agency pricing.

The publisher settlements left Apple as the sole remaining trial defendant in the governments' e-book case. On July 10, 2013, Judge Denise Cote held that Apple participated in and facilitated a horizontal price-fixing conspiracy to raise the price of e-books in violation of the antitrust laws. [6] The court of appeals affirmed the verdict.

New E-Book Pricing Arrangements: Unilateral Moves to Agency Pricing?

The settlement agreements negotiated with the five defendant publishers allow the publishers to enter into agency contracts after about two years, with certain limitations. The agreements expire after five years.

In October 2014, soon after the expiration of the two-year window in the settlement agreement, Amazon and Simon & Schuster agreed to a new multiyear contract under which Simon & Schuster sets the retail prices for its e-books. The contract also provides unspecified incentives for the publisher to deliver lower prices for readers (as reported in Trachtenberg 2014). Hachette concluded a similar multiyear agreement with Amazon in November 2014 effective for sales commencing in 2015 (as reported in Trachtenberg and Bensinger 2014). The Amazon–Hachette negotiations were particularly lengthy and acrimonious. At one point, Amazon gave a glimpse of its monopsony power when it stopped allowing "pre-ordering" of new p-books from Hachette, which meant that customers buying those books often experienced a delay of several weeks before the book arrived.

Although the details of these contracts between Amazon and the publishers are confidential, they appear to reflect a new willingness on the part of Amazon to accept higher retail e-book prices. De los Santos and Wildenbeest (2014) observe a trend toward higher e-book prices at Barnes & Noble and Amazon, especially for newer and more popular titles, following the expiration of the two-year window during which the publisher settlement agreements prohibited the publishers from interfering with retail discounting of their e-books.

The observed trend toward higher e-book prices suggests that publishers successfully nudged Amazon to raise the retail prices of its e-books. A question that is relevant to the impact of the conduct by Apple and the five defendant publishers is whether the adoption of some form of agency contracts would have occurred even if the publishers had not agreed to coordinate their negotiations

[6] The July 10, 2013, ruling is available at http://www.justice.gov/atr/cases/f299200/299275.pdf. Judge Cote issued a Final Judgment with a renewable term of five years that restricts the ability of Apple and the publishers to enter into contracts that impede Apple's ability to set retail e-book prices or to offer discounts or other promotions. The final judgment also prohibits Apple from enforcing a most-favored-nation provision with any e-book publisher (Cote 2012d).

with Amazon but instead had acted individually. Amazon's insistence on very low retail margins for its e-books may have limits now that it has established its command of the e-retail marketplace and the rate of growth of e-book sales has slowed. The "agency-lite" contracts that Amazon has negotiated with Simon & Schuster and Hachette are likely to become a template for other contracts between publishers and Amazon, and establish a role for agency pricing in the e-book marketplace.

Conclusion

Before the advent of the e-book, the major book publishers and Amazon had found common ground as the Internet opened a new distribution channel for publishers to sell their printed books and for Amazon to establish its presence as the Internet superstore. However, the e-book roiled the traditional publishing industry by expanding the reach of online book retailing and further entrenching Amazon's position in this increasingly important retail channel.

Amazon's aggressive pricing of e-books squeezed margins for other book retailers and alarmed publishers who were concerned about their growing reliance on Amazon to distribute their products. Publishers worried that Amazon's low e-book prices and high share of e-book sales threatened the viability of brick-and-mortar booksellers, including the many independent booksellers who offer advice and encouragement to readers of their books.

E-books likely take sales away from independent booksellers. However, the impact is not particularly large compared to the competition that exists from chain bookstores, mass merchandisers, and online sales of printed books, and the number of independent booksellers has recovered from a low in 2009. Online sales diminish the importance of brick-and-mortar bookstores in the distribution chain, but online channels also open up a variety of other promotional opportunities and ways for publishers to contact potential book-buyers. Furthermore, the empirical evidence is that e-books expand total book sales.

As a powerful buyer, Amazon has the incentive and ability to bargain for low wholesale prices. Conditional on the supply of book titles, lower wholesale prices benefit consumers if they are passed on to lower retail prices. In that event, low wholesale prices are pro-competitive on the intensive margin by bringing retail prices closer to the low marginal cost of licensing and selling an e-book. However, low e-book prices may harm competition on the extensive margin by reducing the supply of book titles if they lower the book revenues available to pay author royalties and cover the cost of promoting new books.

After trying various strategies to limit Amazon's influence as an e-book retailer, several major publishers worked with Apple to impose an agency pricing model for e-books in which the publisher sets the retail price and the e-retailer earns a specified commission. The US Department of Justice and a number of states and territories alleged that this conduct was collusive behavior in violation of the antitrust laws.

The publishers accepted settlement agreements that prohibited them from setting retail e-book prices for a period of about two years, after which they are permitted to negotiate contracts that give the publishers a limited ability to control the retail prices of their e-books.

In keeping with the terms of their settlement agreements, which expire after five years, Hachette and Simon & Schuster negotiated contracts with Amazon that allow the publishers to determine the prices of their e-books under certain conditions—although the specifics of these "agency-lite" contracts are not yet clear. Similar agreements with other publishers are likely to follow, and agency pricing for e-books may yet prevail in some form, as publishers seek out ways to diminish their reliance on Amazon and as both publishers and retailers seek higher profits from the sale of digital books.

Publishers also fear that Amazon's entry into publishing is a harbinger of a new industry in which Amazon deals directly with authors, and publishers' traditional roles will be severely compromised. They have reasons to be concerned given the success of self-publishing programs sponsored by Amazon and other e-retailers, such as Apple. The e-book story shows how the traditional players in the book industry are struggling to achieve a new market equilibrium in a time where their industry is facing severe technological disruption and illustrates the hazards they face in attempting to manage the transition to that new equilibrium.

■ *I am grateful for advice and feedback from Jonathan Baker, Ana Carolina Bittar, Kun Huang, Justin Johnson, Hui Li, Joseph Mangan, James Ratliff, Ray Riegert, Daniel Rubinfeld, Richard Steuer, Nadine Vassallo, and the journal editors Gordon Hanson, Enrico Moretti, and Timothy Taylor. I consulted for the US Department of Justice in its antitrust case against Apple and five publishers. Nothing in this article relies on confidential facts or communications in that case.*

References

Ashenfelter, Oley C. 2013. Direct Testimony of Orley C. Ashenfelter, Ph.D. in United States v. Apple, et al. (June 6, 2013). Trial Exhibit PX-1097. Available at http://www.justice.gov/atr/cases/apple/apple-direct.html.

Bittar, Ana Carolina. 2014. "Unlocking the Gates of Alexandria: DRM, Competition and Access to E-Books." July 25. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2620354.

Blair, Roger D., and Jeffrey L. Harrison. 1991. "Antitrust Policy and Monopsony." *Cornell University Law Review* 76(2): 297–340.

Bläsi, Christoph, and Franz Rothlauf. 2013. "On the Interoperability of eBook Formats." European and International Booksellers Federation, April.

Book Industry Study Group. 2011. *BookStats: An Annual Comprehensive Study of the U.S. Publishing Industry*. Vol. 1. Association of American Publishers.

Book Industry Study Group. 2014. *BookStats: An Annual Comprehensive Study of the U.S. Publishing Industry*. Vol. 4. Association of American Publishers.

Carlton, Dennis W., and Jeffrey M. Perloff. 2005. *Modern Industrial Organization*, 4th edition. Addison Wesley.

Cote, Denise. 2012. U.S. Department of Justice v. Apple, et al., U.S. District Court for the Southern District of New York, Final Judgment as to Defendants Hachette, HarperCollins, and Simon & Schuster, September 6, 2012.

Cote, Denise. 2013a. U.S. Department of Justice v. Apple, et al., U.S. District Court for the Southern District of New York, Final Judgment as to Defendants Penguin Group, a Division of Pearson PLC, and Penguin Group (USA), Inc., May 17, 2013.

Cote, Denise. 2013b. U.S. Department of Justice v. Apple, et al., U.S. District Court for the Southern District of New York, Opinion and Order by Judge Denise Cote, July 10, 2013.

Cote, Denise. 2013c. U.S. Department of Justice v. Apple, et al., U.S. District Court for the Southern District of New York, Final Judgment as to Defendants Verlagsgruppe Georg Von Holtzbrinck Gmbh & Holtzbrinck Publishers, llc D/B/A Macmillan, August 12, 2013.

Cote, Denise. 2013d. U.S. Department of Justice v. Apple, et al., U.S. District Court for the Southern District of New York, Plaintiff United States' Final Judgment and Plaintiff States' Order Entering Permanent Injunction, September 5, 2013.

Crocker, Keith J., and Thomas P. Lyon. 1994. "What Do 'Facilitating Practices' Facilitate? An Empirical Investigation of Most-Favored-Nation Clauses in Natural Gas Contracts." *Journal of Law and Economics* 37(2): 297–322.

De los Santos, Babur, and Matthijs R. Wildenbeest. 2014. "E-Book Pricing and Vertical Restraints." NET Institute Working Paper no. 14-18, October. Available at SSRN: http://ssrn.com/abstract=2506509.

Dennis, Anthony J. 1995. "Most Favored Nation Contract Clauses under the Antitrust Laws." *University of Dayton Law Review* 20(3): 821–54.

Foros, Øystein, Hans Jarle Kind, and Greg Shaffer. 2013. "Turning the Page on Business Formats for Digital Platforms: Does Apple's Agency Model Soften Competition?" CESifo Working Paper 4362, August. Available at SSRN: http://ssrn.com/abstract=2317715.

Gans, Joshua S. 2012. "Mobile Application Pricing." *Information Economics and Policy* 24(1): 52–59.

Gaudin, Germain, and Alexander White. 2014. "On the Antitrust Economics of the Electronic Books Industry." September 24. Available at SSRN: http://ssrn.com/abstract=2352495.

Gilbert, Richard J. 2013. Direct Testimony of Richard J. Gilbert, Ph.D., United States v Apple, et al. (April 25). Trial Exhibit PX-1105, Available at http://www.justice.gov/atr/cases/apple/apple-direct.html.

Greenfield, Jeremy. 2013. "Self-Published Titles Dominate Top of Ebook Best-Sellers List." *Digital Book World*, April 30. http://www.digitalbookworld.com/2013/self-published-titles-dominate-top-of-ebook-best-sellers-list/.

Hu, Yu Jeffrey, and Michael D. Smith. 2013. "The Impact of E-book Distribution on Print Sales: Analysis of a Natural Experiment." April 30. Available at SSRN: http://ssrn.com/abstract=1966115.

Johnson, Justin P. 2014. "The Agency Model and MFN Clauses." January 10. Available at SSRN: http://ssrn.com/abstract=2217849.

Karabell, Zachary. 2014. "Why Indie Bookstores Are on the Rise Again." *Slate*, September 9.

Li, Hui. 2015. "Cannibalization or Market Expansion? The Impact of E-books on Print Book Sales." February 20. Available at SSRN: http://ssrn.com/abstract=2613757.

Lieber, Ethan, and Chad Syverson. 2012. "Online versus Offline Competition." Chap. 8 in *The Oxford Handbook of the Digital Economy*, edited by Martin Peitz and Joel Waldfogel. Oxford University Press.

Milliot, Jim. 2014. "BEA 2014: Can Anyone Compete with Amazon?" *Publishers Weekly*, May 28.

OECD. 2012. *E-books: Developments and Policy Considerations*. OECD Digital Economy Paper 208, OECD Publishing.

Reimers, Imke, and Joel Waldfogel. 2014. "Throwing the Books at Them: Amazon's Puzzling Long Run Pricing Strategy." March 4. Available at SSRN: http://ssrn.com/abstract=2442747.

Rich, Motoko. 2009. "Publishers Delay E-Book Releases." *New York Times*, December 9.

Rochet, Jean-Charles, and Jean Tirole. 2006. "Two-Sided Markets: A Progress Report." *Rand Journal of Economics* 37(3): 645–67.

Rysman, Marc. 2009. "The Economics of Two-Sided Markets."*Journal of Economic Perspectives* 23(3): 125–43.

Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *Bell Journal of Economics* 6(2): 417–29.

Steuer, Richard M. 2015. "Online Price

Restraints Under U.S. Antitrust Law." *CPI Antitrust Chronicle* 5(1): 1–9.

Stone, Brad. 2013. *The Everything Store: Jeff Bezos and the Age of Amazon*. New York: Little Brown.

Trachtenberg, Jeffrey A. 2014. "Amazon, Simon & Schuster Reach Book Contract." *Wall Street Journal*, October 20.

Trachtenberg, Jeffrey A., and Greg Bensinger. 2014. "Amazon, Hachette End Publishing Dispute." *Wall Street Journal*, November 13.

US Census. 1997, 2002, 2007, 2012. *Economic Census of the United States*. NAICS code 45211.

US Census. 1997–2012. *Monthly and Annual Retail Trade Report, Retail and Food Service Sales*. NAICS code 45211.

US Department of Justice. 2012. U.S. Department of Justice v. Apple, et al., U.S. District Court for the Southern District of New York, Complaint, April 11, 2012.

Zickuhr, Kathryn, and Lee Rainie. 2014. "E-Reading Rises as Device Ownership Jumps." Pew Research Center Report, January 14. http://pewinternet.org/Reports/2014/E-Reading-Update.aspx.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 185–208

The Indian Gaming Regulatory Act and Its Effects on American Indian Economic Development[†]

Randall K. Q. Akee, Katherine A. Spilde, and Jonathan B. Taylor

The Indian Gaming Regulatory Act (IGRA), passed by the US Congress in 1988, was a watershed in the history of policymaking directed toward reservation-resident American Indians. IGRA set the stage for tribal government-owned gaming facilities. It also shaped how this new industry would develop and how tribal governments would invest gaming revenues. Since then, Indian gaming (the casinos and bingo halls owned by tribal governments in the United States are also sometimes referred to as *tribal gaming* or *tribal government gaming*) has approached commercial, state-licensed gaming in total revenues. Gaming operations have had a far-reaching and transformative effect on American Indian reservations and their economies. Specifically, Indian gaming has allowed marked improvements in several important dimensions of reservation life. For the

■ *Randall Akee is an Assistant Professor of Public Policy, University of California at Los Angeles, Los Angeles, California. He is a Research Fellow, Harvard Project on American Indian Economic Development, Harvard Kennedy School of Government, Cambridge, Massachusetts; Faculty Affiliate, Center for Effective Global Action (CEGA), University of California, Berkeley, California. Katherine Spilde is an Associate Professor of Hospitality and Tourism Management and Endowed Chair of the Sycuan Institute on Tribal Gaming, San Diego State University, San Diego, California. Jonathan B. Taylor is President of the Taylor Policy Group, Sarasota, Florida. He is also a Research Affiliate, Harvard Project on American Indian Economic Development, Harvard Kennedy School of Government, Cambridge, Massachusetts, and Senior Policy Associate, Native Nations Institute, Udall Center for Studies in Public Policy, University of Arizona, Tucson, Arizona. Akee is the corresponding author. The authors' email addresses are rakee@ucla.edu, kspilde@mail.sdsu.edu, and jonathan@taylorpolicy.com.*

[†] To access the Data Appendix, visit
http://dx.doi.org/10.1257/jep.29.3.185

doi=10.1257/jep.29.3.185

first time, some tribal governments have moved to fiscal independence. Native nations have invested gaming revenues in their economies and societies, often with dramatic effect. Table 1 provides selected characteristics of American Indian social and economic conditions over the past two decades: incomes for American Indians grew at six times the US rate; female labor force participation rose; unemployment fell; and reservation housing quality improved. Relative improvement across a range of census indicators was particularly strong in the 1990s, the first census decade after IGRA's passage, and continued in the 2000s, albeit at a slower pace.

While on average there have been large improvements, the effect of Indian gaming varies tremendously across tribes. Some tribes have had spectacular successes; others have found gaming to be a small part of their economic portfolio and of limited importance to their tribal government revenues and communities. Annual Indian gaming revenues increased from about $100 million in 1988 to $28 billion dollars in 2013 (National Indian Gaming Commission 2014; Senate Committee on Indian Affairs 1988). The number of tribal gaming operations went from fewer than 30 to about 450 across 31 states. Tribal gaming affects reservations with fewer than 100 residents to those with populations that number in the tens of thousands. In addition to the variation arising from differential access to markets, corporate governance, and managerial skill, there are instances where state-tribal conflict has held Indian gaming below its potential.

The focus of this paper is on Indian Country, a broad term often used to describe tribal lands in the United States. The term also has specific meaning in US law (18 USC §1151). In 2012, the contiguous 48 states held 324 reservations (or trust lands or joint use areas) in 32 states, home to more than 300 federally recognized tribes (Osier 2012) and 540,000 people self-reporting that they were American Indian or Alaska Native alone (that is, not in combination with other races) (US Census 2011a). An additional 33 federally recognized tribes were affiliated with 33 census tribal statistical areas in California, New York, Oklahoma, and Washington.[1] After the reservations themselves, it is typical to find the next-highest concentration of members of a tribe living in the reservation environs or in nearby cities: say, Navajo living in Flagstaff, Arizona, or Oglala Lakota in Rapid City, South Dakota. Of course, many American Indians maintain civic, economic, social, and cultural ties with reservation communities regardless of where they live. The discussion here focuses on conditions in the contiguous 48 states and does not characterize distinctive Native Hawaiian and Native Alaskan histories, policies, or conditions.

We begin with an overview of policymaking leading up to the political and legal fights for Native self-determination, of which Indian gaming is an outgrowth. We consider the steps, starting in the late 1980s with a key US Supreme Court decision

[1] In all 50 states, the population reporting American Indian or Alaska Native (AIAN) alone was 2,932,248, and the number of Americans reporting AIAN alone or in combination with one or more races was 5,220,579 (US Census 2011a).

Table 1

Selected Indicators of Social and Economic Condition

(Indians on reservations in the contiguous 48 states in bold vs. US all-races averages in parentheses)

	Amount (in percent unless indicated as $)			*Change (in percentage points unless indicated as %)*		
	Census 1990	*Census 2000*	*ACS 2006–10*	*1990s*	*2000s*	*Both decades*
Real per capita income	**$7,673** ($24,951)	**$10,227** ($27,798)	**$11,406** ($26,893)	**33.3%** (11.4%)	**11.5%** (−3.3%)	**48.6%** (7.8%)
Real median household income	**$21,201** ($52,001)	**$28,689** ($54,077)	**$28,298** ($51,076)	**35.3%** (4.0%)	**−1.4%** (−5.5%)	**33.5%** (−1.8%)
Child poverty	**55.6** (18.3)	**44.3** (16.6)	**43.9** (19.2)	**−11.4** (−1.7)	**−0.4** (2.6)	**−11.8** (0.9)
Family poverty	**47.7** (10.0)	**35.7** (9.2)	**32.2** (10.1)	**−12.0** (−0.8)	**−3.5** (0.9)	**−15.4** (0.1)
Unemployment	**25.7** (6.2)	**21.9** (5.7)	**18.9** (7.9)	**−3.9** (−0.5)	**−3.0** (2.1)	**−6.9** (1.6)
Labor force participation	**50.9** (65.3)	**51.5** (63.9)	**52.4** (65.0)	**0.6** (−1.3)	**0.9** (1.1)	**1.5** (−0.3)
Male labor force participation	**57.4** (74.4)	**54.7** (70.7)	**54.1** (70.9)	**−2.7** (−3.7)	**−0.6** (0.2)	**−3.3** (−3.5)
Female labor force participation	**44.8** (56.8)	**48.5** (57.5)	**50.8** (59.4)	**3.7** (0.8)	**2.3** (1.9)	**6.0** (2.6)
Overcrowded homes*	**16.1** (4.7)	**14.7** (5.7)	**8.2** (3.1)	**−1.4** (1.1)	**−6.5** (−2.6)	**−7.9** (−1.6)
Homes without complete plumbing	**20.9** (0.8)	**13.7** (0.6)	**8.6** (0.5)	**−7.2** (−0.1)	**−5.1** (−0.1)	**−12.4** (−0.3)
Homes without complete kitchens*	**11.1** (1.1)	**10.9** (1.3)	**10.7** (2.7)	**−0.2** (0.2)	**−0.2** (1.4)	**−0.4** (1.6)
High school degree only	**29.3** (30.0)	**31.2** (28.6)	**35.0** (29.0)	**1.9** (−1.4)	**3.8** (0.4)	**5.8** (−1.0)
College graduate or more	**4.0** (20.3)	**6.0** (24.4)	**7.4** (27.9)	**2.0** (4.1)	**1.5** (3.5)	**3.4** (7.6)

Notes: Numbers for "Indians on reservations" are in bold; numbers for "all races nationwide" are in parentheses underneath. Dollars are 2009 dollars.

* Due to data limitations, the reservation figures for overcrowded homes and homes without complete kitchens are the all-races, rather than Indian, statistics (Akee and Taylor 2014).

and the Indian Gaming Regulatory Act of 1988, which led to the expansion of Indian gaming. We then turn to a discussion of how the growth of Indian gaming has affected Native Americans living on or near reservations, and how it has affected nearby localities and regions. We conclude with thoughts about the future of Indian gaming and the research agenda in this area.

A Brief Policy History of Indian Country

Most American Indian reservations were established by treaties and executive orders in the 19th century. Since then, Indian policy has oscillated between policies seeking to dissolve American Indian communities and tribes, and policies supportive of American Indian self-rule under duly constituted governments (for overviews, see American Indian Lawyer Training Program 1988, pp. 8–15; Cornell 1988, p. 14; Wilkins 2002, p. 105).

Under the "Allotment Era" inaugurated in 1887 by the Dawes Act, federal law privatized reservation lands (for example, apportioning 160 acres per household) and marked large portions of reservation lands as "surplus" suitable for sale to private citizens. As with many laws, the Dawes Act was supported by a coalition of well-intentioned, as well as opportunistic, political forces (Carlson 1981), but the underlying idea was that individual ownership would usher Indians (and their land) into the mainstream economy. By 1934, 86 million acres of reservation land—62 percent of the total—had transferred out of Indian ownership via sale, foreclosure, lien, and fraud (Wilkinson 1988, p. 20). As a result of the impoverishing effects of the Dawes Act (for example, as documented in Meriam et al. 1928), the Indian Reorganization Act of 1934 (IRA) ushered in a "New Deal for Indians." The law ended land allotment on American Indian reservations, promoted constitutional self-government, and pointed to federally chartered tribal corporations as the primary vehicles for stimulating American Indian economic progress (Wilkins 2002). By the 1950s, policy for American Indians shifted again, to the "Termination Era," which was marked by legislation disbanding particular tribes and by passage of PL 83-280, which transferred certain tribes' criminal (and limited civil) jurisdiction to state governments.

By the late 1960s and early 1970s, American Indian assertions of tribal sovereignty via litigation and political action heralded the contemporary "Self-Determination Era," in which the federal government delegated powers and responsibilities to tribal governments. This era provided greater autonomy to tribal governments in the determination of their political institutions, economic activities, and development (Wilkins 2002). One example of this increased autonomy arose from the Indian Educational Assistance and Self-Determination Act of 1975. Under that act and successive amending legislation, Native nations tailored federal programs (such as education services) to tribal cultures and reservation conditions by contracting to deliver the federal program services directly or by compacting with the US government to operate multiple programs under multifunction arrangements similar to federal block grants to states.

Over the last few decades, executive orders from presidents of both parties have consistently supported principles of Indian self-government and a government-to-government relationship between the federal and tribal governments (Nixon 1970; Carter 1979; Reagan 1983; Bush 1991; Clinton 1994, 2000; Bush 2004; Obama 2009). In addition, federal policy increasingly treats tribes like states, or otherwise gives Indian governments latitude in crafting policies for housing, healthcare, education,

workforce development, crime, and natural resources.[2] In this period, many tribes sued the US government to defend property rights in salmon, oil, water, and timber that had been weakened by non-Indian encroachment or mismanagement by federal officials and agencies.

Through all of the various federal policy approaches toward American Indians, there is consensus that federally directed development has failed to produce sustained economic growth on reservations. Economic bright spots in Indian Country had been few (Cornell and Kalt 1992, p. 3). American Indians residing on reservations have regularly been among the poorest people in the United States. In the 1970 US Census, the per capita income of Indians on major US reservations was 32 percent of the US average. It rose to 41 percent of the national average in 1980 but fell to 32 percent again by 1990 (Akee and Taylor 2014). The decline in the 1980s has been attributed to the pronounced retreat of federal funding directed toward Indian Country in that decade (Trosper 1996).

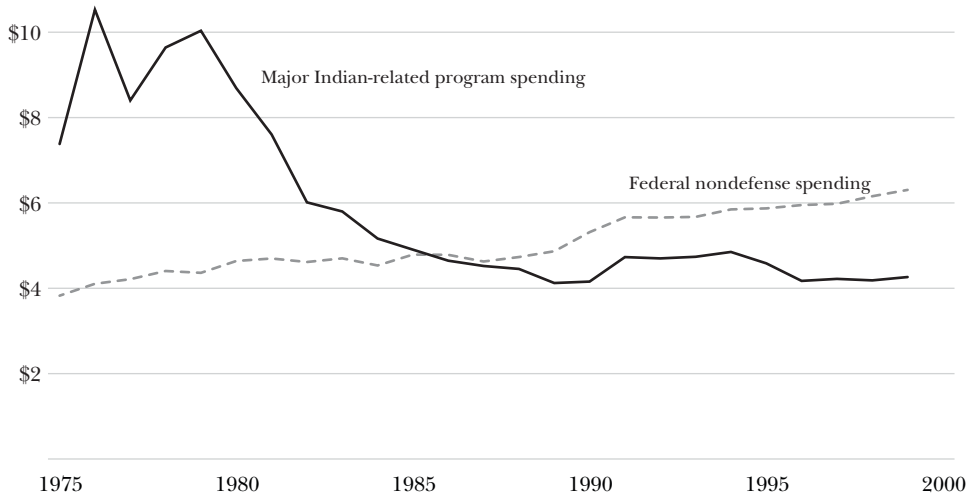
A number of obstacles to effective political rule and economic development help explain the persistence of reservation poverty. The historical legacy of Indian Country involves a loss of indigenous culture and language, the isolation of tribal communities on marginal lands, and the destruction of traditional tribal government structures (Cornell and Kalt 1995, p. 406). Potential investors confronted unfamiliar (or nonexistent) courts, laws, and commercial codes on American Indian reservations. Property interests were often unclear or held in federal trust, hindering transactions. In particular, inheritance rules often led to fractionated ownership, so that sometimes approval had to be sought from scores of owners—some of whom owned only a few square feet—before a property could be bought or sold (GAO 1992; Russ and Stratmann 2013).

Tribal governments were poorly equipped in the 1970s and 1980s to meet these challenges. Weak institutions of self-governance resulted in increased opportunism and corruption in some places. To make matters worse, tribal governments did not generally have the ability to raise revenues via taxation as most states and counties do (Fletcher 2004). For example, tribal governments cannot tax tribal lands held in trust by the federal government (McCullough 1994). Historically, issuing bonds was also prohibitively difficult (Clarkson 2007, p. 1015), although a few tribes have now managed to do so (Brashares and O'Keefe 2013).

Federal programs did not put things right. Expenditures in the "major programs affecting the nation's Indian population, particularly those programs targeting Indians in federally recognized tribes" totaled $4.4 billion in 1999 (Walke 2000), but as shown in Figure 1, this funding had decreased dramatically in the 1980s on a per capita basis (per service-eligible Indian), and did not keep pace with

[2] For example, amendments to the Clean Air Act and Clean Water Act explicitly established rules under which tribes can attain "treatment as state status" for making and enforcing environmental standards. More recently, Title XI of the Wall Street Reform and Consumer Protection Act of 2010 (better-known as the Dodd–Frank Act) defines tribes as states in the definition: "the term "State" means any State, territory, or possession of the United States . . . or any federally recognized Indian tribe, as defined by the Secretary of the Interior under section 479a-1(a) of title 25."

Figure 1
Federal Spending on Major Indian Programs per Capita
(thousands of 2014 dollars)



Source: Walke (2000); and FRED (2014) for deflating nominal dollars.
Notes: Per the Congressional Research Service (CRS), *Indian-related* includes program spending directed at "American Indian and Alaska Native tribes and their members because of their political status as Indians, not because of their racial classification or simply because they are citizens" (Walke 2000, p. 199). It includes the Bureau of Indian Affairs, the Office of Special Trustee for American Indians, the Indian Health Service (IHS), the Administration for Native Americans (Department of Health and Human Services), the Office of Indian Education (Department of Education), the Indian housing development program (Department of Housing and Urban Development), and the Indian and Native American Training Program (Department of Labor). The American Indian population denominator is the Indian Health Service's tabulation of service-eligible Indians—a population smaller than the nationwide American Indian and Alaska Native population but larger than the on-reservation population—both as recorded by the Census Bureau. *Federal nondefense* excludes both national defense expenditures and interest on the federal debt and is divided by intercensal population estimates (Walke 2000, p. 203, 207).

national per capita nondefense spending thereafter. The US Commission on Civil Rights (2003, p. iii) found federal spending for Indians "not sufficient to address the basic and very urgent needs of indigenous peoples." For example, per capita federal Indian healthcare spending was half what the federal government spent on prisoner health care at the time (p. 44).

Given these issues, external *and* internal investors often fled the scene (Cornell and Kalt 1998). The few extant instances of successful economic development in Indian Country were primarily confined to natural resource extraction industries and federal grant-funded projects. Tribes with confirmed treaty rights and large land bases were able to extract resource rents from low-cost, low-sulfur coal (Crow), old-growth timber (Warm Springs), hydropower (Salish & Kootenai), trophy elk (White Mountain Apache), and other resources. Tribes were sometimes able to move downstream: for example, they could collect fees on the right to harvest lumber or to use hydropower or coal, and then invest the proceeds in sawmills, power

plants, and other value-adding segments. Prior to vigorous self-determination, such resource development took place under federal supervision and was often limited in scale and efficiency (Krepps and Caves 1994, p. 134).

Tribal governments sought capital where they could, but often found that federal grants for economic development were the only viable option. Some tribes were able to build motels, industrial parks, and malls with federal grants. But such projects depended upon the grant-making trends of the day and were often poorly matched to competition, labor force, or demand (Cornell and Kalt 2007). These projects typically received only a single cycle of investment and left a swath of white elephants still visible in Indian Country.

Against this backdrop, some tribal governments asserted that they had the right to offer high-stakes bingo or legal card games on reservations in states where such activity was not expressly prohibited to everyone and that state and county gambling regulations did not apply on the reservation. Tribes in the vanguard sometimes sought and received federal approval of their gaming ordinances, as well as federal loans and loan guarantees to underwrite facilities: for an example, see Cattelino's (2008) discussion of the experience of the Seminole tribe in Florida.

Cabazon v. California and the Indian Gaming Regulation Act

As American Indian tribal governments began developing gaming establishments in the late 1970s and early 1980s, local and state officials asserted jurisdiction, and arrests and lawsuits followed. Several court decisions in the 1970s distinguished between criminal/prohibitory and civil/regulatory authority on American Indian reservations. For example, the US Supreme Court held in *Bryan v. Itasca County* (426 US 373 [1976]) that a state could not impose a tax on property (specifically, on a mobile home) located on an Indian reservation. As this legal doctrine evolved, the general rule emerged that if an activity is considered criminal and is prohibited by state laws, then those state prohibitions apply on Indian reservations in the 16 states where Congress had transferred criminal jurisdiction in the Termination Era under PL 83-280.[3] By contrast, if states merely regulate an otherwise legal activity—such as gambling—then the activity is a matter of civil regulatory authority and the state's jurisdiction does not generally extend onto Indian reservations. In 1982, the Supreme Court clarified this distinction when it declined to hear an appeal of a lower court ruling holding that Florida's gaming statute was civil/regulatory rather than criminal/prohibitory and therefore did not apply to the Seminole Tribe's high-stakes bingo operation (*Seminole Tribe of Florida v. Butterworth* 658 F. 2d 310 [US Court of Appeals, 5th Circuit 1981]).

[3] Six states were required by the act to assume jurisdiction over American Indians residing on reservations in their states: Alaska, California, Minnesota (except Red Lake), Nebraska, Oregon (except Warm Springs), and Wisconsin. Ten other opted to do so: Arizona, Florida, Idaho (subject to tribal consent), Iowa, Montana, Nevada, North Dakota (subject to tribal consent), South Dakota, Utah, and Washington (Goldberg, n.d.).

Across the country in southern California, the Morongo and Cabazon Bands built card room facilities that local and state governments sought to shut down, a controversy that eventually reached the US Supreme Court in *California v. Cabazon Band of Mission Indians* (480 US 202 [1987]). The federal government filed an *amicus* brief on behalf of the tribes in the *Cabazon* case, demonstrating that these businesses were supported by federal loans and loan guarantees, that the US Department of Interior had approved the tribal gaming ordinances, and that there was a significant federal interest in the success of these operations. The Court reasoned that because California's gambling laws in general were civil/regulatory—allowing charitable bingo nights and regulating card rooms, for example—rather than criminal/prohibitory, then state statutes could not be applied to tribal gaming operations. Moreover, the Court noted (p. 203):

> The federal interests in Indian self-government, including the goal of encouraging tribal self-sufficiency and economic development, are important, and federal agencies, acting under federal laws, have sought to implement them by promoting and overseeing tribal bingo and gambling enterprises. Such policies and actions are of particular relevance in this case since the tribal games provide the sole source of revenues for the operation of the tribal governments and are the major sources of employment for tribal members.

Thus, the Court ruled that the federal and tribal interests in tribal self-government and economic self-determination outweighed California's stated interest in preventing infiltration of tribal gaming by criminal elements. The state could also not forbid non-Indians from participating in high-stakes bingo and commercial card games on the reservation.

As the *Cabazon* claims wound toward the Supreme Court ruling in 1987, Congress began to discuss legislation that would apply to Indian gaming. The resultant Indian Gaming Regulatory Act passed in 1988. It created a National Indian Gaming Commission (NIGC) and established a three-class structure that delineated the roles of tribal, state, and federal governments. Class I gaming comprises traditional American Indian games of chance, which is considered social gambling for low stakes. Tribal governments regulate Class I exclusively, applying their own customs and traditions. Class II gaming encompasses bingo, pull-tabs, and nonbanked card games such as poker. Tribal governments and the NIGC jointly regulate Class II games, with tribal governments as the primary regulators. Finally, Class III gaming includes all other games, including house-banked card games and casino-style slot machines. Because the Class III games were perceived to be the biggest competitive threat to commercial casino jurisdictions and to hold the most potential to attract gambling customers, before a tribe can offer Class III gaming, it must negotiate a compact governing the scope and regulation of gaming with the state within whose borders the facility will be located.

Congress aimed to design an arrangement that would encourage states—some of which already possessed gaming regulatory expertise—to negotiate Indian gaming regulation in good faith, without diminishing tribal sovereignty or weakening tribal

bargaining power. While it might appear that states should have welcomed tribal gaming since it could potentially bring additional tax revenue, the law forbids states from requesting a share of tribal gaming revenue as a condition of signing a compact. The Indian Gaming Regulatory Act does allow tribal reimbursement of state regulation of Indian gaming and permits voluntary tribal contributions to local governments but does not allow revenue sharing or other indirect state taxation.

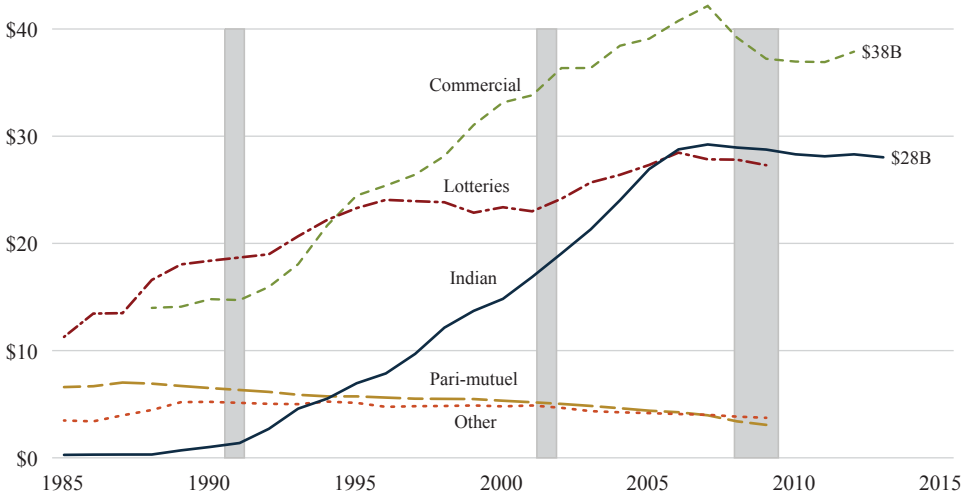
Of course, states could block Class III gaming entirely by refusing to agree to tribal government requests for compact negotiations, but the Indian Gaming Regulatory Act (IGRA) also allowed tribes to sue states for failing to negotiate in good faith. The most common reason a state would refuse to negotiate with a tribe was a disagreement on the permitted scope of gaming in the state, and this conflict delayed compacting for over a decade in a number of states, including California and Florida. However, the power of tribes to sue states under IGRA was ultimately ruled unconstitutional in *Seminole Tribe of Florida v. Florida* (517 US 44 [1996]), making ambiguity and litigation the order of the day in many states. Matters were further complicated in states like South Dakota that had substantial non-Indian gaming that would compete with tribes.

The negotiations between states and tribes over compacts to govern the scope of permitted gaming and the regulation of Class III gaming proceeded smoothly in some states and in some cases yielded results better than the tribes might have expected. In Michigan, for example, the state agreed to defer to tribal regulatory commissions so long as Indian casinos displayed signs explaining that Michigan did not regulate them (GAO 1998). The tribes of Minnesota and Mississippi negotiated compacts without an expiration date, virtually eliminating the "hold-up problem" that makes it more difficult to attract investment funds for casinos if the state leaves open the possibility of revisiting the compact in the future—a problem that continued to affect tribal casino development elsewhere. From 1991 to 1995, new compacts between tribes and states were successfully negotiated at a pace of about two dozen per year. By the end of the 1990s, compacts concerning Class III operations had been agreed for about 140 reservations that were home to about half of the population of American Indians living on reservations in 2000 (Taylor and Kalt 2005). As of 2010, reservations that were home to more than 90 percent of Indians living on reservations had gaming operations (Akee and Taylor 2014).

Among the tribes that have not signed a compact, some chose not to develop casinos for internal reasons such as religious or moral opposition to gaming industries. For instance, the Hopi Tribe has chosen repeatedly to reject casino development. Seneca, Navajo, and others chose not to pursue gaming compacts for a long period and then reversed course later. In some instances, tribes opened casinos and then closed them due to low consumer demand (for example, the Lummi Nation, the Hualapai Tribe, and the La Posta and Santa Ysabel Bands).

In states with permitted gaming, tribes could generally open Class II gaming operations without a compact. Class III gaming, however, involves a significant house advantage in card games and electronic gaming devices, more employment, and therefore more governmental revenue for tribes. These revenues are

Figure 2
Indian Gaming Revenues in Comparison to Other Sectors'
(billions of 2013 dollars)



Sources: National Indian Gaming Commission (2014); American Gaming Association (2014); International Gaming and Wagering Business (various years); Christiansen (1999); Christiansen (2001); National Bureau of Economic Research (2012); US Census (2011c); FRED (2014); GAO (1997).
Notes: "Lotteries" are state lotteries. "Pari-mutuel" wagering most commonly takes the form of horse and dog racing. "Other" includes charitable gaming, charitable bingo, legal bookmaking, and card rooms. The grey areas represent recessions.

the ultimate goal for many tribes. As the owners of the gaming facility, tribal governments generally earmark gaming revenues for specific tribal budget items, offsetting federal funding shortfalls across myriad programs. Tribal governments are obligated by the Indian Gaming Regulatory Act of 1988 to invest 100 percent of net gaming revenues in ways that improve tribal welfare. Section 11 of IGRA requires that net revenues from "any tribal gaming" be used for five primary purposes: 1) to fund tribal government operations or programs; 2) to provide for the general welfare of the Indian tribe and its members; 3) to promote tribal economic development; 4) to donate to charitable organizations; or 5) to help fund operations of local government agencies. Consistent with IGRA's requirements, tribal governments are investing gaming revenues into a variety of tribal programs and services (health, law enforcement, and education, to name a few) and promoting economic diversification in ways that seek to benefit tribal citizens.

In the aftermath of the 1988 legislation, Indian gaming revenues grew at a rapid pace, as shown in Figure 2. By 1992, the revenues from Indian gaming eclipsed charitable bingo and other charitable gambling (not independently displayed). Three years later, Indian gaming revenues overtook those of pari-mutuel wagering, which most commonly takes the form of horse and dog racing. In 2006, Indian gaming outpaced state lotteries. More recently, revenues have plateaued both for

commercial gaming and Indian gaming. At present, revenues from Indian gaming are roughly three-quarters of the size of commercial gaming.

While the tribal gaming industry has grown substantially, the political requirements imposed by the Indian Gaming Regulatory Act, specifically the tribal-state compact process, have meant that more than 25 years later, the tribal gaming industry has not grown to meet market demand in all locations. Tribal-state disputes have concerned the types of allowable games (Washington), demands for revenue sharing (New Mexico), the terms of intergovernmental gambling competition (South Dakota), and conflict over the permitted scope of games (Florida). Compacts in states like California and South Dakota placed binding constraints on the number of electronic gaming machines, and the experience of tribes nationwide suggests that tribes in those states could have developed bigger facilities earlier.

Perhaps the biggest constraint is that the Indian Gaming Regulatory Act required tribal governments to locate the facilities exclusively on tribal trust lands. While section 20 of IGRA specifies a process for tribal, state, and federal approval of gaming facilities on subsequently acquired lands (in recognition of the complex history of Indian land claims), it has proven arduous to do so. As of 2013, only eight tribes had applied for and received approval from the US Secretary of the Interior to have such lands taken into trust ownership status by the federal government for tribal government gaming.[4] Consequently, the geographic distribution of Indian gaming reflects the historic contingencies of American Indian land cessions and federal reservation-making, not the market demand for an early 21st century leisure industry.

As of year-end 2013, one commercial directory identified 468 open Indian gambling establishments in 31 states. Their sizes ranged from the Forest County Potawatomi Community's 780,000 square-foot Potawatomi Hotel & Casino in Milwaukee, Wisconsin, to very small travel-mart slot rooms of only a few hundred square feet (Casino City 2013). As the range in sizes implies, the ability of tribes to reach customers varies widely. The National Indian Gaming Commission (2014) publishes data on the distribution of tribal gaming revenue. For fiscal year 2013, the 252 tribal gaming facilities that earned $25 million or less represented 56 percent of all operations but only 7.4 percent of all Indian gaming revenue. By contrast, the 78 operations that took in $100 million or more represented 17 percent of the facilities but 71 percent of the sector's revenues. A skewed distribution is not surprising, arising as it does from access to urban population centers. It is similarly unsurprising that some populous reservations have large casinos (for example, the Gila River Indian Community in Chandler, Arizona, near Phoenix) and others have small ones (for example, the Pine Ridge Reservation in South Dakota). The converse is true too. Small reservation communities are located across the market spectrum; some have access to urban areas (the San Manuel Band in California) and some are in remote locations (the Campo Band in California).

[4] They are the Enterprise Rancheria of Maidu Indians (CA), Forest County Potawatomi Community (WI), Fort Mojave Indian Tribe (AZ, CA, NV), Kalispel Indian Community (WA), Kaw Nation (OK), Keweenaw Bay Indian Community (MI), Northern Cheyenne Tribe (MT), and Northfork Rancheria of Mono Indians (CA) (Hart 2014).

The Consequences of Gaming for Indian Nations

The effects of tribal gaming on American Indian nations have been profound. Kevin Washburn (2008), Assistant Secretary of Indian Affairs at the US Department of the Interior, has said, "Indian gaming is simply the most successful economic venture ever to occur consistently across a wide range of American Indian reservations." While there is considerable heterogeneity of results across different tribal communities, gaming has been welcome for the vast majority.

In contrast to grant-funded federal development efforts, Indian gaming yielded sustained revenues for almost all tribes that built facilities. This break with the past was possible for a number of reasons. First, tribes entered early in the gaming industry's growth cycle. Outside the state of Nevada and Atlantic City, New Jersey, only a few non-Indian governments had begun to allow gaming in the 1980s. Second, while a few regions witnessed multiple tribes introducing gaming, in many cases a given tribe might be the sole operator for miles. Third, tribes worked hard to capture margins by starting conservatively, sometimes with temporary buildings, to avoid overcapitalizing their businesses while assessing what was, in the early 1990s, a poorly understood opportunity. Fourth, tribes went to capital markets, retained attorneys, hired management consultants, and developed the facilities on their own initiative to exploit opportunities they themselves evaluated. Not all tribes succeeded. But in contrast to federally conceived, single-cycle, grant-funded investments in hotels, mini-malls, and other flavors-of-the-month, gaming development was self-determined and grew with internal consistency checks and market feedback.

One of the measures of achievement of the Indian Regulatory Gaming Act of 1988 is that many tribal governments now have an ample flow of revenues for the first time. Indian gaming revenues have allowed tribes to invest in new programs to address poverty and provide public goods.[5] One of the most common investments has been in education, including school construction (for example, Mille Lacs Ojibwe), college scholarships (for example, the Osage Nation 2015), and Native language revitalization programs (Cherokee). Tribes have developed "wrap-around services" to help their citizens get jobs and keep them (Sisseton-Wahpeton Oyate). Tribes have combined conventional, traditional Native religious and non-Indian religious treatment in drug rehabilitation programs (Taylor 2006). Improvements in tribal services have resulted from an increase in government resources and employment. As a result, tribes have reduced emergency response times from hours to minutes (at Gila River Indian Community, HPAIED 2008, p. 152). Tribes have invested in their cultural lives, specifically museums, ceremonial grounds (Kalispel) (Taylor 2006, p. 36), artifact repatriation (San Carlos Apache), and arts patronage. Services have increased dramatically across reservations. There have been improvements in elder care services (Tohono O'odham), foster care (Fond du Lac), policing

[5] Unless otherwise cited, the examples in this paragraph are drawn from the reports of Honoring Nations, an awards program for excellence in Native governance housed at Harvard's Kennedy School of Government (HPAIED 2014).

(Flandreau), endangered species management (Nez Perce), water quality (Sandia Pueblo), financial literacy (Umatilla), and public works (Lummi).

Tribal governments have also used the revenues from gaming to fund other economic development, based on the widely shared view that Indian gaming will not provide sustained economic growth indefinitely. Typically, the pattern begins with developing adjacent hotels, conference halls, amphitheaters, and other amenities that increase the drawing power and visit durations of gaming facilities. In many cases, tribes have invested in nearby retail businesses, outlet malls, and other businesses that take advantage of customer traffic. Finally, they turn toward more distinct sectors as varied as banking (Citizen Potawatomi Nation), commercial real estate (San Manuel), and federal facilities management (Winnebago), often redeploying the management experience gained in tribal gaming development.

The operation of tribal gaming facilities has also changed labor markets on reservations. Opening tribal gaming facilities increases the demand for both high- and low-skill labor on the reservation. New employment opportunities exist in management and professional positions in the gaming and tourism industries. Over time, tribes have replaced external executives with internal tribal members in those management positions as citizens have gained relevant experience and education in the industry. Cozzetto (1995) found a decline in Indian welfare dependence coincident with gaming facility openings. Others have found that a substantial fraction of American Indian employees of tribal gaming come from the ranks of the unemployed (Cornell, Kalt, Krepps, and Taylor 1998). As programs and government services have grown, so too has tribal government employment. In the past 20 years, the proportion of American Indians on reservations employed in public service (including tribal government employment) has increased by 5 percentage points, a 20 percent increase (Akee and Taylor 2014). A similar increase is not observed in other sectors of the tribal economy, nor is this duplicated in the non-Indian population in the same time period. It is also important to note that the number of gaming management contracts (often with external, non-Indian casino companies) has decreased over time, indicating that tribal employees are now managing tribal enterprises. No new external management contracts have been approved by the National Indian Gaming Commission since 2010 (National Indian Gaming Commission 2015). For instance, the San Pasqual Mission Band of Indians bought out their five-year management contract after just one year and began to manage gaming operations with their own hires (Contreras 2005), a pattern that repeats across Indian Country.

Tribal gaming affects local and regional migration patterns as well. Tribal member income and employment have increased (Reagan and Gitter 2007) and therefore helped to stop or reverse "brain drain" off of the reservation. Improving economic opportunities appear to have brought return migration as well. In the first decade after the Indian Gaming Regulatory Act of 1988, there was an increase in tribal populations (Evans and Topoleski 2002). American Indians increasingly view their tribal governments as capable of creating desirable places to live and work. There are instances of interest rates falling when these revenue-generating tribal governments choose to borrow, as well. The Squaxin Island Tribe north of

Olympia, Washington, for example, found that its cost of capital dropped several percentage points after the introduction of gaming operations (Taylor 2006, p. 44).

Reservation life has improved in measurable ways in the wake of tribal gaming. There was a relatively large convergence in the average conditions of American Indians on reservations towards that of the rest of the US population in the 1990s, as shown earlier in Table 1. Convergence continued, though more moderately in the 2000s. Real per capita income earned by Indians living on reservations in the contiguous 48 states grew by 33.3 percent in the 1990s (compared to the national average of 11.4 percent) and by 11.5 percent over the 2000s (compared to the national average of −3.3 percent). Consistent gains were made over the 1990–2010 period for educational attainment, income, and female labor participation, accompanied by similar reductions in poverty and overcrowded homes. In most instances, improvements on Indian reservations outpaced national changes over the period. Larger gains were observed for those reservations operating a casino or bingo hall by 2000 (Taylor and Kalt 2005).

Some tribal governments—typically ones without very large populations—have distributed a percentage of their gaming revenues to citizens. These per capita disbursements typically take the form of annual or semi-annual checks sent directly to tribal members above the age of 18 (or held in escrow for minors). As of 2009, 120 tribes had filed revenue allocation plans with the Bureau of Indian Affairs, a prerequisite under the Indian Gaming Regulatory Act for tribes' allocating revenue per capita in this way (Taggart and Conner 2011). The amounts distributed may vary according to the revenue in a given year. The total amount of payments is not typically disclosed publicly; however, several tribal governments announce the size of their payments, which range from a few hundred to thousands of dollars per person annually. This change in household income can have profound effects on previously poverty-stricken households. Cornell et al. (2008) provide an overview of determining eligibility and other issues confronting tribes that make these kinds of per capita payments.

A few empirical studies have examined the effects of the per capita income disbursements or casino operations on American Indian populations and adjacent non-Indian communities. For example, Akee, Copeland, Keeler, Angold, and Costello (2010) found that an increase in unearned income from per capita payments resulted in increased educational attainment for children in poverty-stricken households. For each additional $1,000 in unearned income at the household level, there was an increase of about 6 to 7 percentage points in high school graduation rates for children from previously poverty-stricken households. Additionally, American Indian children in households with higher incomes due to the per capita transfer payments attended school about four more days per quarter. In related work examining the effect of casino operations on American Indians, Wolfe, Jakubowski, Haveman, and Courey (2012) found that casino operations are correlated with decreases in smoking by 9 percent, in heavy drinking by 5 percent, and in obesity by 2.7 percent. Evans and Topoleski (2002) found that reservations with gaming experienced increases in employment of about 26 percent and an increase in population size of about 11 percent, four or more years after casino operations began.

Although the vast majority of empirical research on Indian gaming has found benefits to those living on or near the reservations, Indian casinos have been associated with controversial and even deleterious effects in some communities. Tribal governments vary in their capacity to withstand political division, to administer programs effectively, and to produce public goods that their citizens want. One controversial outcome has been the disenrollment of tribal citizens, which has resulted in significant conflicts in a number of American Indian communities (Gonzales 2003). Reducing the size of the tribal population can potentially benefit existing tribal members if there are per capita distributions of gaming revenues. Fights over control of the gaming facility itself have also accentuated factional division in Indian communities leading, in extreme cases, to standoffs (Picayune Rancheria) and even constitutional crises (Winnebago of Wisconsin). On occasion, casino competition has intensified intertribal conflict, especially where off-reservation casinos are proposed. For example, in November 2014 California citizens voted against Proposition 48, which would have ratified a tribal-state gaming compact for the Northfork tribe to open a gaming facility away from its reservation land but closer to population centers. Some of the opposition came from other tribes whose facilities would have faced heightened competition from the proposed new facility.

Finally, it should be noted that for all the good news coming from Indian Country since the passage of the Indian Gaming Regulatory Act of 1988, the accumulated economic and social deficits on reservations are so large that even if Indian income growth keeps its pace, it will take decades for American Indians to close the gap with the average American (Taylor and Kalt 2005, p. 7; Akee and Taylor 2014, p. 36). Indeed, given that standards of living in the United States are recovering from the Great Recession and given that there is no apparent successor to gaming waiting in the wings for Indian Country, it will remain critical for tribal policymakers to get other aspects of development right.

Consequences for State and Local Economies

During the late 1980s, at the time of the *Cabazon* decision and the debates over the Indian Gaming Regulatory Act, state and local governments expressed concerns that Indian gaming facilities would produce negative externalities in two broad categories. First, it was argued that rising visitation to the reservations would have an adverse impact on local governments' infrastructure and services, clogging highways, overloading emergency services, or overtaxing waste treatment facilities. Second, it was argued, Indian gaming facilities would market an inherently risky product—gambling—which would have negative social impacts in host communities such as bankruptcy, organized crime infiltration, disordered gambling, drug abuse, suicide, and other social ills.

The Indian Gaming Regulatory Act contained explicit provisions to address potential adverse effects of the tribal gaming industry. Among five sanctioned uses of net tribal gaming revenues are: "to donate to charitable organizations" and "to help

fund operations of local government agencies" (25 USC §2710(b)(2)(B)). In addition, IGRA envisioned that tribes could reimburse states' regulatory costs (25 USC §2710(d)(3)(C)(iii)). Indeed, many state-tribal compacts have clauses governing payments for local impact mitigation or regulatory reimbursement clauses. A number of state-tribal compacts also have clauses governing investment in responsible gaming initiatives, including corporate and tribal policies and procedures that help prevent or ameliorate the consequences of disordered gambling (for definitions, see National Center for Responsible Gaming 2011, p. 3). Broadly speaking, IGRA and its compacting process encourage cooperation in the production of intergovernmental public goods. Comprehensive or national-level research about the relationship between tribes and local governments is thin. However, the available evidence does not suggest that the early fears of state and local government have been borne out.

For example, what of the initial fears related to the social costs of disordered gambling behavior resulting from increased access to gambling through the expansion of Indian gaming? Empirical research of gambling pathology has failed to identify large net costs. For example, a 16-year, 100-community randomized multilevel regression performed by the National Opinion Research Center (NORC) at the University of Chicago for the National Gambling Impact Study Commission found that when a casino is opened, communities near the casinos experienced reductions in unemployment (one percentage point), some changes in wage distribution across sectors, and no discernible change in bankruptcy, crime, or infant mortality (Johnson 1999). For comparison, NORC calculated that the national annual costs of problem and pathological gambling, $5 billion in 1998, were 3 percent of the estimated $166.5 billion in annual national costs for alcohol abuse (Gerstein et al. 1999, p. 53). Himmelstein, Warren, Thorne, and Woolhandler (2005, p. 67) found that about half of all bankruptcy filers cited medical emergencies as a contributing cause, whereas uncontrolled gambling was listed as a contributing cause by only 1 percent of bankruptcy filers.

Indeed, some research at the state level reveals that newly expanded opportunities to gamble offer casino guests access to information about problem gambling that they previously lacked, while having little long-term effect on the prevalence of problem gambling. A study in California found that between 1990 and 2006, when more than 40 new tribal facilities opened in the state, California experienced a reduction in gambling participation generally (Volberg, Nysse-Carris, and Gerstein 2006, p. 54). This finding is not all that unexpected once one considers that access to other forms of gambling in the state, including the lottery, card rooms, and horse racing, existed in 1990, along with proximity to full-scale gambling in neighboring Nevada. The report finds that "[based] on the survey data, it is possible to compare lifetime participation rates for several gambling activities in 1990, 1999 and 2006 . . . Casino gambling increased slightly between 1990 and 1999 but then decreased between 1999 and 2006" (p. 53). This decline in participation rates and duration reflect what is known as the "novelty effect," wherein gamblers are initially drawn to a new gambling product or service but their overall participation then reverts to the mean over time.

Another claim often made by state and local governments against Indian gaming argues that Indian casinos diminish state and local tax collections (Washington Research Council 2002; Anders, Siegel, and Yacoub 1998). Much of the empirical support for the claim remains unpersuasive.[6] After all, reservation economic activity requires goods and services from off-reservation communities, which incur local and state taxes on sales and income. Survey data from Washington State tribes, for example, indicate that two-thirds of the 27,376 workers employed in tribal casinos, governments, and nongaming enterprises in 2010 were non-Indians (Taylor 2012). Detailed procurement information from four of those tribes indicates that at least 94 percent of all tribal goods and services in 2004 came from off-reservation suppliers (Taylor 2006). Thus, even when consumer spending shifts from off-reservation (and state-taxable) restaurants, movie theaters, and bars to Indian casinos, spas, and hotels, the overall effects on input markets may be negligible. Indian gaming may cause a shift in spending patterns, but it is likely that state revenue from taxes on input labor, goods, and services would be virtually unchanged. In one study, Taylor (2005) found no discernible effect of the introduction of casinos on taxable sales and property in the state of Washington for 268 communities over 13 years.

Moreover, tribe–state gaming compacts often contain revenue-sharing provisions. Although state insistence on tax revenue or revenue-sharing as a condition of compact approval was prohibited by the Indian Gaming Regulatory Act, the US Secretary of the Interior has approved compacts with revenue-sharing provisions under the condition that the states contribute to the economic value of the tribe's facilities in some way (Martin 2003). Such contributions range from giving tribes statewide casino exclusivity (for example, Mashantucket Pequot and Mohegan in Connecticut) to deploying condemnation powers to allow a tribe to purchase property for their business and selling a state-owned convention center for $1 (Seneca in New York). Such terms make states quasi-joint venture partners—contributors to and beneficiaries of Indian gaming development. Over the years, such revenue flows have in certain places been substantial, for example: $1 billion in 11 years to Arizona (Arizona Department of Gaming 2014), and $6.7 billion in 22 years to Connecticut (Connecticut Department of Consumer Protection 2014). In 2012, nationwide Indian gaming revenue sharing with states was estimated to be $1.5 billion (Meister 2014).

In addition to direct fund transfers, nearby off-reservation communities also benefit from Indian gaming's economic spillovers—spillovers that may exceed those of commercial gaming for at least three reasons. First, in many places, Indian gaming attracts customers from further away than more competitively distributed amenities, making Indian gaming facilities net contributors to the local or regional economies, all else equal. Oklahoma's Indian gaming, for example, recruits customers heavily

[6] As one example, an article on the subject mistook Maricopa County (Arizona) tax revenue declines coincident with tribe-state *compacting* for the effects of Indian casino *openings* (Anders, Siegel, and Yacoub 1998). The examples in the analysis actually pre- and post-date a purported casino-driven fall in revenue by many months and appear, by the paper's own data, to have actually left contemporaneous Maricopa County revenue undisturbed (Taylor 2005).

from neighboring Texas—which does not have casinos. The opening of Seneca Niagara Falls Casino at year's end 2002 coincided with precipitous revenue decline across the border in Ontario (Gardner 2005; Niagara Falls Canada 2006), as western New Yorkers pulled leisure spending back from Canada. Even within state borders, destination effects can be pronounced. Second, Indian gaming often takes place in poorer-than-average regions of the country—not just the reservations are poorer, but the surrounding counties, too. In such regions, chances are better that underutilized resources, especially labor, see net gains in utilization, with larger consequences for the regional economy. Third, the investment of tribal gaming revenue is geographically restricted to the tribe's governing jurisdiction rather than distributed wherever in the global economy a commercial casino company's shareholders might be.

Evidence on these effects is accumulating. In one study, the presence of an Indian casino in an adjacent California county was associated with greater real median family income growth from 1990 to 2000 (Center for California Native Nations 2006). A follow-up to that parsimonious difference-in-difference analysis found a diminished but persistent effect in the subsequent decade (Akee, Spilde, and Taylor 2014). Evans and Kim (2006) found that Indian casinos reduced unemployment and increased wages for low-skilled workers. A re-examination of the National Opinion Research Center (NORC) study discussed above (Johnson 1999), which examined more closely the counties proximate to Indian gaming introductions, found that the effects were more positive than those of commercial non-Indian casinos and that those counties had a reduced reliance on welfare (Taylor, Krepps, and Wang 2000).

Indian gaming often does attract funds that could have been spent on entertainment at other casinos or on nongaming leisure activity. But of course, the same can be said of a wide variety of entertainment-related destinations. One would not want to overstate the social welfare benefits of Indian gaming by treating every job in the industry or every dollar of revenue flowing to the tribes as an addition to social welfare. But neither would an economist argue that an entertainment venue has zero social benefit on the grounds that the entertainment dollars could have been spent somewhere else. The true gain to social welfare, of course, lies somewhere in-between.

Where Indian gaming development increases unreimbursed infrastructure burdens on surrounding governments, such costs are the consequences of growth in regional economic activity, the state taxation of which would at least partially rectify the harm. Of course, the degree to which incremental taxes exceed, meet, or fall short of the burden depends upon the tribal-state compact terms governing local impact mitigation and revenue sharing, intrastate fiscal allocation mechanisms, and the attributes of the burden itself. The Indian Gaming Regulatory Act has specific clauses that allow for the reimbursement of non-Indian infrastructure burdens under the terms of the state-tribal compact. On the other hand, there may be adverse effects for other leisure activities and businesses in a region. As gaming operations begin in a region, consumers may shift their leisure spending towards the new, previously unavailable gaming activities. Assessing whether the overall benefits to consumer surplus from the introduction of a new leisure activity

outweigh potential losses to other pre-existing leisure activity businesses has not been adequately examined.

Conclusion

Indian gaming is no longer in its infancy. Indian tribes will face new competitors as state-sanctioned casinos continue to spread. As Eadington pointed out in this journal (1999, p. 190), overall casino gambling as an industry has been undergoing a long progression from concentrated availability in Las Vegas and Atlantic City to dispersed localities around the country. Technological change is now raising the possibility of online gaming operations that may rival or complement brick-and-mortar operations. These changes mean that the days of regional exclusivity for a large number of Indian gaming operations are probably numbered, and so too are the days of build-it-and-they-will-come operations.

In the years ahead, tribal governments will face stronger incentives to improve tribal gaming performance. At various times and places, certain Indian gaming facilities have faced competitive pressures that have been severe (Ohkay Owingeh), devastating (Penobscot), and unsustainable (Lummi). Tribes will benefit from research exploring these cases and generally explaining the variation observed in casino performance. Market access to large numbers of nearby customers is a first-order explanation, of course, but beyond that governance quality, management abilities, amenity diversity, and service quality all play a role.

Tribal incentives to diversify the nongaming aspects of their governance and economies will strengthen, too. The low-hanging fruits of self-administration—such as correcting principal-agent slippage in federal timber management (Krepps and Caves 1994)—may already have been harvested in many places. Likewise, tribes may have already reaped the bulk of the benefits of tailoring federal programs to local needs and conditions. Tribal leaders increasingly confront the politically difficult work of cutting underperforming programs, improving performance from tribal agencies, and reducing popular budget items. Tribally owned enterprises face the challenges that government-owned businesses face around the globe (Grant and Taylor 2007). Native fertility is higher than for Americans generally (US Census 2011b), and to reverse the incentives for emigration from tribal areas, tribal governments will need both to diversify the tribally owned sector and to develop policies that encourage private business formation and recruitment on the reservations as well (Cornell, Jorgensen, Record, and Timeche 2007).

While commercial casino gaming is spreading to new jurisdictions across the United States, it is not clear that this type of gaming expansion will bring the pronounced social and economic development benefits that tribal gaming brings to communities that are on or near tribal lands. The requirements under the Indian Gaming Regulatory Act of 1988 that tribal gaming facilities be owned by tribal governments and that revenues be invested in the general welfare of the community and take place on tribal trust lands has resulted in an intense and particularly local concentration of tribal gaming's benefits that may be difficult to replicate.

The requirements of the Indian Gaming Regulatory Act have triggered the development of tribal institutions too. For example, IGRA requires tribes to establish independent gaming commissions for licensing casino personnel and regulating gaming facilities. National Indian Gaming Commission regulations further specify minimum internal control standards governing cash-handling and customer blandishments. On their own initiative, tribal governments have added to these mandatory structures and created independent boards that separate the governance of the tribal polity from that of tribal businesses, and many have promulgated policies that handle everything from personnel disputes to budgeting, appropriating, and investing tribal gaming revenues. A steady flow of gaming revenues also loosened a tight liquidity constraint holding back the development of institutions unrelated to gaming operations. For example, the Tulalip Tribes north of Seattle were able to take back criminal jurisdiction from the state of Washington by developing competent judicial, policing, and prosecutorial staffs. The Osage and Citizen Potawatomi Nations of Oklahoma (and many others) have modernized their constitutions. Moreover, the preponderance of tribal programs winning Harvard's Honoring Nations awards for excellence in tribal governance have been created by tribes that operate gaming facilities. Most such reforms and innovations might not have been accomplished as quickly or successfully (or at all) without gaming revenues for salaries and professional services.

It is also the case that on a few reservations, gaming revenues have raised the stakes of internal political conflict, straining to the breaking point the weak political institutions bequeathed by historical federal policies. Some tribes have emerged from such crises with stronger constitutions (for example, the Ho-Chunk Nation in Wisconsin), but tribes have also been deeply riven by disenrollment controversies and constitutional crises. Generally, we see that institutional reforms and programmatic innovation are the norm and deleterious crises the exception, but more systematic research is needed to link gaming and institutional change.

There continues to be a great need for research on the impact of the gaming industry on long-run outcomes for American Indians. Evaluations of gaming are typically general in scope, not focused on Indian gaming in particular (for example, Grinols 2004; Walker 2007; Eadington 1999). How are the spread of Indian gaming and the rise in local incomes related to factors such as Native family composition, indigenous language proficiency, reservation brain drain, or expectations and beliefs about the future? After nearly three decades of additional investments in educational and social programs, what lessons can we extract for socioeconomic recovery in other Native and non-Native populations (Besaw et al. 2004)? A generation of American Indians born after the 1987 *Cabazon* decision and the passage of the Indian Gaming Regulatory Act of 1988 is coming of age. Indian gaming has profoundly changed the trajectories of many individual lives and the patterns of economic development on American Indian reservations. Additionally, it has laid the institutional foundation for sustained change and provided an environment across Indian Country that is attractive for investment of capital and human resources, in some cases for the first time in generations.

References

Akee, Randall K. Q., William E. Copeland, Gordon Keeler, Adrian Angold, and E. Jane Costello. 2010. "Parents' Incomes and Children's Outcomes: A Quasi-Experiment Using Transfer Payments from Casino Profits." *American Economic Journal: Applied Economics* 2(1): 86–115.

Akee, Randall K. Q., Katherine A. Spilde, and Jonathan B. Taylor. 2014. "Social and Economic Changes on American Indian Reservations in California: An Examination of Twenty Years of Tribal Government Gaming." *UNLV Gaming Research & Review Journal* 18(2).

Akee, Randall K. Q., and Jonathan B. Taylor. 2014. *Social and Economic Change on American Indian Reservations: A Databook of the US Censuses and American Community Survey, 1990–2010.* Sarasota, FL: Taylor Policy Group, Inc. Available at: http://taylorpolicy.com/us-databook/.

American Gaming Association. 2014. "State of the States." Available at: http://www.americangaming.org/industry-resources/research/state-states.

American Indian Lawyer Training Program. 1991. *Indian Tribes as Sovereign Governments: A Sourcebook on Federal-Tribal History, Law, and Policy.* Oakland, CA: American Indian Resources Institute Press.

Anders, Gary C., Donald Siegel, and Munther Yacoub. 1998. "Does Indian Casino Gambling Reduce State Revenues? Evidence From Arizona." *Contemporary Economic Policy* 16(3): 347–55.

Arizona Department of Gaming. 2014. "Tribal Contributions from Gaming Revenue to the State, Cities, Towns and Counties as of October 31, 2013. *Arizona Department of Gaming.* Retrieved December 3, 2014, from https://gaming.az.gov/sites/default/files/documents/files/cumulativetc10-31.pdf.

Besaw, Amy, Joseph P. Kalt, Andrew Lee, Jasmin Sethi, Julie Boatright Wilson, and Marie Zemler. 2004. *The Context and Meaning of Family Strengthening in Indian America: A Report to the Annie E. Casey Foundation.* Cambridge, MA: Harvard Project on American Indian Economic Development.

Brashares, Edith, and Siobhan O'Keefe. 2013. "Indian Tribal Government Access to Tax-Exempt Bond Financing." Available at SSRN: http://ssrn.com/abstract=2370079.

Bush, George H. W. 1991. "Statement by the President Reaffirming the Government-to-Government Relationship Between the Federal Government and Tribal Governments." June 14. Office of the Press Secretary, The White House, Washington D.C.

Bush, George W. 2004. "Memorandum for the Heads of Executive Departments and Agencies: Government-to-Government Relationship with Tribal Governments." September 23. The White House, Washington, DC.

Carlson, Leonard A. 1981. *Indians, Bureaucrats, and Land: The Dawes Act and the Decline of Indian Farming.* Contributions in Economics and Economic History. Westport, Conn.: Greenwood Press.

Carter, Jimmy [James E.]. 1979. "The State of the Union Address Delivered Before a Joint Session of the Congress." January 23. http://www.presidency.ucsb.edu/ws/?pid=32657.

Casino City. 2013. "Gaming Directory.com." Casino City Press. http://www.gamingdirectory.com/.

Cattelino, Jessica R. 2008. *High Stakes: Florida Seminole Gaming and Sovereignty.* Duke University Press Books.

Center for California Native Nations. 2006. *An Impact Analysis of Tribal Government Gaming in California.* University of California, Riverside.

Christiansen, Eugene Martin. 1999. "The Gross Annual Wager of the United States 1998." Manuscript for *International Gaming and Wagering Business.*

Christiansen, Eugene Martin. 2001. "The Gross Annual Wage of the United States 2000." Manuscript for *International Gaming and Wagering Business.*

Clarkson, Gavin. 2007. "Tribal Bonds: Statutory Shackles and Regulatory Restraints on Tribal Economic Development." *North Carolina Law Review* 85(4): 1009–85.

Clinton, William J. 1994. "Government-to-Government Relations with Native American Tribal Governments." Memorandum, April 29. *Federal Register* 59(84).

Clinton, William J. 2000. "Executive Order 13175: Consultation and Coordination with Indian Tribal Governments." *Federal Register* 65(218).

Connecticut Department of Consumer Protection. 2014. "DCP - Gaming Division. Transfers to General Fund, Accumulative to Date - Through June 2014." August 8. A table. DCP - Gaming Division, Hartford, CT. Retrieved December 3, 2014, from http://ct.gov/dcp/lib/dcp/pdf/gaming/stmt2014.pdf.

Contreras, Michael. 2005. "Back In the Driver's Seat: San Pasqual Tribe Now Running Valley View Casino." *Indian Gaming* 15(3).

Cornell, Stephen. 1988. *The Return of the Native: American Indian Political Resurgence.* Oxford University Press.

Cornell, Stephen, Miriam Jorgensen, Stephanie Carroll Rainie, Ian Record, Ryan Seelau, and Rachel Rose Starks. 2008. "Per Capita Distributions of American Indian Tribal Revenues: A

Preliminary Discussion of Policy Considerations." *Joint Occasional Papers on Native Affairs*, no. 2008-02. http://www.nni.arizona.edu/whatsnew/pdfs/jopna_2008-02percapita.pdf.

Cornell, Stephen, Miriam Jorgensen, Ian Record, and Joan Timeche. 2007. "Citizen Entrepreneurship: An Underutilized Development Resource." In *Rebuilding Native Nations: Strategies for Governance and Development*, 197–222. Tucson, AZ: University of Arizona Press.

Cornell, Stephen, and Joseph P. Kalt. 1992. "Reloading the Dice: Improving the Chances for Economic Development on American Indian Reservations." In *What Can Tribes Do? Strategies and Institutions in American Indian Economic Development*, edited by Stephen Ellicott Cornell and Joseph Peggs Kalt, 1–59. Los Angeles, CA: American Indian Studies Center, University of California Los Angeles.

Cornell, Stephen, and Joseph P. Kalt. 1995. "Where Does Economic Development Really Come From? Constitutional Rule Among the Contemporary Sioux and Apache." *Economic Inquiry* 33(3): 402–26.

Cornell, Stephen, and Joseph P. Kalt. 1998. "Sovereignty and Nation-Building: The Development Challenge in Indian Country Today." *American Indian Culture and Research Journal* 22(3): 187–214.

Cornell, Stephen, and Joseph P. Kalt. 2007. "Two Approaches to the Development of Native Nations: One Works, the Other Doesn't." In *Rebuilding Native Nations: Strategies for Governance and Development*, edited by Miriam Jorgensen, 3–32. Tucson, AZ: University of Arizona Press.

Cornell, Stephen, Joseph P. Kalt, Matthew B. Krepps, and Jonathan B. Taylor. 1998. *American Indian Gaming Policy and Its Socioeconomic Effects: A Report to the National Gambling Impact Study Commission*. Cambridge, MA: The Economics Resource Group, Inc.

Cozzetto, Don A. 1995. "The Economic and Social Implications of Indian Gaming: The Case of Minnesota." *American Indian Culture and Research Journal* 19(1): 119–31.

Eadington, William R. 1999. "The Economics of Casino Gambling." *Journal of Economic Perspectives* 13(3): 173–92.

Evans, William N., and Julie H. Topoleski. 2002. "The Social and Economic Impact of Native American Casinos." NBER Working Paper 9198.

Evans, William, and Wooyoung Kim. 2006. "The Impact of Local Labor Market Conditions on the Demand for Education: Evidence from Indian Casinos." Working Paper 06-14, U.S. Census Bureau, Center for Economic Studies.

Fletcher, Matthew L. M. 2004. "In Pursuit of Tribal Economic Development as a Substitute for Reservation Tax Revenue." *North Dakota Law Review* 80(4).

FRED. 2014. Consumer price index for all urban consumers. *FRED Economic Data*, 2014, August 1. Federal Reserve Bank of St. Louis. Retrieved October 10, 2014, from http://research.stlouisfed.org/fred2/series/CPIAUCSL.

GAO. 1992. "Indian Programs: Profile of Land Ownership at 12 Reservations." Briefing Report to the Chairman, Select Committee on Indian Affairs, U.S. Senate. Washington, DC: US General Accounting Office.

GAO. 1997. *Tax Policy: A Profile of the Indian Gaming Industry*. Washington, DC: US General Accounting Office.

GAO. 1998. *Casino Gaming Regulation: Roles of Five States and the National Indian Gaming Commission*. Washington, DC: US General Accounting Office.

GAO. 2004. *Impact of Gambling: Economic Effects More Measurable than Social Effects*. Washington, DC: US General Accounting Office.

Gardner, Kent. 2005. *Seneca Niagara Casino: Fiscal & Economic Impact on Niagara Falls, NY*. Rochester, NY: Center for Governmental Research, Inc.

Gerstein, Dean, et al. 1999. *Gambling Impact and Behavior Study: Report to the National Gambling Impact Study Commission*. Chicago, IL: National Opinion Research Center at the University of Chicago.

Goldberg, Carole. No date. "Questions and Answers about Public Law 280." *Tribal Court Clearinghouse*. Retrieved May 29, 2015, from http://www.tribal-institute.org/articles/goldberg.htm.

Gonzales, Angela A. 2003. "Gaming and Displacement: Winners and Losers in American Indian Casino Development." *International Social Science Journal* 55(175): 123–33.

Grant, Kenneth W., and Jonathan B. Taylor. 2007. "Managing the Boundary Between Business and Politics: Strategies for Improving the Chances for Success in Tribally Owned Enterprises." In *Rebuilding Native Nations: Strategies for Governance and Development*, edited by Miriam Jorgensen, 175–96. Tucson, AZ: University of Arizona Press.

Grinols, Earl L. 2004. *Gambling in America: Costs and Benefits*. Cambridge University Press.

Hart, Paula. Letter. 2014. "Section 20 Approvals," October 7.

Himmelstein, David U., Elizabeth Warren, Deborah Thorne, and Steffie Woolhandler. 2005. "MarketWatch: Illness and Injury as Contributors to Bankruptcy." *Health Affairs*. Published online before print. doi:10.1377/hlthaff.W5.63.

Harvard Project on American Indian Economic Development (HPAIED). 2008. *The State of the Native Nations: Conditions under U.S. Policies of Self-Determination*. New York, NY: Oxford University Press.

Harvard Project on American Indian Economic Development (HPAIED). 2014. "About Honoring Nations." (Honoring Nations is "An awards program that identifies, celebrates, and shares outstanding examples of tribal governance.) From the HPAIED website: http://hpaied.org/index.php?option=com_content&view=article&id=7&Itemid=8.

International Gaming and Wagering Business. Various years. "Gross Annual Wages." *International Gaming and Wagering Business*.

Johnson, Robert. 1999. "Chapter 5. Impacts of Casino Proximity on Social and Economic Outcomes, 1980–1997: a Multilevel Time-Series Analysis." In *Gambling Impact and Behavior Study: Report to the National Gambling Impact Study Commission*, 65–72. Chicago, IL.

Krepps, Matthew B., and Richard E. Caves. 1994. "Bureaucrats and Indians: Principal-Agent Relations and Efficient Management of Tribal Forest Resources." *Journal of Economic Behavior & Organization* 24(2): 133–51.

Marks, Mindy, and Kate Spilde Contreras. 2007. "Lands of Opportunity: Social and Economic Effects of Tribal Gaming on Localities." *Policy Matters* 1(4), Summer.

Martin, Aurene M. 2003. "Statement of Aurene M. Martin, Acting Assistant Secretary, Indian Affairs, Department of the Interior." July 9, 2003. In *Oversight Hearing on Indian Gaming Regulatory Act; Role and Funding of the National Indian Gaming Commission, One Hundred Eighth Congress First Session, Part 2*, 89–154. Committee on Indian Affairs, United States Senate.

McCulloch, Anne M. 1994. "The Politics of Indian Gaming: Tribe/State Relations and American Federalism. *Publius: The Journal of Federalism* 24(3): 99–111.

Meister, Alan P. 2014. *Casino City's Indian Gaming Industry Report*. 2015 edition. Newton, MA: Casino City Press.

Meriam, Lewis, Ray A. Brown, Henry Roe Cloud, Edward Everett Dale, Emma Duke, Herbert R. Edwards, Fayette Avery McKenzie, Mary Louise Mark, W. Carson Ryan, and William J. Spillman. 1928. *The Problem of Indian Administration: Report of a Survey Made at the Request of Honorable Hubert Work, Secretary of the Interior, and Submitted to Him, February 21, 1928*. Baltimore, MD: Johns Hopkins Press.

National Bureau of Economic Research. 2012. "US Business Cycle Expansions and Contractions." *NBER.org*. Cambridge, MA. http://www.nber.org/cycles.html.

National Center for Responsible Gaming. 2011. "A Guide to Gambling Disorders and Responsible Gaming." http://www.ncrg.org/sites/default/files/uploads/docs/researchandresources2011.pdf.

National Indian Gaming Commission. 2014. Gaming Revenue Reports. http://www.nigc.gov/Gaming_Revenue_Reports.aspx.

National Indian Gaming Commission. 2015. "Approved Management Contracts." http://www.nigc.gov/Reading_Room/Management_Contracts/Approved_Management_Contracts.aspx.

National Research Council. 1999. *Pathological Gambling: A Critical Review*. Washington, DC: National Research Council.

Niagara Falls Canada. 2006. *Tourist and Economic Impact Statistics*. Niagara Falls, ON: Business Development Department, Niagara Falls.

Nixon, Richard M. 1970. "Special Message to Congress on Indian Affairs." *Public Papers of the Presidents of the United States*, July 8.

Obama, Barack H. 2009. "Memorandum for the Heads of Executive Departments and Agencies." *Federal Register* , November 9, 74(215): 57879–82.

Osage Nation. 2015. "Education Scholarship." https://www.osagenation-nsn.gov/what-we-do/education-department/education-scholarship (retrieved May 29, 2015).

Osier, Vince. 2012. "American Indian and Alaska Native Area to Tribe 2012." Washington, DC: US Census Bureau.

Reagan, Patricia B, and Robert J. Gitter. 2007. "Is Gaming the Optimal Strategy? The Impact of Gaming Facilities on the Income and Employment of American Indians." *Economics Letters* 95(3): 428–32.

Reagan, Ronald. 1983. "Executive Order 12401—Presidential Commission on Indian Reservation Economies." *The American Presidency Project*. http://www.presidency.ucsb.edu/ws/?pid=41365.

Russ, Jacob W., and Thomas Stratmann. 2013. "Creeping Normalcy: Fractionation of Indian Land Ownership." GMU Working Paper in Economics no. 13-28, George Mason University.

Senate Committee on Indian Affairs. 1988. "Senate Report 100-446 on the Indian Gaming Regulatory Act, to Accompany S. 555." In *Congressional Record* 134: 3071–3106. WestLaw.

Taggart, William A., and Thaddieus W. Conner. 2011. "Indian Gaming and Tribal Revenue Allocation Plans: A Case of 'Play to Pay.'" *Gaming Law Review and Economics* 15(6): 355–63.

Taylor, Jonathan B. 2005. *Tribal Self-Government and Gaming Policy: The Consequences for Indians and Washington State*. (Indian Self-Government in Washington, Vol. 1.) Washington Indian Gaming Association and Taylor Policy Group. http://www.washingtonindiangaming.org/images/content/WA-Indian-Gaming-Policy-2-14-05.pdf.

Taylor, Jonathan B. 2006. *The Character and Effects of the Indian Economy in Washington State*. (Indian

Self-Government in Washington, Vol. 2.) Washington Indian Gaming Association and Taylor Policy Group. http://www.washingtonindiangaming.org/images/content/WIGA-Vol-2-Report-8-16-06-Lo Res.pdf.

Taylor, Jonathan B. 2012. "The Economic and Fiscal Impacts of Indian Tribes in Washington." Washington Indian Gaming Association and Taylor Policy Group. http://www.washingtonindiangaming.org/images/content/wigaeconseptupt3.pdf.

Taylor, Jonathan B., and Joseph Peggs Kalt. 2005. "American Indians on Reservations: A Databook of Socioeconomic Change Between the 1990 and 2000 Censuses." Harvard Project on American Indian Economic Development. https://nnidatabase.org/db/attachments/text/2005_TAYLOR_kalt_HPAIED_databook.pdf.

Taylor, Jonathan B., Matthew B. Krepps, and Patrick Wang. 2000. *The National Evidence on the Socioeconomic Impacts of American Indian Gaming on Non-Indian Communities.* Cambridge, MA: Harvard Project on American Indian Economic Development.

Trosper, Ronald L. 1996. "American Indian Poverty on Reservations, 1969–1989." In *Changing Numbers, Changing Needs: American Indian Demography and Public Health*, edited by Gary D. Sandefur, Ronald R. Rindfuss, and Barney Cohen, 172–95. Washington, DC: National Academy Press.

US Census. 2011a. *2010 Census Summary File 1 United States.* US Census Bureau.

US Census. 2011b. *Facts for Features: American Indian and Alaska Native Heritage Month: November 2011.* https://www.census.gov/newsroom/releases/archives/facts_for_features_special_editions/cb11-ff22.html.

US Census. 2011c. "Table 1258. Gaming Revenues by Industry: 2000 to 2009." In *Statistical Abstract of the United States: 2011*, p. 774. https://www.census.gov/compendia/statab/2011/tables/11s1258.pdf.

US Commission on Civil Rights. 2003. *A Quiet Crisis: Federal Funding and Unmet Needs in Indian Country.* Washington, DC: US Commission on Civil Rights.

Volberg, Rachel, Kari Nysse-Carris, and Dean Gerstein. 2006. *2006 California Problem Gambling Prevalence Survey: Final Report.* August. National Opinion Research Center, University of Chicago. http://www.standupca.org/reports/CA_Problem_Gambling_Prevalence_Survey-Final_Report.pdf.

Walke, Roger. 2000. "Indian-Related Federal Spending Trends, FY 1975–2001, US Congressional Research Service Memorandum, March 1, 2000." In *Report of the Committee on the Budget, United States Senate to Accompany S. Con. Res. 101 Together with Additional and Minority Views*, Senate Report 106-251, p. 199–250.

Walker, Douglas M. 2007. *The Economics of Casino Gambling.* 1st edition. Springer.

Washburn, Kevin K. 2008. "The Legacy of Bryan v. Itasca County: How an Erroneous $147 County Tax Notice Helped Bring Tribes $200 Billion in Indian Gaming Revenue." *Minnesota Law Review* 92(4): 919–970.

Washington Research Council. 2002. *Special Report: Untaxed and Lightly Regulated.* Seattle, WA: Washington Research Council.

Wilkins, David E. 2002. *American Indian Politics and the American Political System.* New York, NY: Rowman & Littlefield.

Wilkinson, Charles F. 1988. *American Indians, Time, and the Law: Native Societies in a Modern Constitutional Democracy.* Yale University Press.

Wolfe, Barbara, Jessica Jakubowski, Robert Haveman, and Marissa Courey. 2012. "The Income and Health Effects of Tribal Casino Gaming on American Indians." *Demography* 49(2): 499–524.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 209–216

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., Saint Paul, Minnesota, 55105.

Potpourri

An IMF staff team considers "Is the Glass Half Empty or Half Full? Issues in Managing Water Challenges and Policy Instruments." "Lack of proper management exacerbates water challenges, even in countries with abundant water endowment. A case in point is Pakistan, where, despite an abundance of water a few decades ago, lagging policies have raised the prospect of water scarcity that could threaten all aspects of the economy. The bulk of Pakistan's farmland is irrigated through a canal system, but canal water is vastly underpriced, recovering only one-quarter of annual operating and maintenance costs. Meanwhile, agriculture, which consumes almost all annual available surface water, is largely untaxed. The combination of these

■ *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives*, based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

http://dx.doi.org/10.1257/jep.29.3.209 doi=10.1257/jep.29.3.209

policies leads to overuse of water. In the Democratic Republic of the Congo (DRC), a country with an extensive system of rivers and lakes, years of poor management, conflicting water sector regulations, and low cost recovery have created a situation in which consumption of drinking water is far below the regional average and only a fraction of agricultural land is irrigated. . . . Experiences in some countries with naturally limited water resources have shown that sound water management can be achieved and water challenges are not insurmountable. . . . One notable innovation in Burkina Faso is the Bagre 'growth pole,' in which a huge manmade reservoir supports diverse activities, such as fishing and irrigation for crops. . . . For example, Burkina Faso introduced a progressive tariff grid for drinking water based on the volume of use, with the higher tiers subsidizing the lowest tier as well as part of sanitation activities." IMF Staff Discussion Note SDN/15/11. The listed authors are Kalpana Kochhar, Catherine Pattillo, Yan Sun, Nujin Suphaphiphat, Andrew Swiston, Robert Tchaidze, Benedict Clements, Stefania Fabrizio, Valentina Flamini, Laure Redifer, and Harald Finger. June 2015, http://www.imf.org/external/pubs/ft/sdn/2015/sdn1511.pdf.

Ravi Kanbur and Adam Wagstaf ask: "How Useful Is Inequality of Opportunity as a Policy Construct?" "In policy and political discourse, 'equality of opportunity' is the new motherhood and apple pie. It is often contrasted with equality of outcomes, with the latter coming off worse. Equality of outcomes is seen variously as Utopian, as infeasible, as detrimental to incentives, and even as inequitable if outcomes are the result of differing efforts. Equality of opportunity, on the other hand, is interchangeable with phrases such as 'leveling the playing field', 'giving everybody an equal start' and 'making the most of inherent talents.' In its strongest form, the position is that equality of outcomes should be irrelevant to policy; what matters is equality of opportunity. . . . However, attempts to quantify and apply the concept of equality of opportunity in a policy context have also revealed a host of problems of a conceptual and empirical nature, problems which may in the end even question the practical usefulness of the concept. . . . Health inequality is emblematic of the difficulties that current approaches face. If children's health is truly outside their control, then all of the inequality in their health is a legitimate objective of policy, not just that part which is explained by variables which measure parental circumstances. Similarly, especially for children but also for adults, if bad luck leads to ill health then wiping out this inequality as illegitimate for policy concern does not sit well with moral intuition—and yet that is what the present procedures which calculate inequality of opportunity in health tend to do." July 2014, World Bank Policy Research Working Paper 6980, http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2014/07/28/000158349_20140728112400/Rendered/PDF/WPS6980.pdf.

Bengt Holmstrom offers a framework for "Understanding the Role of Debt in the Financial System." "Panics always involve debt. Panics happen when information-insensitive debt (or banks) turns into information-sensitive debt . . . A regime shift occurs from a state where no one feels the need to ask detailed questions, to a state where there is enough uncertainty that some of the investors begin

to ask questions about the underlying collateral and others get concerned about the possibility. . . . These events are cataclysmic precisely because the liquidity of debt rested on over-collateralisation and trust rather than a precise evaluation of values. Investors are suddenly in the position of equity holders looking for information, but without a market for price discovery. Private information becomes relevant, shattering the shared understanding and beliefs on which liquidity rested . . . [T]here is a danger in the logic of money markets: if their liquidity relies on no or few questions being asked, how will one deal with the systemic risks that build up because of too little information and the weak incentives to be concerned about panics. I think the answer will have to rest on over-collateralisation, stress tests and other forms of monitoring banks and bank-like institutions. But my first priority has been to exposit the current logic and hope that it will be useful for the big question about systemic risk as we move forward." January 2015, BIS Working Paper 479, Bank of International Settlements, http://www.bis.org/publ/work479.pdf.

Liran Einav and Jonathan Levin discuss "Economics in the Age of Big Data." "Even 15 or 20 years ago, interesting and unstudied data sets were a scarce resource. Gathering data on a specific industry could involve hunting through the library or manually extracting statistics from trade publications. Collaborations with companies were unusual, as were experiments, both in laboratory settings and in the field. Nowadays the situation is very different along all of these dimensions. . . . The first feature is that data are now often available in real time. Government surveys and statistics are released with a lag of months or years. . . . However, administrative and private data that are continuously updated have great value for helping to guide economic policy. . . . The second feature is that data are available on previously unmeasured activities. Much of the data now being recorded is on activities that were previously difficult to quantify: personal communications, social networks, search and information gathering, and geolocation data. These data may open the door to studying issues that economists have long viewed as important but did not have good ways to study empirically, such as the role of social connections and geographic proximity in shaping preferences, the transmission of information, consumer purchasing behavior, productivity, and job search. Finally, data come with less structure. Economists are used to working with 'rectangular' data, with N observations and $K << N$ variables per observation and a relatively simple dependence structure between the observations. New data sets often have higher dimensionality and less-clear structure. For example, Internet browsing histories contain a great deal of information about a person's interests and beliefs and how they evolve over time. But how can one extract this information? The data record a sequence of events that can be organized in an enormous number of ways, which may or may not be clearly linked and from which an almost unlimited number of variables can be created. Figuring out how to organize and reduce the dimensionality of large-scale, unstructured data is becoming a crucial challenge in empirical economic research." *Science*, November 7, 2014, vol. 346, no. 6210, pp. 1243089-1 to 1243089-6.

The African Progress Panel is a group of 10 prominent individuals ranging from Kofi Annan to Bob Geldof. Leading staff members include Caroline Kende-Robb,

Kevin Watkins, and Maria Quattri. The group has published *People, Power, Planet: Seizing Africa's Energy and Power Opportunities*. "Measured on a global scale, electricity consumption in Sub-Saharan Africa excluding South Africa is pitifully low, averaging around 162 kilowatt hours (kWh) per capita a year. . . . The global average consumption figure is 2,800kWh, rising to 5,700kWh in the European Union and 12,200kWh in the United States. Electricity consumption for Spain exceeds that of the whole of Sub-Saharan Africa (excluding South Africa). To put the figures in a different context, 595 million Africans live in countries where electricity availability per person is sufficient to only light a single 100-watt light bulb continuously for less than two months. It takes the average Tanzanian around eight years to consume as much electricity as an American uses in one month. When American households switch on to watch the Super Bowl, the annual finale of the football season, they consume 10 times the electricity used over the course of a year by the more than 1 million people living in Juba, capital city of South Sudan. Ethiopia, with a population of 94 million, consumes one-third of the electricity supplied to the 600,000 residents of Washington D.C. . . . Around 30 countries in the region have grid-connected power systems smaller than 500 megawatts (MW), while another 13 have systems smaller than 100MW. For purposes of comparison, a single large-scale power plant in the United Kingdom generates 2,000MW." *Africa Progress Report 2015*. http://app-cdn.acwupload.co.uk/wp-content/uploads/2015/06/APP_REPORT_2015_FINAL_low1.pdf.

The Future of Children has devoted a nine-paper issue to the subject of "Policies to Promote Child Health." From the introductory essay by Janet Currie and Nancy Reichman: "Unfortunately, the fragmentation of children's health care services and resources in the United States, combined with a crisis response approach to child health, has produced an inefficient system. Moreover, because this fragmentation results in a lack of data about the cost effectiveness of various interventions and policies, it's hard to make informed policy choices. We suspect that, for many dimensions of child health, an ounce of prevention would be worth a pound of cure, but it's difficult to prove this without hard evidence on the costs and benefits of different approaches. . . . [S]pending on child health has increased over time, but that the largest share of the increased spending over the past century has been for health care, while spending on other determinants of child health, which may be as or more important, has not kept pace. . . . Many child health problems start early in life, in utero, or perhaps even before mothers conceive . . . " Spring 2015, http://www.princeton.edu/futureofchildren/.

Asli Demirguc-Kunt, Leora Klapper, Dorothe Singer, and Peter Van Oudheusden report results from "The Global Findex Database 2014: Measuring Financial Inclusion around the World." "The Global Financial Inclusion (Global Findex) database provides in-depth data showing how people save, borrow, make payments, and manage risk. . . . The indicators are based on interviews with about 150,000 nationally representative and randomly selected adults age 15 and above in more than 140 economies. . . . Between 2011 and 2014, 700 million adults became account holders while the number of those without an account—the unbanked—dropped

by 20 percent to 2 billion. What drove this increase in account ownership? A growth in account penetration of 13 percentage points in developing economies and innovations in technology—particularly mobile money, which is helping to rapidly expand access to financial services in Sub-Saharan Africa. Along with these gains, the data also show that big opportunities remain to increase financial inclusion, especially among women and poor people. Governments and the private sector can play a pivotal role by shifting the payment of wages and government transfers from cash into accounts. . . . In developing economies 1.3 billion adults with an account pay utility bills in cash, and more than half a billion pay school fees in cash. Digitizing payments like these would enable account holders to make the payments in a way that is easier, more affordable, and more secure." April 2015, World Bank Policy Research Working Paper 7255, http://www.worldbank.org/en/programs/globalfindex.

Interviews

Alvin Roth is interviewed by Douglas Clement. "God makes wheat, but the Chicago Board of Trade makes #2 hard red winter wheat. It has a lot less variance than wheat. You know what you're going to get and, therefore, you don't have to care who you're buying it from. You don't have to inspect it. But before wheat was commodified, you had to have someone look at the wheat to see what you were buying. . . . In those markets, you can make an offer to the entire market. I want #2 hard red winter wheat from whomever; it doesn't matter who I get it from. But, of course, labor markets aren't like that, and many other markets aren't like that—because you care not just about the price, but also about who you're dealing with. . . . Instead, it's personalized prices, maybe doubly personalized prices. How much will Google pay me to work for them? How much would I need to take their offer, rather than a different salary from Facebook? . . . There isn't a sharp line between matching markets and commodity markets. I think there is sort of a continuum. There are markets where price does all the work: the New York Stock Exchange, for instance. Its job is to define at any moment the price at which supply equals demand for each of a bunch of financial commodities. The labor market is very personal, but price also matters a lot, so it's somewhere in the middle of the continuum. For school choice and kidney exchange, we don't let prices work at all. And lots of markets fall somewhere between kidney exchange and the market for wheat." *The Region,* Federal Reserve Bank of Minneapolis, June 2015, pp. 14–25, https://www.minneapolisfed.org/publications/the-region/interview-with-alvin-roth.

Dani Rodrik is interviewed by Aaron Steelman. "The root of it is the problem that the profession has more or less the wrong idea about how economics as a science works. If you ask most economists, 'What kind of a science is economics?,' they will give a response that approximates natural sciences like physics, which is that we develop hypotheses and then we test them, we throw away those that are rejected, we keep those that cannot be rejected, and then we refine our hypotheses

and move in their direction. This is not how economics works—with newer and better models succeeding models that are older and worse in the sense of being empirically less relevant. The way we actually increase our understanding of the world is by expanding our collection of models. We don't throw out models, we add to them; the library of models expands. Social reality is very different from natural reality in that it is not fixed; it varies across time and place. The way that an economy works in the Congo is very different from the way that it works in the United States. So the best that we can do as economists is try to understand social reality one model at a time. Each model identifies one particular salient causal mechanism, and that salient effect might be very strong in the Congo but it may be very weak at any point in time in the United States, where we may need to apply a different model. . . . Economists know how to think about various causal mechanisms that operate as part of social reality, but what they're very bad at in practice is navigating among the models describing them. How exactly do I pick the right model for a given setting? This is a craft because the evidence never settles it in real time. We have these periods of fads where we say the New Keynesian or the Neoclassical model explains everything. We lose sight of the fact that models are highly context-specific and we need to be syncretic, simultaneously carrying many models in our mind." *Econ Focus,* Federal Reserve Bank of Richmond, Third Quarter 2014, https://www.richmondfed.org/publications/research/econ_focus/2014/q3/interview.

Claudia Goldin is interviewed by Jessie Romero. "Across the wage distribution, the vast majority of the gender gap is occurring within occupations, not between occupations. There's considerable discussion about occupational segregation, but you could get rid of all occupational segregation and reduce the gender gap by only a small amount. . . . So then the question is, why are there some occupations with large gender gaps and others with very narrow gaps? There are some occupations where people face a nonlinear function of wages with respect to hours worked; that is, people earn a disproportionate premium for working long and continuous hours. For example, someone with a law degree could work as a lawyer in a large firm, and that person would make a lot of money per unit of time. But if that person worked fewer than a certain number of hours per week, the pay rate would be cut quite a bit. Or someone could work fewer or more flexible hours as general counsel for a company and earn less per unit of time than the large-firm lawyer. Pharmacy is the opposite—earnings increase linearly with hours worked. There's no part-time penalty." *EconFocus,* Federal Reserve Bank of Richmond, Fourth Quarter 2014, pp. 24–28, https://www.richmondfed.org/publications/research/econ_focus/2014/q4/interview.

Discussion Starters

Elizabeth R. Berman and Rachel K. Johnson tell the story of "The Unintended Consequences of Changes in Beverage Options and the Removal of Bottled Water on a University Campus." "Policy changes related to the types of bottled beverages

sold at the University of Vermont in Burlington, Vermont, provided an opportunity to study how changes in beverage offerings affected the beverage choices as well as the calorie and total and added sugar consumption of consumers. First, in August 2012, all campus locations selling bottled beverages were required to provide a 30% healthy beverage ratio in accordance with the Alliance for a Healthier Generation's beverage guidelines. Then, in January 2013, campus sales locations were required to remove bottled water while still maintaining the required 30% healthy beverage ratio. . . . However, between fall 2012 and spring 2013, when bottled water was banned, the per capita number of bottles shipped to campus increased significantly. Thus, the bottled water ban did not reduce the number of bottles entering the waste stream from the university campus, which was the ultimate goal of the ban. Furthermore, with the removal of bottled water, people in the university community increased their consumption of other, less healthy bottled beverages." *American Journal of Public Health,* July 2015, vol. 105, no. 7, pp. 1404–08.

IEEE Spectrum has published a "Special Report: 50 Years of Moore's Law," with a dozen short articles looking back at Moore's original formulation of the law, how it has developed over time, and prospects for the law continuing. March–April 2015, at http://spectrum.ieee.org/static/special-report-50-years-of-moores-law. As one example, Chris Mack writes about "The Multiple Lives of Moore's Law: Why Gordon Moore's Grand Prediction Has Endured for 50 Years": "A half century ago, a young engineer named Gordon E. Moore took a look at his fledgling industry and predicted big things to come in the decade ahead. In a four-page article in the trade magazine *Electronics*, he foresaw a future with home computers, mobile phones, and automatic control systems for cars. All these wonders, he wrote, would be driven by a steady doubling, year after year, in the number of circuit components that could be economically packed on an integrated chip. A decade later, the exponential progress of the integrated circuit—later dubbed 'Moore's Law'—showed no signs of stopping. And today it describes a remarkable, 50-year-long winning streak that has given us countless forms of computers, personal electronics, and sensors. The impact of Moore's Law on modern life can't be overstated. We can't take a plane ride, make a call, or even turn on our dishwashers without encountering its effects." At http://spectrum.ieee.org/semiconductors/processors/the-multiple-lives-of-moores-law. As another example, in "Graphic: Transistor Production Has Reached Astronomical Scales," Dan Hutcheson writes: "In 2014, semiconductor production facilities made some 250 billion billion (250×10^{18}) transistors. This was, literally, production on an astronomical scale. Every second of that year, on average, 8 trillion transistors were produced. That figure is about 25 times the number of stars in the Milky Way and some 75 times the number of galaxies in the known universe. The rate of growth has also been extraordinary. More transistors were made in 2014 than in all the years prior to 2011." At http://spectrum.ieee.org/computing/hardware/transistor-production-has-reached-astronomical-scales.

Tomáš Hellebrandt and Paolo Mauro forecast "The Future of Worldwide Income Distribution." They look at a wide array of household-level evidence on the distribution of income in more than 100 countries, and use estimates of

economic growth to project what the distribution of global income will look like in the future. They write: "Global income inequality started declining significantly at the turn of the century, and we project that this trend will continue for the next two decades, under what we consider the profession's 'consensus' projections for the growth rates of output and population." April 2015, Peterson Institute for International Economics, Working Paper 15-7, http://www.piie.com/publications/wp/wp15-7.pdf.

James Bessen discusses the interaction between technology and employment in "Toil and Technology." Here's one example: "Just because computers can perform some job tasks does not mean that jobs will be eliminated. Consider bank tellers. Automated teller machines (ATMs) were first installed in the United States and other developed economies in the 1970s. These machines handle some of the most common tasks bank tellers performed, such as dispensing cash and taking deposits. Starting in the mid-1990s, banks rapidly increased their use of ATMs; over 400,000 are installed in the United States alone today. One might expect such automation to decimate the ranks of bank tellers, but in fact the number of bank teller jobs did not decrease as the ATMs were rolled out. Instead, two factors combined to preserve teller jobs. First, ATMs increased the demand for tellers because they reduced the cost of operating a bank branch. Thanks to the ATM, the number of tellers required to operate a branch office in the average urban market fell from 20 to 13 between 1988 and 2004. But banks responded by opening more branches to compete for greater market share. Bank branches in urban areas increased 43 percent. Fewer tellers were required for each branch, but more branches meant that teller jobs did not disappear. Second, while ATMs automated some tasks, the remaining tasks that were not automated became more valuable. As banks pushed to increase their market shares, tellers became an important part of the 'relationship banking team.' Many bank customers' needs cannot be handled by machines—particularly small business customers'. Tellers who form a personal relationship with these customers can help sell them on high-margin financial services and products. The skills of the teller changed: cash handling became less important and human interaction more important." *Finance & Development*, March 2015, vol. 52, no. 1, pp. 16–19, http://www.imf.org/external/pubs/ft/fandd/2015/03/bessen.htm.

Journal of Economic Perspectives—Volume 29, Number 3—Summer 2015—Pages 217–218

Editorial Note
Correction to Jeffrey B. Liebman's "Understanding the Increase in Disability Insurance Benefit Receipt in the United States"

In the paper "Understanding the Increase in Disability Insurance Benefit Receipt in the United States," by Jeffrey B. Liebman, in the Spring 2015 issue (pp. 123–150), the author discovered some coding errors soon after publication. These errors do not substantially change anything in the analysis, and changes to the graphs are almost imperceptible. A corrected version is now posted online with the supplementary materials at the *JEP* website at https://www.aeaweb.org/articles.php?doi=10.1257/jep.29.2.123.

AMERICAN
ECONOMIC
ASSOCIATION

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 *JEP* for your e-reader, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $20.00, $30.00, or $40.00, 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

Summer 2015, Volume 29, Number 3

Symposia

Automation and Labor Markets

Pre-Analysis Plans

Doing Business

Articles

AMERICAN ECONOMIC ASSOCIATION