The Journal of Economic Perspectives

A journal of the
American Economic Association

Fall 2013

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), Fall 2013, Vol. 27, No. 4. 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 http://www.vanderbilt.edu/AEA/. 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 27 • Number 4 • Fall 2013

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 27, Number 4—Fall 2013—Pages 3–16

A Century of US Central Banking: Goals, Frameworks, Accountability

Ben S. Bernanke

Several key episodes in the 100-year history of the Federal Reserve have been referred to in various contexts with the adjective "Great" attached to them: the Great Experiment of the Federal Reserve's founding, the Great Depression, the Great Inflation and subsequent disinflation, the Great Moderation, and the recent Great Recession. Here, I'll use this sequence of "Great" episodes to discuss the evolution over the past 100 years of three key aspects of Federal Reserve policymaking: the goals of policy, the policy framework, and accountability and communication. The changes over time in these three areas provide a useful perspective, I believe, on how the role and functioning of the Federal Reserve have changed since its founding in 1913, as well as some lessons for the present and for the future.

The Great Experiment

The original goal of the Great Experiment that was the founding of the Fed was the preservation of financial stability. In the words of one of the authors of the Federal Reserve Act, Robert Latham Owen (1919, p. 24), the Federal Reserve was

■ *Ben S. Bernanke is Chairman of the Board of Governors of the Federal Reserve System, Washington, DC.*

http://dx.doi.org/10.1257/jep.27.4.3 doi=10.1257/jep.27.4.3

established to "provide a means by which periodic panics which shake the American Republic and do it enormous injury shall be stopped."[1]

At the time, the standard view of financial panics was that they were triggered when the needs of business and agriculture for liquid funds outstripped the available supply—as when seasonal plantings or shipments of crops had to be financed, for example—and that panics were further exacerbated by the incentives of banks and private individuals to hoard liquidity during such times (Warburg 1914). The new institution was intended to relieve such strains by providing an "elastic" currency: that is, by providing liquidity as needed to individual member banks through the discount window. Commercial banks, in turn, would then be able to accommodate their customers. Interestingly, although congressional advocates hoped the creation of the Fed would help prevent future panics, they did not fully embrace the idea that the Fed should help end ongoing panics by serving as lender of last resort, as had been famously recommended by the British economist and writer Walter Bagehot (1873 [1897]), the source of the classic dictum that central banks should address panics by lending freely at a penalty rate (see also Willis 1923, p. 1407; Carlson and Wheelock 2012; Bordo and Wheelock 2013). Instead, legislators imposed limits on the Federal Reserve's ability to lend in response to panics, for example, by denying nonmember banks access to the discount window and by restricting the types of collateral that the Fed could accept.[2]

Soon after the Federal Reserve was founded in 1913, its mission shifted to supporting the war effort and then to managing the unwinding of that support. The year 1923 was thus one of the first in which the Federal Reserve confronted normal peacetime financial conditions, and it took the opportunity to articulate its views on the appropriate conduct of policy in such conditions in the *Tenth Annual Report of the Federal Reserve Board* (Board of Governors 1924).

The framework that the Federal Reserve employed in these early years to promote financial stability reflected in large measure the fact that the United States was on the gold standard as well as the influence of the so-called "real bills" doctrine.[3]

[1] A 1929 book review by the financial editor of the *New York Times*, making reference both to the idea of a "great experiment" and to the broad responsibilities for financial stability of the new central bank, observed: "The Federal Reserve System has from the first necessarily been a great experiment, bound to adjust its general policies to the requirements of such novel and varying situations as should arise in the course of our financial history and which could not possibly be foreseen" (Noyes 1929). To be sure, the US Treasury carried out some central banking functions before the creation of the Federal Reserve, and the First and Second Banks of the United States represented early attempts to establish a central bank. By 1913, however, it had been about 75 years since the latter institution had ceased fulfilling that purpose. Moreover, the Federal Reserve operated somewhat differently from the prior institutions, as well as from existing central banks abroad, and thus its creation amounted to an experiment.

[2] The collateral acceptable to be pledged to the discount window has been expanded significantly over time; in particular, various pieces of banking legislation in the early 1930s enabled the Federal Reserve to make advances to member banks so long as the loans were "secured to the satisfaction" of the Federal Reserve Bank extending the loan. The Monetary Control Act of 1980 gave all depository institutions access to the discount window.

[3] Humphrey (1982) discusses the historical evolution of the real bills doctrine. He notes that, in its simplest form, the doctrine contends that banks should lend against short-term commercial paper

In the real bills doctrine, the Federal Reserve saw its function as meeting the needs of business for liquidity—consistent with the idea of providing an elastic currency—with the ultimate goal of supporting financial and economic stability. When business activity was increasing, the Federal Reserve would seek to accommodate the need for credit by supplying liquidity to banks; when business was contracting and less credit was needed, the Fed would then reduce the liquidity in the system. The policy framework of the Fed's early years has been much criticized in retrospect. Economic historians have pointed out that, under the real bills doctrine, the Fed increased the money supply precisely at those times at which business activity and upward pressures on prices were strongest; that is, monetary policy was procyclical. Thus, the Fed's actions tended to increase rather than decrease the volatility in economic activity and prices (Friedman and Schwartz 1963; Humphrey 1982; Meltzer 2003).

As noted, the Federal Reserve pursued its real bills approach in the context of the gold standard. In the 1920s, Federal Reserve notes were redeemable in gold on demand, and the Fed was required to maintain a gold reserve equal to 40 percent of outstanding notes. In principle, the gold standard should limit discretion by monetary policymakers, but in practice US monetary policy did not appear to be greatly constrained in the years after the Fed's founding. Indeed, the large size of the US economy, together with the use of market interventions that prevented inflows and outflows of gold from being fully translated into changes in the domestic money supply, gave the Federal Reserve considerable scope during the 1920s to conduct monetary policy according to the real bills doctrine without much hindrance from the gold standard.[4]

I've discussed the original mandate and early policy framework of the Federal Reserve. What about its accountability to the public? When the Federal Reserve was established, the question of whether it should be a private or a public institution was highly contentious. The compromise solution created a hybrid Federal Reserve System. The system was headed by a federally appointed Board of Governors, which initially included the Secretary of the Treasury and the Comptroller of the Currency. However, the 12 regional Reserve Banks were placed under a mixture of public and private oversight, including board members drawn from the private sector, and they

---

associated with real business transactions (as opposed to other activities such as speculative investment). According to this doctrine, central banks should expand the money supply to facilitate this type of bank lending, by buying commercial paper from banks or accepting as collateral banks' holdings of such paper. Thus, the doctrine implies that the money supply should expand and contract along with business activity.

[4] Specifically, the Fed was able to sterilize the effects of gold flows on the domestic money supply through open market operations—the purchase and sale of government securities in the open market. Initially, the Fed's main tools were the quantity of its lending through the discount window and the interest rate at which it lent—the discount rate. Open market operations were "discovered" when, to generate earnings to finance its operations, the Federal Reserve began in the 1920s to purchase government securities. Fed officials soon found that these operations affected the supply and cost of bank reserves and, consequently, the terms on which banks extended credit to their customers. Subsequently, of course, open market operations became a principal monetary policy tool, one that allowed the Fed to interact with the broader financial markets, not only with banks (Strong 1926).

were given considerable scope to make policy decisions that applied to their own districts. For example, Reserve Banks were permitted during this time to set their own discount rates, subject to a minimum set by the Board of Governors.

While the founders of the Federal Reserve hoped that this new institution would provide financial and hence economic stability, the policy framework and the institutional structure would prove inadequate to the challenges the Fed would soon face.

The Great Depression

The Great Depression was the Federal Reserve's most difficult test. Tragically, the Fed failed to meet its mandate to maintain financial stability. In particular, although the Fed provided substantial liquidity to the financial system following the 1929 stock market crash, its response to the subsequent banking panics of the 1930s was limited at best; the widespread bank failures and the collapse in money and credit that ensued were major sources of the economic downturn. Bagehot's dictum to lend freely at a penalty rate in the face of panic appeared to have few adherents at the Federal Reserve of that era (Friedman and Schwartz 1963).

Economists have also identified a number of instances from the late 1920s to the early 1930s when Federal Reserve officials, in the face of the sharp economic contraction and financial upheaval, either tightened monetary policy or chose inaction. Some historians trace these policy mistakes to the early death in 1928 of Benjamin Strong, Governor of the Federal Reserve Bank of New York, which left the decentralized system without an effective leader (for example, Friedman and Schwartz 1963, chapter 7). This hypothesis, whether valid or not, raises the interesting question of what intellectual framework an effective leader would have drawn on at the time to develop and justify a more activist monetary policy. The degree to which the gold standard actually constrained US monetary policy during the early 1930s is debated; but, in any case, the gold standard philosophy clearly did not encourage the sort of highly expansionary policies that were needed.[5] The same can be said for the real bills doctrine, which apparently led policymakers to conclude, on the basis of low nominal interest rates and low borrowings from the Fed, that monetary policy was appropriately supportive and that further actions would be fruitless (Meltzer 2003; Romer and Romer 2013). Historians have also noted the prevalence at the time of yet another counterproductive doctrine: the so-called "liquidationist view" that depressions perform a necessary cleansing function (as discussed, for

[5] The US commitment to the gold standard might have constrained policy if looser monetary conditions, by encouraging capital outflows and a higher demand for imports, induced sufficient gold outflows to threaten the gold backing of the dollar (Eichengreen 1992). Wicker (1965) and Temin (1989) suggest that US policymakers in the early 1930s indeed felt constrained by the gold standard. In contrast, Hsieh and Romer (2006), as well as Bordo, Choudhri, and Schwartz (2002), focus on the short-lived monetary expansion in 1932 as evidence against the idea that the gold standard imposed important constraints on the Federal Reserve.

example, in DeLong 1990). It may be that the Federal Reserve suffered less from lack of leadership in the 1930s than from the lack of an intellectual framework for understanding what was happening and what needed to be done.

The Fed's inadequate policy framework ultimately collapsed under the weight of economic failures, new ideas, and political developments. The international gold standard was abandoned during the 1930s. The real bills doctrine lost prestige after the disaster of the 1930s; for example, the Banking Act of 1935 amended section 12A(c) of the Federal Reserve Act so as to instruct the Federal Reserve to use open market operations with consideration of "the general credit situation of the country," not just to focus narrowly on short-term liquidity needs. The Congress also expanded the Fed's ability to provide credit through the discount window, allowing loans to a broader array of counterparties, secured by a broader variety of collateral.[6]

The experience of the Great Depression had major ramifications for all three aspects of the Federal Reserve I am discussing: its goals, its policy framework, and its accountability to the public. With respect to goals, the high unemployment of the Depression—and the fear that high unemployment would return after World War II—elevated the maintenance of full employment as a goal of macroeconomic policy. The Employment Act of 1946 made the promotion of employment a general objective for the federal government. Although the Fed did not have a formal employment goal until the Federal Reserve Reform Act of 1977 codified "maximum employment," along with "stable prices," as part of the Fed's so-called dual mandate, earlier legislation nudged the central bank in that direction.[7] For example, legislators described the intent of the Banking Act of 1935 as follows: "To increase the ability of the banking system to promote stability of employment and business, insofar as this is possible within the scope of monetary action and credit administration" (US Congress 1935). At the same time, the Federal Reserve became less focused on its original mandate of preserving financial stability, perhaps in part because it felt superseded by the creation during the 1930s of the Federal Deposit Insurance Corporation and the Securities and Exchange Commission, along with other reforms intended to make the financial system more stable.

In the area of governance and accountability to the public, policymakers also recognized the need for reforms to improve the Federal Reserve's structure and decision-making. The Banking Act of 1935 simultaneously bolstered the legal independence of the Federal Reserve and provided for stronger central control by the Federal Reserve Board. In particular, the act created the modern configuration of

[6] For example, section 10B enhanced the powers of the Federal Reserve to lend to member banks, and sections 13(3) and 13(13) enabled the Federal Reserve to provide short-term credit to a wide range of potential borrowers in specific circumstances.

[7] More precisely, the three statutory objectives for monetary policy set forth in the Federal Reserve Reform Act of 1977 are maximum employment, stable prices, and moderate long-term interest rates. The dual mandate refers to the first two goals, and the long-term interest rate goal is viewed as likely to emerge from the macroeconomic environment associated with achievement of the employment and price stability goals (Mishkin 2007). Thus, the interest rate goal of the Federal Reserve Reform Act can be regarded as subsumed within the dual mandate.

the Federal Open Market Committee (FOMC), giving the Board the majority of votes on the Committee, while removing the Secretary of the Treasury and the Comptroller of the Currency from the Board. In practice, however, the US Treasury continued to have considerable sway over monetary policy after 1933, with Meltzer (2003) describing the Fed as "in the back seat." During World War II, the Federal Reserve used its tools to support the war financing efforts by holding interest rates and government borrowing costs low. Even after the war, Federal Reserve policy remained subject to considerable Treasury influence. It was not until the 1951 Accord with the Treasury that the Federal Reserve began to recover genuine independence in setting monetary policy.

The Great Inflation and Disinflation

Once the Federal Reserve regained its policy independence, its goals centered on the price stability and employment objectives laid out in the Employment Act of 1946. In the early post–World War II decades, the Fed used open market operations and the discount rate to influence short-term market interest rates; the federal funds interest rate (that is, the interest rate that depository institutions pay each other for loans, usually overnight, to make sure that they hold sufficient reserves at the Fed) gradually emerged as the preferred target for conducting monetary policy. Low and stable inflation was achieved for most of the 1950s and the early 1960s. However, beginning in the mid-1960s, inflation began a long climb upward, partly because policymakers proved to be too optimistic about the economy's ability to sustain rapid growth without inflation (for discussion, see Orphanides 2003; Meltzer 2009a).

Two mechanisms might have mitigated the damage from that mistaken optimism. First, a stronger policy response to rising inflation—more like that observed in the 1950s—certainly would have helped (Romer and Romer 2002b). Indeed, empirical estimates of the response of the federal funds rate to inflation for the 1970s generally show only a weak reaction (Judd and Rudebusch 1998; Taylor 1999a; Clarida, Galí, and Gertler 2000). Second, Fed policymakers could have reacted to continued high readings on inflation by adopting a more realistic and less optimistic assessment of the economy's productive potential (Lars Svensson in the discussion following Stokey 2003, p. 63). Instead, policymakers chose to emphasize so-called cost-push and structural factors as sources of inflation and saw wage- and price-setting as having become insensitive to economic slack (for example, Poole 1979; Romer and Romer 2002a, 2013; Bernanke 2004; Nelson 2005). This perspective, which contrasted sharply with Milton Friedman's (1963, p. 17) famous dictum that "inflation is always and everywhere a monetary phenomenon," led to Fed support for measures such as wage and price controls rather than monetary solutions to address inflation. A further obstacle was the view among many economists during the 1970s, as discussed in DeLong (1997) and Taylor (1997), that the gains from low inflation did not justify the costs of achieving it.

The consequence of the monetary framework of the 1970s was two bouts of double-digit inflation during that decade. Moreover, by the end of the decade, lack of commitment to controlling inflation had clearly resulted in inflation expectations becoming "unanchored," or unstable, with high estimates of trend inflation embedded in longer-term interest rates.

Under the leadership of Chairman Paul Volcker, the Federal Reserve in 1979 fundamentally changed its approach to the issue of ensuring price stability. This change involved an important rethinking on the part of policymakers. By the end of the 1970s, Federal Reserve officials increasingly accepted the view that inflation is a monetary phenomenon, at least in the medium and longer term; they became more alert to the risks of excessive optimism about the economy's potential output; and they placed renewed emphasis on the distinction between real—that is, inflation-adjusted—and nominal interest rates (for discussion, see Meltzer 2009b). The change in policy framework was initially tied to a change in operating procedures that put greater focus on growth in bank reserves, but the critical change—the willingness to respond more vigorously to inflation—endured even after the Federal Reserve resumed its traditional use of the federal funds rate as the policy instrument (Axilrod 1982). The new regime also reflected an improved understanding of the importance of providing a firm anchor for the inflation expectations of the private sector, secured by the credibility of the central bank.[8] Finally, it entailed a changed view about the dual mandate, in which policymakers regarded achievement of price stability as helping to provide the conditions necessary for sustained maximum employment (Lindsey, Orphanides, and Rasche 2005).

The Great Moderation

Volcker's successful battle against inflation set the stage for the so-called Great Moderation of 1984 to 2007, during which the Fed enjoyed considerable success in achieving both objectives of its dual mandate. Financial stability remained a goal, of course. The Federal Reserve monitored threats to financial stability and responded when the financial system was upset by events such as the 1987 stock market crash and the terrorist attacks of 2001. More routinely, the Fed shared supervisory duties with other banking agencies. Nevertheless, for the most part, financial stability did not figure prominently in monetary policy discussions during these years. In retrospect, it is clear that, during that period, macroeconomists—both inside and outside central banks—relied too heavily in their modeling and

[8] The emphasis of central banks on management of inflation expectations partly reflected lessons from the rational expectations literature of the 1970s. Monetary policy implications of the rational expectations literature were further clarified by later research. For example, Sargent (1982) brought out dramatically the dependence of inflation expectations on the monetary policy regime in his study of major disinflations, while rational expectations models were extended to include sticky prices (Fischer 1977; Taylor 1980; Rotemberg 1982; Calvo 1983) and interest rate rules (Sargent and Wallace 1975; McCallum 1981; Taylor 1993, 1999b; Woodford 2003).

analysis on variants of the so-called Modigliani and Miller (1958) theorem, which shows that—under a number of restrictive assumptions—the value of a firm is not related to how that firm is financed.[9] Influenced by the logic of Modigliani–Miller, many monetary economists and central bankers concluded that the details of the structure of the financial system could be largely ignored when analyzing the behavior of the broader economy.

An important development of the Great Moderation was the increasing emphasis that central banks around the world put on communication and transparency, as economists and policymakers reached consensus on the value of communication in attaining monetary policy objectives (Woodford 2005). Federal Reserve officials, like those at other central banks, had traditionally been highly guarded in their public pronouncements. They believed, for example, that the ability to take markets by surprise was important for influencing financial conditions (for example, Goodfriend 1986; Cukierman and Meltzer 1986). Although Fed policymakers of the 1980s and early 1990s had become somewhat more explicit about policy objectives and strategy (Orphanides 2006), the same degree of transparency was not forthcoming on monetary policy decisions and operations. The release of a post-meeting statement by the Federal Open Market Committee, a practice that began in 1994, was therefore an important watershed. Over time, these statements were expanded to include more detailed information about the reason for the policy decision and an indication of the balance of risks (Lindsey 2003).

In addition to improving the effectiveness of monetary policy, these developments in communications also enhanced the public accountability of the Federal Reserve. Accountability is, of course, essential for continued policy independence in a democracy. Moreover, central banks that are afforded policy independence in the pursuit of their mandated objectives tend to deliver better economic outcomes (Alesina and Summers 1993; Debelle and Fischer 1994).

One cannot look back at the Great Moderation today without asking whether the sustained economic stability of the period somehow promoted the excessive risk-taking that followed. The idea that this long period of relative calm lulled investors, financial firms, and financial regulators into paying insufficient attention to risks that were accumulating must have some truth in it. I don't think we should conclude, though, that we therefore should not strive to achieve economic stability. Rather, the right conclusion is that, even in (or perhaps, especially in) stable and prosperous times, monetary policymakers and financial regulators should regard safeguarding financial stability to be of equal importance as—indeed, a necessary prerequisite for—maintaining macroeconomic stability.

[9] Specifically, Modigliani and Miller (1958) argue that, under certain conditions, firms will be indifferent between obtaining funds via equity finance and obtaining funds via debt issue. As noted in the text, some researchers have taken their result as implying that detailed modeling of the financial sector may not be central for understanding private sector decisions or the effects of monetary policy. However, as also noted in the text, Modigliani's and Miller's result depends on restrictive assumptions, including no effects of taxes on financing choices, no bankruptcy costs, no agency problems, and no asymmetric information.

Macroeconomists and historians will continue to debate the sources of the remarkable economic performance during the Great Moderation: for a sampling of the debate, one might start with Stock and Watson (2003); Ahmed, Levin, and Wilson (2004); Dynan, Elmendorf, and Sichel (2006); and Davis and Kahn (2008). My own view is that the improvements in the monetary policy framework and in monetary policy communication, including, of course, the better management of inflation and the anchoring of inflation expectations, were important reasons for that strong performance. However, we have learned in recent years that while well-managed monetary policy may be necessary for economic stability, it is not sufficient.

The Financial Crisis, the Great Recession, and Today

It has now been about six years since the first signs of the financial crisis appeared in the United States in 2007, and the economy still has not fully recovered from its effects. What lessons should we take for the future from this experience, particularly in the context of a century of Federal Reserve history?

The financial crisis and the ensuing Great Recession reminded us of a lesson that we learned both in the nineteenth century and during the Depression, but had forgotten to some extent, which is that severe financial instability can do grave damage to the broader economy. The implication is that a central bank must take into account risks to financial stability if it is to help achieve good macroeconomic performance. Today, the Federal Reserve sees its responsibilities for the maintenance of financial stability as coequal with its responsibilities for the management of monetary policy, and we have made substantial institutional changes in recognition of this change in goals. In a sense, we have come full circle, back to the original goal of the Federal Reserve of preventing financial panics (Bernanke 2011).

How should a central bank seek to enhance financial stability? One means is by assuming the lender-of-last-resort function that Bagehot (1873 [1897]) understood and described 140 years ago, under which the central bank uses its power to provide liquidity to ease market conditions during periods of panic or incipient panic. The Fed's many liquidity programs played a central role in containing the crisis of 2008 to 2009. However, putting out the fire is not enough; it is also important to foster a financial system that is sufficiently resilient to withstand large financial shocks. Toward that end, the Federal Reserve, together with other regulatory agencies and the Financial Stability Oversight Council, is actively engaged in monitoring financial developments and working to strengthen financial institutions and markets. The reliance on stronger regulation is informed by the success of New Deal regulatory reforms, but current reform efforts go even further by working to identify and defuse risks not only to individual firms but to the financial system as a whole, an approach known as "macroprudential regulation."

Financial stability is also linked to monetary policy, though these links are not yet fully understood. Here the Fed's evolving strategy is to make monitoring, supervision, and regulation the first line of defense against systemic risks; to the

extent that risks remain, however, the Federal Open Market Committee strives to incorporate these risks in the cost–benefit analysis applied to all monetary policy actions (Bernanke 2002).

What about the monetary policy framework? In general, the Federal Reserve's policy framework inherits many of the elements put in place during the Great Moderation. These features include the emphasis on preserving the Fed's inflation credibility, which is critical for anchoring inflation expectations, and a balanced approach in pursuing both parts of the Fed's dual mandate in the medium term. We have also continued to increase the transparency of monetary policy. For example, the Federal Open Market Committee's communications framework now includes a statement of its longer-run goals and monetary policy strategy. In the statement issued January 25, 2012, (http://www.federalreserve.gov/newsevents/press/monetary/20120125c.htm), the Committee indicated that it judged that inflation at a rate of 2 percent (as measured by the annual change in the price index for personal consumption expenditures) is most consistent over the longer run with the FOMC's dual mandate. FOMC participants also regularly provide estimates of the longer-run normal rate of unemployment; those estimates currently have a central tendency of 5.2 to 6.0 percent. By helping to anchor longer-term expectations, this transparency gives the Federal Reserve greater flexibility to respond to short-run developments. This framework, which combines short-run policy flexibility with the discipline provided by the announced targets, has been described as constrained discretion (for example, as discussed in Bernanke and Mishkin 1997, in this journal). Other communication innovations include early publication of the minutes of FOMC meetings and quarterly post-meeting press conferences by the Chairman.

The framework for implementing monetary policy has evolved further in recent years, reflecting both advances in economic thinking and a changing policy environment. Notably, following the ideas of Svensson (2003) and others, the Federal Open Market Committee has moved toward a framework that ties policy settings more directly to the economic outlook, a so-called forecast-based approach. In a forecast-based approach, monetary policymakers inform the public of their medium-term targets—say, a specific value for the inflation rate—and attempt to vary the instruments of policy as needed to meet that target over time. In contrast, an instrument-based approach involves providing the public information about how the monetary policy committee plans to vary its policy instrument—typically, a short-term interest rate, like the federal funds interest rate—in response to economic conditions. In particular, the FOMC has released more detailed statements following its meetings that have related the outlook for policy to prospective economic developments and has introduced regular summaries of the individual economic projections of FOMC participants (including for the target value of the federal funds interest rate). The provision of additional information about policy plans has helped Fed policymakers deal with the constraint posed by the effective lower bound on short-term interest rates; in particular, by offering guidance about how policy will respond to economic developments, the Committee has been able to increase policy accommodation, even

when the short-term interest rate is near zero and cannot be meaningfully reduced further (as elaborated in Yellen 2012). The Committee has also sought to influence interest rates of securities that mature farther into the future (that is, farther out on the "yield curve"), notably through its securities purchases. Other central banks in advanced economies that also confronted the situation that short-term interest rates had been lowered to their effective lower bound of near-zero percent have taken similar measures.

In short, the recent crisis has underscored the need both to strengthen monetary policy and financial stability frameworks and to better integrate the two. We have made progress on both counts, but more needs to be done. In particular, the complementarities among regulatory and supervisory policies (including macroprudential policy), lender-of-last-resort policy, and standard monetary policy are increasingly evident. Both research and experience are needed to help the Fed and other central banks develop comprehensive frameworks that incorporate all of these elements. The broader conclusion is what might be described as the overriding lesson of the Federal Reserve's history: that central banking doctrine and practice are never static. We and other central banks around the world will have to continue to work hard to adapt to events, new ideas, and changes in the economic and financial environment.

■ *This paper is a revised version of remarks presented at "The First 100 Years of the Federal Reserve: The Policy Record, Lessons Learned, and Prospects for the Future," a conference sponsored by the National Bureau of Economic Research in Cambridge, Massachusetts, on July 10, 2013. I am indebted to Mark Carlson, Edward Nelson, and Jonathan Rose of the Board's staff for their substantial contributions to the preparation of this article.*

References

Ahmed, Shaghil, Andrew Levin, and Beth Anne Wilson. 2004. "Recent U.S. Macroeconomic Stability: Good Policies, Good Practices, or Good Luck?" *Review of Economics and Statistics* 86(3): 824–32.

Alesina, Alberto, and Lawrence H. Summers. 1993. "Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence." *Journal of Money, Credit and Banking* 25(2): 151–62.

Axilrod, Stephen H. 1982. "Monetary Policy, Money Supply, and the Federal Reserve's Operating Procedures." In *Central Bank Views on Monetary Targeting*, edited by Paul Meek, 32–41. New York: Federal Reserve Bank of New York.

Bagehot, Walter. 1873 [1897]. *Lombard Street: A Description of the Money Market.* New York: Charles Scribner's Sons.

Bernanke, Ben S. 2002. "Asset-Price 'Bubbles' and Monetary Policy." Speech delivered at the New York Chapter of the National Association for Business Economics, New York, NY, October 15. http://www.federalreserve.gov/boarddocs/speeches/2002/20021015.

Bernanke, Ben S. 2004. "The Great Moderation." Speech delivered at the meetings of the Eastern Economic Association, Washington, DC, February 20. http://www.federalreserve.gov/boarddocs/speeches/2004/20040220/.

Bernanke, Ben S. 2011. "The Effects of the Great Recession on Central Bank Doctrine and Practice." Speech delivered at the Federal Reserve Bank of Boston 56th Economic Conference, Boston, MA, October 18. http://www.federalreserve.gov/newsevents/speech/bernanke20111018a.htm.

Bernanke, Ben S., and Frederic S. Mishkin. 1997. "Inflation Targeting: A New Framework for Monetary Policy?" *Journal of Economic Perspectives* 11(2): 97–116.

Board of Governors of the Federal Reserve System. 1924. *Tenth Annual Report of the Federal Reserve Board, Covering Operations for the Year 1923.* Washington, DC: Government Printing Office.

Bordo, Michael D., Ehsan U. Choudhri, and Anna J. Schwartz. 2002. "Was Expansionary Monetary Policy Feasible during the Great Contraction? An Examination of the Gold Standard Constraint." *Explorations in Economic History* 39(1): 1–28.

Bordo, Michael D., and David C. Wheelock. 2013. "The Promise and Performance of the Federal Reserve as Lender of Last Resort 1914–1933." In *The Origins, History, and Future of the Federal Reserve: A Return to Jekylll Island,* edited by Michael D. Bordo and William Roberds, 59–98. New York: Cambridge University Press.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12(3): 383–98.

Carlson, Mark A., and David C. Wheelock. 2012. "The Lender of Last Resort: Lessons from the Fed's First 100 Years." Working Paper 2012-056B, Research Division, Federal Reserve Bank of St. Louis, November. http://research.stlouisfed.org/wp/more/2012-056.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115(1): 147–80.

Cukierman, Alex, and Allan H. Meltzer. 1986. "A Theory of Ambiguity, Credibility, and Inflation under Discretion and Asymmetric Information." *Econometrica* 54(5): 1099–1128.

Davis, Steven J., and James A. Kahn. 2008. "Interpreting the Great Moderation: Changes in the Volatility of Economic Activity at the Macro and Micro Levels." *Journal of Economic Perspectives* 22(4): 155–80.

Debelle, Guy, and Stanley Fischer. 1994. "How Independent Should a Central Bank Be?" In *Goals, Guidelines, and Constraints Facing Monetary Policymakers,* edited by Jeffrey C. Fuhrer, 195–221, proceedings of a conference held in North Falmouth, MA, June. Boston: Federal Reserve Bank of Boston. http://www.bos.frb.org/economic/conf/conf38/conf38f.pdf.

DeLong, J. Bradford. 1990. "'Liquidation' Cycles: Old-Fashioned Real Business Cycle Theory and the Great Depression." NBER Working Paper 3546, December.

DeLong, J. Bradford. 1997. "America's Peacetime Inflation: The 1970s." In *Reducing Inflation: Motivation and Strategy,* edited by Christina D. Romer and David H. Romer, 247–76. Chicago: University of Chicago Press.

Dynan, Karen E., Douglas W. Elmendorf, and Daniel E. Sichel. 2006. "Can Financial Innovation Help to Explain the Reduced Volatility of Economic Activity?" *Journal of Monetary Economics* 53(1): 123–50.

Eichengreen, Barry. 1992. *Golden Fetters: The Gold Standard and the Great Depression, 1919–1939.* New York: Oxford University Press.

Fischer, Stanley. 1977. "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule." *Journal of Political Economy* 85(1): 191–205.

Friedman, Milton. 1963. *Inflation: Causes and Consequences.* New York: Asia Publishing House.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960.* Princeton, NJ: Princeton University Press.

Goodfriend, Marvin. 1986. "Monetary Mystique: Secrecy and Central Banking." *Journal of Monetary Economics* 17(1): 63–92.

Hsieh, Chang-Tai, and Christina D. Romer. 2006. "Was the Federal Reserve Constrained by the Gold Standard during the Great Depression? Evidence from the 1932 Open Market Purchase Program." *Journal of Economic History* 66(1): 140–76.

Humphrey, Thomas M. 1982. "The Real Bills Doctrine." Federal Reserve Bank of Richmond *Economic Review* 19 (September–October): 3–13. http://www.richmondfed.org/publications/research/economic_review/1982/pdf/er680501.pdf.

Judd, John P., and Glenn D. Rudebusch. 1998. "Taylor's Rule and the Fed: 1970–1997." Federal Reserve Bank of San Francisco *Economic Review,* no. 3, pp. 3–16. http://www.frbsf.org/economic-research/publications/98-3/3-16.pdf.

Lindsey, David E. 2003. *A Modern History of FOMC Communication: 1975–2002.* Unpublished report, Board of Governors of the Federal Reserve System, Division of Monetary Affairs, June. http://fraser.stlouisfed.org/docs/publications/books/20030624_lindsey_modhistfomc.pdf.

Lindsey, David E., Athanasios Orphanides, and Robert H. Rasche. 2005. "The Reform of October

1979: How It Happened and Why." Federal Reserve Bank of St. Louis *Review* 87(2): 187–236, http://research.stlouisfed.org/publications/review/05/03/part2/MarchApril2005Part2.pdf.

McCallum, Bennett T. 1981. "Price Level Determinacy with an Interest Rate Policy Rule and Rational Expectations." *Journal of Monetary Economics* 8(3): 319–29.

Meltzer, Allan H. 2003. *A History of the Federal Reserve*, Vol. 1: *1913–1951*. University of Chicago Press.

Meltzer, Allan H. 2009a. *A History of the Federal Reserve,* Vol. 2, Book 1: *1951–1969*. University of Chicago Press.

Meltzer, Allan H. 2009b. *A History of the Federal Reserve,* Vol. 2, Book 2: *1970–1986*. University of Chicago Press.

Mishkin, Frederic S. 2007. "Monetary Policy and the Dual Mandate." Speech delivered at Bridgewater College, Bridgewater, VA, April 10. http://www.federalreserve.gov/newsevents/speech/mishkin20070410a.htm.

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance, and the Theory of Investment." *American Economic Review* 48(3): 261–97.

Nelson, Edward. 2005. "The Great Inflation of the Seventies: What Really Happened?" *Advances in Macroeconomics* 5(1): Article 3.

Noyes, Alexander D. 1929. "Wall Street's Controversy with the Reserve Board: Professor Lawrence Takes the Side of the Stock Market in the Quarrel of the Past Year." A review of *Wall Street and Washington,* by Joseph Stagg Lawrence, in *New York Times*, Book Review, August 25.

Orphanides, Athanasios. 2003. "The Quest for Prosperity without Inflation." *Journal of Monetary Economics* 50(3): 633–63.

Orphanides, Athanasios. 2006. "The Road to Price Stability." *American Economic Review* 96(2): 178–81.

Owen, Robert L. 1919. *The Federal Reserve Act: Its Origin and Principles*. New York: Century Company.

Poole, William. 1979 "Burnsian Monetary Policy: Eight Years of Progress?" *Journal of Finance* 32(2): 473–84.

Romer, Christina D., and David H. Romer. 2002a. "The Evolution of Economic Understanding and Postwar Stabilization Policy." In *Rethinking Stabilization Policy,* proceedings of a symposium sponsored by the Federal Reserve Bank of Kansas City, August 29–31. http://www.kc.frb.org/publicat/sympos/2002/pdf/S02RomerandRomer.pdf

Romer, Christina D., and David H. Romer. 2002b. "A Rehabilitation of Monetary Policy in the 1950s." *American Economic Review* 92(2): 121–27.

Romer, Christina D., and David H. Romer. 2013. "The Most Dangerous Idea in Federal Reserve History: Monetary Policy Doesn't Matter." *American Economic Review* 103(3): 55–60.

Rotemberg, Julio J. 1982. "Sticky Prices in the United States." *Journal of Political Economy* 90(6): 1187–1211.

Sargent, Thomas J. 1982. "The Ends of Four Big Inflations." In *Inflation: Causes and Consequences,* edited by Robert E. Hall, 41–97. Chicago: University of Chicago Press.

Sargent, Thomas J., and Neil Wallace. 1975. "'Rational' Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule." *Journal of Political Economy* 83(2): 241–54.

Stock, James H., and Mark W. Watson. 2003. "Has the Business Cycle Changed? Evidence and Explanations." In *Monetary Policy and Uncertainty: Adapting to a Changing Economy,* proceedings of a symposium sponsored by the Federal Reserve Bank of Kansas City, Kansas City, MO, August 28–30, pp. 9–56. http://www.kc.frb.org/publicat/sympos/2003/pdf/Stockwatson2003.pdf.

Stokey, Nancy L. 2003. "'Rules vs. Discretion' after Twenty-Five Years." In *NBER Macroeconomics Annual 2002*, vol. 17, edited by Mark Gertler and Kenneth Rogoff; specifically the "Discussion" of the paper on pp. 62–64.

Strong, Benjamin. 1926. "Open Market Operations," in hearing before the US House of Representatives Committee on Banking and Currency, April. Reprinted in *Interpretations of Federal Reserve Policy in the Speeches and Writings of Benjamin Strong* (1930), edited by W. Randolph Burgess. New York: Harper and Brothers.

Svensson, Lars E. O. 2003. "What Is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules." *Journal of Economic Literature* 41(2): 426–77.

Taylor, John B. 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy* 88(1): 1–23.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39(1): 195–214.

Taylor, John B. 1997. Comment on "America's Peacetime Inflation: The 1970s," in *Reducing Inflation: Motivation and Strategy,* edited by Christina D. Romer and David H. Romer, 276–80. University of Chicago Press.

Taylor, John B. 1999a. "A Historical Analysis of Monetary Policy Rules." In *Monetary Policy Rules,* edited by John B. Taylor, 319–41. University of Chicago Press.

Taylor, John B., ed. 1999b. *Monetary Policy Rules*. University of Chicago Press.

Temin, Peter. 1989. *Lessons from the Great Depression*. Cambridge, MA: MIT Press.

US Congress, House Committee on Banking and Currency. 1935. "Banking Act of 1935, Report No. 742 to Accompany H.R. 7617." 74th Congress. Washington, DC: Government Printing Office.

Warburg, Paul M. 1914. "A United Reserve Bank of the United States." *Proceedings of the Academy of Political Science in the City of New York*, 4(July): 75–115.

Wicker, Elmus R. 1965. "Federal Reserve Monetary Policy, 1922–33: A Reinterpretation." *Journal of Political Economy* 73(4): 325–43.

Willis, H. Parker. 1923. *The Federal Reserve System: Legislation, Organization, and Operation*. New York: Ronald Press Company.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton: Princeton University Press.

Woodford, Michael. 2005. "Central Bank Communication and Policy Effectiveness." In *The Greenspan Era: Lessons for the Future*, proccedings of a symposium sponsored by the Federal Reserve Bank of Kansas City, pp. 399–474. Kansas City, MO: Federal Reserve Bank of Kansas City. http://www.kc.frb.org/publicat/sympos/2005/pdf/Woodford2005.pdf.

Yellen, Janet. 2012. "Revolution and Evolution in Central Bank Communications." Speech delivered at the Haas School of Business, University of California, Berkeley, CA, November 13. http://www.federalreserve.gov/newsevents/speech/yellen20121113a.htm.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 17–44

Central Bank Design[†]

Ricardo Reis

Starting with a blank slate, how could one design the institutions of a central bank for the United States? This question is not as outlandish as it may seem. As soon as the Iraq war ended in 2003, "the first major issue that Coalition economists confronted: What should be done with the Iraqi dinar?" (Foote, Block, Crane, and Gray 2004, p. 60). The economists involved stated that adopting a new central bank law in March 2004 was one of their first and most important economic accomplishments, and a similar judgment would hold true when independent central banks were created in most transition countries as well. Even looking at high-income economies, in 1992, Europeans had to answer to this question after they signed the Maastricht Treaty (von Hagen 1997). The US Federal Reserve has not been an institution set in stone; slowly, and with turns in different directions, its structure has been molded over 100 years into what it is today.

My goal here is not to describe these historical developments; for those who would like a detailed history, Friedman and Schwartz (1963) is the classic account of the history of the Federal Reserve, and Meltzer (2003, 2009a, 2009b) offers a more recent alternative. Instead, this paper explores the question of how to design a central bank, drawing on the relevant economic literature and historical experiences while staying free from concerns about how the Fed got to be what it is today or the short-term political constraints it has faced at various times. The goal is to provide an opinionated overview that puts forward the trade-offs associated with different choices and identifies areas where there are clear messages about

■ *Ricardo Reis is Professor of Economics, Columbia University, New York City, New York. His email address is rreis@columbia.edu.*

[†] To access the disclosure statement, visit
http://dx.doi.org/10.1257/jep.27.4.17

doi=10.1257/jep.27.4.17

optimal central bank design. Romer and Romer (1997) and Blinder (2006) are important precursors.

Stripped to its core, a central bank is the sole institution in a country with the power to borrow from banks in the form of reserves while committing to exchange these reserves on par with banknotes that the central bank can freely issue. More broadly, the central bank can choose some policy instruments that it controls directly, as well as a set of announcements about its knowledge of the economy or future policy intentions. Designing the central bank then consists of specifying three elements: First is the *objective function*, which comes from somewhere or someone, and includes only a few macroeconomic variables, which serve as goals for the central bank, potentially at different horizons, matching the small set of instruments at its disposal. Second, the central bank faces a *resource constraint*, limiting both its ability to distribute dividends as well as the set of policies that it can pursue. Third, there is a set of *equilibrium constraints* mapping policy actions and announcements onto the simultaneous evolution of private agents' beliefs and macroeconomic outcomes, so that commitments by the central bank and transparency about its future intentions can have an effect right away. In the course of exploring these three broad categories, I will discuss twelve dimensions in central bank design.

The Central Bank's Goals

Choosing goals includes reflecting on who makes those choices, which macroeconomic variables are included and at what time horizon, and how to consider differing views.

Dimension 1: The Strictness of the Central Bank's Mandate

A central bank is an agent of the government that should serve society. Basic democratic principles suggest that society should give it a clear set of goals.

However, the mandate of central banks has traditionally been vague. In the United States, the Federal Reserve Reform Act of 1977 established certain goals for the central bank: "maximum employment, stable prices and moderate long-term interest rates." Before spending too much time at their job, most Fed governors give at least one official speech in which they state their interpretation of this mandate: after all, maximum employment does not mean that every able man or woman must have a job, stable prices do not mean average measured inflation is exactly zero, and the third goal—moderate long-term interest rates—is often a consequence of the first two. The mere fact that the governors feel compelled to make their goals clear shows that they have a great deal of discretion in setting the yardsticks by which their own performance is measured.

An active literature has studied the benefits of giving the central bank more precise mandates. Some of the suggestions are to set an objective function that puts a higher weight on inflation relative to other components of social welfare or that explicitly links the central banker's salary or chances of dismissal to

numerical measures of performance (Rogoff 1985; Walsh 1995; Svensson 1997). A well-established consensus argues for central banks to adopt a numerical nominal anchor, even if an active debate remains on how to pick that anchor and on the strictness and speed at which to reach the target (Bernanke and Mishkin 1997; Woodford 2012). These proposals require a mandate that makes society's goals clear and that offers some direction on how to weight each goal relative to the others.

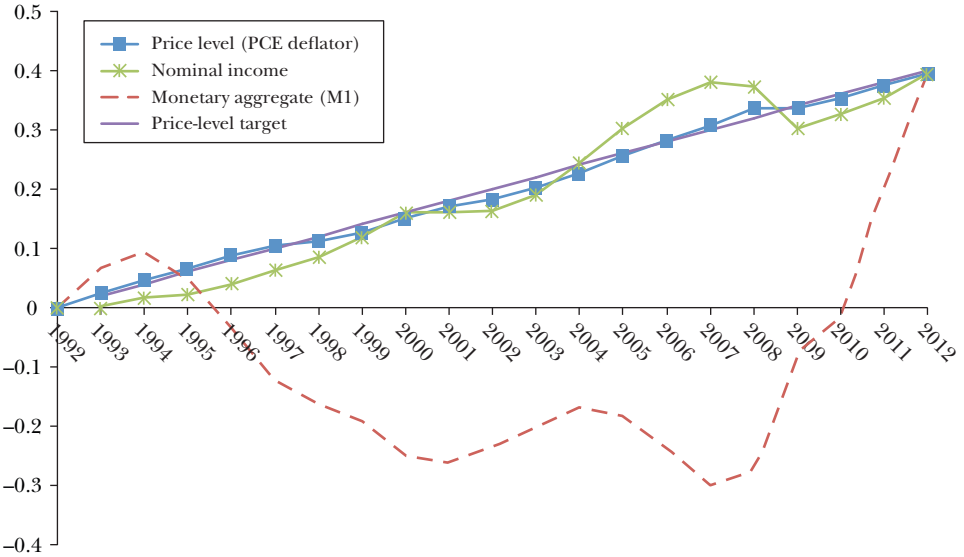
At the same time, questions about what the optimal inflation rate is, whether to target inflation or the price level, or how aggressively to adjust policy in response to unexpected changes in output, involve mostly technical considerations and to a lesser extent disputes on social value. Bureaucrats do not just implement policies, but also shape them. The central bank may be more effective in technical tasks where ability to incorporate quickly changing knowledge is more important than effort at meeting the goals in a strict mandate, and where redistribution is not an important consideration (Alesina and Tabellini 2007). If this is the case, some discretion may achieve an outcome that is closer to fulfilling the overall mandate, even if there is a thin line separating the principles handed to the central bank and the operational targets it sets for itself.

Dimension 2: The Choice of Long-Run Goals

Whether across time, or across countries, there is a strikingly high correlation between the change in the monetary base, the nominal interest rate, and the change in the price level over a period of 30 years (for example, Benati 2009). As Milton Friedman (1970) famously put it: "Inflation is always and everywhere a monetary phenomenon." Because long-run inflation imposes social costs and the central bank has almost perfect control over the quantity of banknotes in circulation and the amount of reserves that banks hold, and can also freely set the interest rate it pays on those reserves and the interest rate at which it lends to banks, it then follows that price stability is a natural long-run goal for a central bank. Indeed, this is true for all central banks of which I am aware, even though with remarkable frequency, actual policy gets focused on a succession of urgent short runs, and prices end up drifting away, as perhaps happened in the 1970s in the United States (Goodfriend 2007). Therefore, it is worth repeating that the central bank should be the agency responsible for establishing a stable nominal anchor.

This mandate leaves open the choice of the appropriate nominal anchor: for example, it can be based in some way on prices, on the quantity of the money supply, or on a measure of national income. My reading of the current literature is that price-level targets bring about less volatile long-run inflation without necessarily higher short-run volatility of output, when compared with measures of either money or nominal income. Monetary aggregates suffer from the important pitfall that financial innovation invariably leads to large fluctuations in the relationship between most broad measures of money and the price level. Moreover, while there are strong arguments for why price instability lowers welfare—for instance, because of the opportunity cost it imposes on holders of currency, or because of inefficient variability in relative prices if prices are set infrequently—research has struggled to

Figure 1
Comparing Long-run Nominal Anchors for the United States, 1992–2012



Source: Author using data from FRED database of the Federal Reserve Bank of Saint Louis.
Notes: Figure 1 uses US data for the last 20 years to plot the price level (the deflator on personal consumption expenditures), a monetary aggregate (M1), and nominal GDP, subtracting a trend from each of these last two series so that their value in 2012 is exactly the same as that for the price level. It also plots a hypothetical price-level target of 2 percent per year.

come up with arguments that are both persuasive and quantitatively large for why instability in monetary aggregates is costly per se, independent of price stability (Williamson and Wright 2010; Woodford 2010). As for nominal income, especially outside the United States over the past century, there was considerable uncertainty on the long-run rate of economic growth in many countries. A central bank can do little about this long-run rate of economic growth, but if it sought to achieve a pre-set nominal income target, this would lead to an unstable price level.[1]

Figure 1 illustrates this point using US data for the last 20 years. It plots the price level, a monetary aggregate (M1), and nominal GDP, subtracting a trend from each of these last two series so that their beginning value in 1992 and their ending value in 2012 is exactly the same as that for the price level. This detrending exercise generously assumes that, in designing policies that target monetary aggregates or nominal income, the central bank would know the long-run trends in velocity and real output. Nevertheless, as the figure shows, while the Federal Reserve policy of focusing on prices has led to a reasonably steady rate of increase in price level, it has

[1] This distinction should not be overstretched; in the short run, a flexible price-level target that responds to the output gap with a coefficient α is equivalent to a nominal income rule with a coefficient $1 - \alpha$ on the output gap.

produced large fluctuations in money and, to a lesser degree, in nominal income. If the Fed had instead pursued these alternative long-run goals, price levels would have likely been more volatile.

If the central bank is focused on a nominal anchor rooted in price data, choices still remain: should it consider the rate of inflation, or should it consider a target for the price level? A price level target might aim to have the price level rise at 2 percent a year, for example. This is not equivalent to a policy targeting inflation at 2 percent. Under an inflation target, bygones are bygones: if inflation exceeds the target in one period, the price level stays higher forever. With price-level targets, higher-than-planned inflation in one period must be followed by commensurately lower inflation in a later period to get back on the target.[2] The literature has identified at least six distinct theoretical arguments for why price-level targets dominate inflation targets at reducing macroeconomic volatility.

First, if a main cost of price variability is that it disrupts people's plans, which they only infrequently or imperfectly update, then it is undesirable to propagate a one-time mistake in inflation forever by making it a part of a permanently higher price level. If the price level is to provide a standard of measurement, much like the meter or the foot, but policymakers cannot prevent deviations in the real counterpart of these units, then they can at least strive to make these deviations short-lived (Hall 1984; Ball, Mankiw, and Reis 2005).

Second, if firms' plans involve sticky prices, then a price-level target has the virtue that forward-looking price setters will moderate how much they increase their prices following a positive inflationary shock today. After a positive inflationary shock, a price-level target translates into a commitment to bring prices back down in the future. Since price setters anticipate they may not be able to change their prices again for a while, they raise their prices by less in response to the shock, so the deviation of inflation from target is lower to start with (Woodford 2003).

Third, a price-level target provides a stronger commitment against the temptation that a central bank continually faces to surprise private agents with inflation. A central bank is tempted to surprise private agents with inflation because a surge of unexpected inflation can increase output and employment in the short run as well as reduce the real cost of government debt. Price-level targeting reduces the incentives for this classic inflation bias as it commits the central bank to undo any positive deviations of inflation from the announced target (Svensson 1999; Clarida, Galí, and Gertner 1999).

Fourth, with a price-level target, there is a smaller benefit of indexing contracts to past inflation than with an inflation target. With a price-level target, workers and firms know that the price level will revert back to its original path after any deviation,

[2] Here is a numerical example: Imagine you have a 0 percent inflation target, but this year some shock hits such that inflation ends up being 1 percent. Then the price level goes from 100 to 101. The next year, the policy says aim for 0 percent inflation again, so prices stay at 101 the following year, so the price level is 101 forever after. If instead your policy is to have the price-level target rise 0 percent, after the price goes to 101, your policy would have you get back to 100 resulting in –1 percent inflation the next year. You would get back to 100 right away and forever.

while with an inflation target, they realize that it requires indexation to have wages and prices keep up with past inflation. Therefore, because fewer firms and workers choose to index their prices and wages, a price-level target frees these prices to become more flexible to react to other shocks (Amano, Ambler, and Ireland 2007).

Fifth, price-level targeting results in a lower cost of capital for the economy relative to inflation targeting. With inflation targeting, the price level follows a random walk: since surprises are never reversed, as one looks further into the future, prices can drift further from the starting point and the variance becomes unbounded. The real payment on nominal assets becomes riskier, which raises their risk premium and therefore the cost of capital in the economy (Fischer 1981).

Sixth, and particularly relevant today, a price-level target is an effective way to guarantee that if a shock pushes the economy into low inflation and zero nominal interest rates, then the central bank automatically commits to higher future inflation escaping from the liquidity trap (Eggertsson and Woodford 2003).

In spite of all of these theoretical virtues, price-level targets have only very rarely been adopted by actual central banks. While each theoretical mechanism above has some evidence to back it, the policy of adopting a price-level target has not been tested. One common objection is that the central bank would have trouble communicating the ever-changing goal for the inflation rate that comes with a price-level target (where positive inflation surprises must come with negative inflation adjustment), to a public that is accustomed nowadays to focusing on 2 percent inflation every year. Yet, over the past few decades, the Federal Reserve has shifted from using targets for monetary aggregates, to targets for the federal funds rate, to targets for inflation. People adapted. In the last few years, the Federal Reserve has offered more frequent speeches, policy announcements about the "quantitative easing" bond purchases, and forward guidance about commitments to a future path for the federal funds interest rates. Again, people adapted. Price-level targets do not seem like such a radical change, in comparison. Another objection is that if agents form expectations of future inflation adaptively as a function of past inflation, price-level targeting will increase instability (Ball 1999). But this begs the question of why agents, even backward-looking ones, would use past inflation instead of the past price level to form their expectations in an economy with a price-level target.

The main reason why the Fed has not discussed adopting a price-level instead of an inflation target is probably empirical. Figure 1 plots what a 2 percent annual rise in the price-level would look like, and it is very close to the actual evolution of the price level. The distinction between inflation targeting and price-level targeting may therefore seem empty of empirical substance. This conclusion would be incorrect. In modern macroeconomics, policy targets and rules affect the expectations and choices of private rational agents, so that even if by chance the path of the price level under policies of price-level and inflation targeting is the same, the six channels described above will lead to higher welfare and more stable real activity under a price-level target. Moreover, modern econometrics teaches us that it would take a great deal more data than 20 years to distinguish

between stable inflation at 2 percent and a stable price level growing at 2 percent a year. A more intriguing possibility revealed by Figure 1 is that the Fed's actions may already be close to a price-level target even as it describes its actions as following an inflation target.

Whether a central bank chooses the inflation rate or the price level, it still faces a question of how to measure the price level. Most measures of inflation are strongly correlated at low frequencies, like a decade or more, but they can differ from each other substantially over shorter periods of several years. Having to wait more than a decade for feedback would obviously make it difficult to assess the central bank's performance. Another difficulty is that even if a broad measure of changes in the cost of living captures social welfare, its variation is dominated by relative-price changes associated with structural changes to potential GDP. If the central bank accepts that monetary policy does not affect potential GDP, then these long-run changes in relative prices are beyond its control. This last argument suggests two ways to construct an appropriate yearly measure of long-run target inflation: look at the change in prices that are by construction uncorrelated with output at low frequencies such as a decade or more (Quah and Vahey 1995), or find a measure of "pure inflation" that filters out all relative-price movements and captures only the changes in the unit of account that the central bank can affect (Reis and Watson 2010).

Finally, since real outcomes are what matter to people, it is tempting to suggest that the central bank should also have a real long-run target. However, there is almost a consensus around the Friedman–Phelps claim that the long-run Phillips curve is vertical, meaning there is no permanent trade-off between inflation and real activity.[3] This implies that the central bank cannot use its power over the price level to affect output or employment in the long run, so there is no point in asking it to focus on a long-run real target. Moreover, even if the central bank had such a target, if we do not understand reasonably well the specifics of the long-run tradeoff between prices and output, setting monetary policy in a way that seeks to achieve a real long-run target could have the undesired consequences of inflation or deflation. These two arguments have convinced most central banks not to have a real long-run target, and the large bulk of the literature supports this choice.

However, it is worth remembering that the empirical evidence that there is zero association between the rate of inflation and the rate of economic growth and employment is quite weak. If inflation goes well into the two-digits, the data seem to suggest that there is a negative association with growth and employment. For inflation rates below 10 percent, the failure to reject the null hypothesis of no association involves confidence intervals wide enough that this failure should not be confused for positive evidence that the long-run Phillips curve is truly vertical.[4]

[3] For recent theoretical arguments for why the long-run Phillips curve may instead be upward or downward sloping, see Berentsen, Menzio, and Wright (2011) and Akerlof, Dickens, and Perry (2000), respectively.
[4] See Bruno and Easterly (1998) for the long-run evidence, and Svensson (2013) for a recent empirical argument for a non-vertical Phillips curve in Sweden.

Dimension 3: The Potential Role of Additional Short-Term Goals

In the short run, should a central bank focus solely on a nominal measure, or should it have a dual mandate and also consider some measure of real activity? There is compelling evidence, using multiple methods, time periods, and datasets, that monetary policy has a large and prolonged effect on real activity (among many others, see Christiano, Eichenbaum, and Evans 1999; Romer and Romer 2004b). Steering the economy using nominal interest rates is neither easy nor mechanical, and the debates over the strength and stability of the monetary transmission mechanism may at times seem endless (Boivin, Kiley, and Mishkin 2010). Yet, the history of the Federal Reserve suggests that whenever the central bank neglected the effect of its actions on output and employment, the economy suffered (Romer and Romer 2013). Because social welfare likely depends at least as much on people having a job and food on the table as it does on inflation, there is a strong argument for including some measure of real activity, like output or employment, in the objective function of the central bank (Friedman 2008).

However, including real activity as a goal is only relevant if 1) there is a short-run trade-off between unemployment and inflation—a downward-sloping Phillips curve—that the central bank can exploit, even if only imperfectly, and 2) stabilizing inflation per se does not guarantee by "divine coincidence" that real activity will also perform at its best possible level (Blanchard and Galí 2007). These issues are hotly debated today, but the current state of knowledge leans towards there being a Phillips curve that is downward-sloping in the short run, as well as a trade-off between price stability and real stability such that giving up some price stability can increase the real stability of an economy. As a starting point to exploring this literature, Mankiw and Reis (2010) offer a modern treatment of the theory behind the Phillips curve and Woodford (2010) of optimal stabilization policy. Therefore, this research suggests the case for a dual mandate that looks at both nominal and real outcomes, like the one for the Federal Reserve.

How to weight the nominal or price-based targets and the real-output or employment-based targets when they are in conflict remains open for discussion. At one extreme, the central bank could be quite patient at reversing increases in inflation, with the hope that gradualism will minimize the potential resulting recession, so that the long-term goal of price stability is reached with a lag of several years. At the other extreme, price stability can receive primacy over economic growth and employment, as in the case of the legislation guiding the European Central Bank. Different societies may choose different extents to which the price level is allowed to deviate from target if there is an output gap; this decision will be based on different social weights on the two goals, different opinions on the slope of the Phillips curve, and different estimates of how quickly inflation expectations adjust to news. Whatever the choice, the central bank can adjust to advances in knowledge in these parameters by changing the degree of gradualism in policy while remaining within what is known as flexible price-level targeting (Woodford 2007; Svensson 2010).

A more contentious debate is whether to have a tripartite mandate that also includes financial stability. After all, the two largest US recessions in the last

century—the Great Depression of the 1930s and the Great Recession that started in 2007—were associated with financial crises. Similar to the question of real targets discussed above, if financial stability is to be included as a separate goal for the central bank, it must pass certain tests: 1) there must be a measurable definition of financial stability, 2) there has to be a convincing case that monetary policy can achieve the target of bringing about a more stable financial system, and 3) financial stability must pose a trade-off with the other two goals, creating situations where prices and activity are stable but financial instability justifies a change in policy that potentially leads to a recession or causes inflation to exceed its target.

Older approaches to this question did not fulfill these three criteria, and thus did not justify treating financial stability as a separate criterion for monetary policy. Before the Fed was founded, seasonal and random changes in the demand for currency and reserves led to wide fluctuations in interest rates and to occasional bank failures and panics. The Fed was in part founded to supply an "elastic currency"—that is, to adjust the supply of money to accommodate these demand shocks. Yet the volatility of interest rates in these cases almost always comes with volatile inflation and real activity, so financial stability was aligned with the other goals, and in that sense did not seem to merit separate consideration. Moreover, deposit insurance and financial regulation conducted outside of the central bank already address many of the stability concerns related to shifts in the demand for liquidity. Another approach to defining financial stability was in terms of large asset price movements. Yet, at most dates, there seems to be someone crying "bubble" at one financial market or another, and the central bank does not seem particularly well equipped to either spot the fires in specific asset markets, nor to steer equity prices (Blinder and Reis 2005; Blinder 2006).

A more promising modern approach begins with thinking about how to define financial stability: for example, in terms of the build-up of leverage, or the spread between certain key borrowing and lending rates, or the fragility of the funding of financial intermediaries (for example, Gertler and Kiyotaki 2010; Cúrdia and Woodford 2010; Adrian and Shin 2010; Brunnermeier and Sannikov, forthcoming; among many others). This literature has also started gathering evidence that when the central bank changes interest rates, reserves, or the assets it buys, it can have a significant effect on the composition of the balance sheets of financial intermediaries as well as on the risks that they choose to take (Kashyap, Berner, and Goodhart 2011; Jimenez, Ongena, Peydro, and Saurina 2012). As a result, even for fixed output and prices, changes in the funding structure of banks, in their net worth, or in their perception of tail risk, can create a misallocation of resources that significantly lowers social welfare. While it is not quite there yet, this modern approach to financial stability promises to be able to deliver a concrete recommendation for a third mandate for monetary policy that can be quantified and implemented.

Dimension 4: The Choice of Central Banker(s)

Society can give a central bank a clear mandate with long and short-run goals, but eventually it must appoint individuals to execute that mandate, and they will

always have some discretion. Choosing the central banker is a complementary way to pick an objective function for the central bank. For example, Romer and Romer (2004a) argue in this journal that different chairmen of the Federal Reserve chose very different policies, in spite of an essentially unchanged legal mandate, mostly due to different views on the role and effects of monetary policy.

Most countries do not pick a single person to have absolute power over the central bank, but prefer a committee of several people. A committee has several advantages including the ability to pool information, the gains from having a diversity of views that must be argued for and against, the checks it provides against autocratic power, and the experimental evidence that committees make less volatile decisions (Blinder 2004). For these potential virtues to be realized requires that the committee members have different perspectives, supported by independent staffs, while sharing a common framework to communicate effectively and to come to agreements (Charness and Sutter 2012).

When a committee makes decisions, there needs to be a rule to aggregate the separate preferences of individuals. There is a long literature on voting rules that have some desirable properties, and there is little about the Federal Open Market Committee that requires special treatment (Vandenbussche 2006). A more interesting question is who should have a vote in the committee if the goal is to elicit talent and bring together different information. For example, is it useful to draw at least some of the membership of the committee from different regions of the country? On the twelve-member Federal Open Market Committee, the seven Washington-based members of the Board of Governors are joined by five heads of the existing twelve regional Federal Reserve banks. These regional Federal Reserve banks are not just local offices of the central bank, spread around the country to interact with and provide services to local communities, but actually have some autonomy and a say in monetary policy. The locations of the regional banks, and even the fact that there are twelve districts, resulted from delicate political equilibriums that only partly reflected economic considerations (Hammes 2001). The Federal Reserve Act leaves vague how the twelve regional banks should interact and work together (Eichengreen 1992).

In considering how regional interests are represented at the Fed, one should note there is evidence that US states share most of their risks (Asdrubali, Sorensen, and Yosha 1996). So even if regional governors had only the consumption of people in their region in mind, this fact would justify a focus on eliminating aggregate risk and ignoring idiosyncratic regional shocks. Might regional governors bring additional information that originates from or pertains to their region? Looking at the forecast performance for key macroeconomic variables, the members of the Fed Open Market Committee seem to add little value to the forecast produced by the staff at the Board of Governors (Romer and Romer 2008). Therefore, the case for having regional governors relies more strongly on promoting different perspectives and stimulating original thinking. Geographical distance and separate staffs and budgets may help to cultivate competition in the market for ideas in interpreting the data and arriving at policy proposals (Goodfriend 1999).

Monetary policy not only responds to economic shocks, but it can also be a source of aggregate risk that agents cannot insure against and that induces redistributions of wealth.[5] In a representative democracy, different age cohorts or business sectors may legitimately ask to be represented when these decisions are made. There are two counterarguments to such a request. First, the literature has so far not been able to determine the systematic direction in which monetary policy redistributes wealth across industries or stable groups in the population. While some people may be hurt in each decision to raise or lower the interest rate, if there is no persistent conflict, then it is hard to defend that some groups should permanently have a say when monetary policy decisions are made. Instead, policymakers can, and perhaps should, take into account the redistributions of wealth that their policies induce without having some members of the Federal Open Market Committee designated to stand for the interests of one group. Second, fiscal policy is a more targeted tool when it comes to distributing resources. Even if redistribution is a side effect of monetary policy, fiscal policies may be a preferable tool to undo its effects on the distribution of income, wealth, or consumption.

The Central Bank's Resources and Policy Tools

Like any other economic agent, central banks have limited real resources that constrain their policies (Reis 2013b). The tools of central banks include interest-rate policies that try to control one or more interest rates, quantitative policies choosing the size of the Fed's liabilities and its dividends, and credit policies determining the composition of the assets of the central banks. Designing the central bank requires making sure that these policies all respect the resource constraint of the central bank, and this suggests four more dimensions of central bank design.

Dimension 5: The Role of the Central Bank as a Dependable Source Of Revenue

It is an old adage in monetary policy that the central bank should not monetize fiscal deficits. History teaches that the surest way to produce inflation is to finance government budgets by printing money. Yet these statements are not quite correct. Almost all central banks issue reserves to buy government debt as part of their open-market operations. Printing money that pays for budget deficits is not a taboo, but rather the day-to-day workings of monetary policy. Moreover, when the interest paid on bank reserves is the same as the short-term return on government bonds, as it is today, then when a central bank uses reserves to buy these bonds it is just exchanging one government liability for another, making it hard to argue that this will have any dramatic impact on anything of relevance.

[5] See Bullard and Waller (2004) for some theory applied to central bank design, and Doepke and Schneider (2006), Berriel (2013), and Coibion, Gorodnichenko, Kueng, and Silvia (2012) for evidence on redistribution.

There is a clearer way to state this important wisdom. As part of its activities, the central bank will generate resources, which have three properties: First, these resources, in present value, come exclusively from the seignorage arising from money creation: that is, the resources arise because the central bank pays less-than-market interest on some of its liabilities in exchange for the service that they provide and at the same time earns market interest rates on the assets backed by these liabilities (Reis 2013b). Second, seignorage depends primarily on the level of inflation, since higher inflation taxes the holders of currency by lowering the value of this government liability relative to the goods it can buy; but the generation of substantial revenue requires very high inflation (Hilscher, Raviv, and Reis, in progress). Third, if the central bank pays out its net income every period, then its budget constraint will be respected regardless of the monetary policy that is chosen (Hall and Reis 2013). Governments will always, under fiscal stress, be tempted to demand that the central bank generate more resources and transfer them to the Treasury. Combining the three properties above, we know that 1) the resources come from seignorage, 2) which requires higher inflation, and 3) the central bank can feasibly make the transfers desired by the Treasury. This suggests that to keep prices stable in the long run, central bank design should allow the bank to resist fiscal demands.

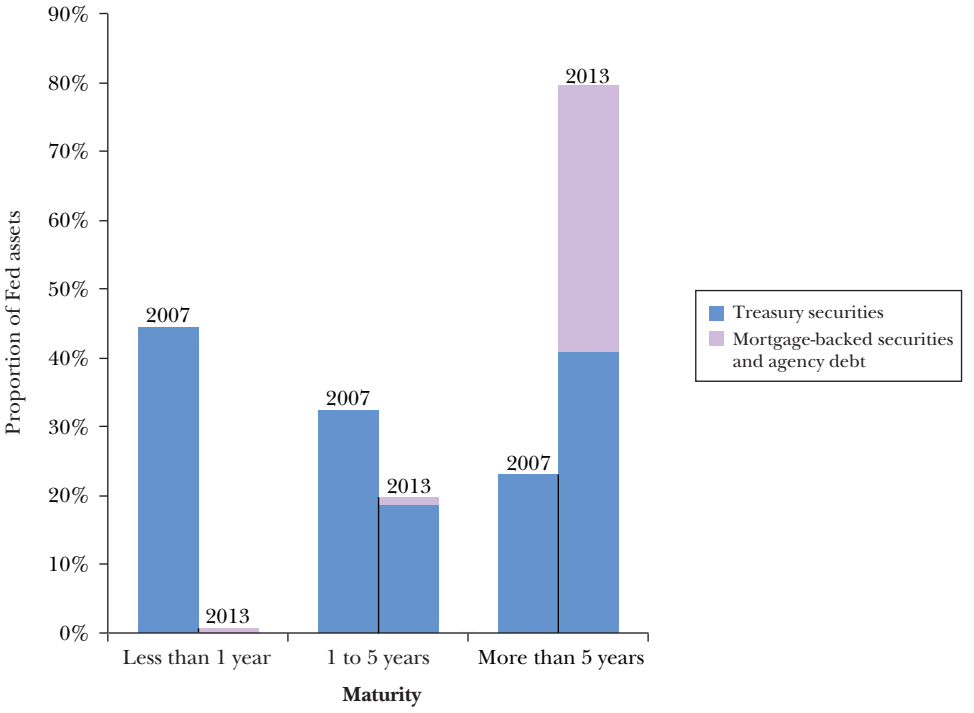
This lesson does not preclude considering the interaction between monetary and fiscal policy in determining inflation (for example, Sims 2013). It also does not deny that it may be optimal in some states of the world to generate fiscal revenues via inflation (Sims 2001; Chari, Christiano, and Kehoe 1991). It simply distinguishes between seignorage revenues, which are small and require high expected inflation, from the fiscal benefits from unexpected inflation that arise, for instance, by lowering the real value of public debt outstanding.

Dimension 6: The Importance of Fiscal Backing for the Central Bank

In conventional times, the Federal Reserve mostly holds government bonds of short maturities and implements monetary policy by buying and selling them from banks in exchange for reserves (Friedman and Kuttner 2010). Under this "old-style" central banking, using open market operations, the assets and liabilities of the Fed are close to riskless and they are matched in their maturity, so net income will almost always be positive (Hall and Reis 2013).

However, if the central bank pays interest on reserves and, especially, if it holds other assets that create a risk–maturity mismatch with its liabilities, sometimes the net income of the central bank will be negative. This is true of the Federal Reserve today as it has embraced a "new-style" central banking where long-term securities that are either issued by the Treasury or implicitly guaranteed by it (agency debt and mortgage-backed securities) now dominate its assets, as shown in Figure 2. The figure shows that in 2007, almost all Federal Reserve assets were in the form of Treasury securities, mainly of short maturities. By 2013, a large share of Federal Reserve assets were in the form of mortgage-backed securities and agency debt, and the Fed primarily held long-term securities.

Figure 2
The Assets of the Federal Reserve by Maturity and Type: Old-Style (December 31, 2007) versus New Style (September 26, 2013)



Source: Author using data from the Federal Reserve statistical release, table H.41.

Most central banks have a rule, more or less explicit, of handing over their positive net income to the Treasury, and the Fed has done so in every year of its existence. However, if there is no transfer in the other direction when income is negative, then the budget constraint of the central bank will not hold (Hall and Reis 2013). Something must give. One plausible consequence is that inflation rises above target so that seignorage is higher and net income does not become negative. Even if this event is rare, expectations of higher inflation can set in even while net income is positive.

Preventing this outcome requires giving fiscal backing to the central bank. One design principle that achieves this backing is to commit the Treasury to transfer resources to the central bank if net income is negative. An alternative is to allow the central bank to build a deferred account against the Treasury when net income is negative, which is then offset against future positive income. Such steps require strict audits of the Fed's accounts, limits to the risks it can take, and an upper bound on this backing, none of which are easy to specify.

A bolder measure that is simpler to implement would be to sever completely the resource link between the central bank and the Treasury. In that case, instead of sending its net income to the fiscal authority, the central bank would directly fund a public good or a public trust fund. As long as the use of funds does not require a stable stream of income, such that periods of negative net income and held-back dividends are not too disruptive, and if its direct recipients do not have the political power to try to extract more from the central bank, the problems raised above would be reduced. (Funding basic research in the social sciences is a provocative candidate!) Under this structure, the central bank would not need the Treasury to provide fiscal backing. The present value of seignorage would become the relevant constraint to cover possible losses and to restrain the risks the central bank takes.

Dimension 7: The Set of Assets Held by the Central Bank

Usually, the Federal Reserve only intervenes directly in the small federal funds market for overnight funds, where not even most banks can trade. Yet the central bank ultimately wants to affect the spending, pricing, and investment decisions of many or most economic agents in order to reach its macroeconomic goals. It must therefore rely on investors, working individually, to move resources across financial markets given the new federal funds rate, ultimately resulting in financial returns moving in all financial markets, both across types of risk and across maturities. Profit seeking will transmit monetary policy choices to the relevant interest rates for the agents' marginal decisions as long as financial markets function reasonably well, meaning that the relationships across securities with different risk and liquidity premia are fairly constant and predictable so that altering one interest rate will create a series of arbitrage opportunities that will end with all interest rates being affected. As long as these conditions hold, even if the central bank could buy other assets, this would make no difference in the effects of policy (Wallace 1981).

Between 2007 and 2009, the Federal Reserve more than doubled its liabilities, from less than $900 billion to slightly more than $2 trillion, acquiring a myriad of other assets that had different risks, maturities, and counterparties. The Fed started making loans to banks, primary dealers, and money market funds; it accepted as collateral commercial bonds as well as auto, student, and small business loans; and it participated in the government bailouts of Bear Sterns and AIG (Reis 2009; Bernanke 2012). Most of these assets were quickly sold as the financial crisis subsided and none of them are left on the balance sheet today. While setting interest rates, and choosing or adjusting the size of its assets are necessarily part of monetary policy, credit policies that change the composition of the assets that a central bank holds are more controversial (Goodfriend 1994). However, in a financial crisis, the central bank has a need, a means, and an ambition to do more than usual and engage in these credit policies. The need arises if cuts in the interest rate in the federal funds market do not lower rates in other financial markets. Perhaps because investors are constrained in their ability to borrow, arbitrage across financial markets will not function well. In this case, the usual mechanism for transmitting monetary policy across interest rates is broken, and purchasing other assets is a

way to bypass it. The means comes because, if markets are quite illiquid, then even the relatively small-scale purchases by central banks can significantly raise security prices and lower their yields (Krishnamurthy and Vissing-Jorgensen 2011). Finally, the ambition is that, if a combination of illiquidity and limits to arbitrage leads relative prices of financial assets to be distorted, then there will be a misallocation of resources that the central bank may be able to correct.

On the other side, there are several objections to a central bank engaging in credit policies. The central bank may realize significant losses, a risk which is greatly magnified with credit policies. Furthermore, if the markets are illiquid enough for the central bank's purchases to make a difference when buying, they are also potentially likewise illiquid enough for it to have trouble selling when it wants to—at least without incurring large losses. Moreover, even when the central bank lends against strong collateral to failed banks, if this keeps nonviable entities operating with growing losses and deposits, it increases the potential losses that deposit insurance will eventually have to bear (Goodfriend 2011). It is also tempting for the central bank to become overconfident about its ability to detect and correct financial market mispricings and to jeopardize the focus on its macroeconomic objectives. Moreover, correcting market distortions is typically the domain of tax and regulatory policy, not central banking.

A final objection is that aggressive credit policy exposes the central bank to legitimate political questions of why some firms, markets, or securities were chosen for support and not others. While conventional buying and selling of government bonds does not clearly benefit one firm or sector, credit policies have clear redistributive effects. At the same time, they also expose the central bank to lobbying pressure from financial market participants. Both will likely get in the way of the central bank's goals (Reis 2013a). A different type of pressure and temptation may come from within the central bank. Without a clear policy rule forbidding the bailing out of systemically important financial institutions, it will always be optimal to do so to avoid a larger crisis; however, the expectation of a bailout may create incentives for banks to become larger, take on more risk, and correlate their exposure so that they become systemically important and thus prime candidates for bailouts (Goodfriend 1994; Stern and Feldman 2004; Farhi and Tirole 2012; Chari and Kehoe 2013).

Given so many virtues as well as objections to credit policy, thoughtful design of a central bank likely puts some restrictions on the assets that the central bank can buy. At one extreme, the policy could be the one that the Federal Reserve faced in 2007, of having to justify unconventional policies to Congress as being due to "unusual and exigent circumstances," which is a fairly vague standard and thus not difficult to meet. At the other extreme, if we judge that there is too much of a temptation for the central bank to find a way to get around the rules, then a strict "buy only Treasuries" rule may be the answer (Goodfriend 2011). Even in this case, the central bank would still be able to shift between short-term and long-term government bonds. These "quantitative easing" policies expose the central bank to maturity risk—when policy becomes contractionary and markets start expecting an upward-sloping path for short-term interest rates, long-term bond prices will fall,

inducing capital losses on the Fed's portfolio—but most empirical estimates of this risk come up with relatively small losses in worst-case scenarios that could easily be written off against a few future years of positive earnings (Hall and Reis 2013; Carpenter, Ihrig, Klee, Quinn, and Boote 2013; Greenlaw, Hamilton, Hooper, and Mishkin 2013).

Between these two extremes, many alternatives are plausible. One concrete restriction would be to prevent the central bank from taking part in ad hoc interventions targeted at specific institutions: that is, the central bank would have to stick to a general policy that is applied uniformly at arms-length across the entire financial sector. This would prevent the Federal Reserve from being able to resolve a particular financial institution, as happened in the bailouts of Bear Stearns in March 2008 and AIG in September 2008. A tighter restriction would require the central bank to purchase only securities for which there is a market price, with enough market participants that compete for the central bank's funds. A stronger version of this rule would prevent the Federal Reserve from intervening in any over-the-counter financial markets. A weaker version could draw from the experience in industrial organization and require the central bank to run a reverse auction, with even a small set of institutions, designed to ensure that its purchases are allocated efficiently.

Dimension 8: The Payment of Interest on Reserves

When a central bank pays interest on the reserves deposited by banks, it can use quantitative policy to satisfy the liquidity needs of the economy. By choosing both the interest on reserves and the federal funds rate, the central bank can at the same time set the short-term interest rate that will determine inflation, as well as affect the amount of liquidity held in the banking sector (Kashyap and Stein 2012). Separately from its interest-rate policy, the Fed can have a large balance sheet, like it does at present, if society wants to keep a larger share of wealth in money-like investments, or the balance sheet can quickly shrink to the pre-crisis levels, all without consequences or dangers for the rate of inflation. Most central banks around the world have the authority to pay interest on reserves, and the Fed joined them in October 2008.

The central bank could go one step further along these lines (Hall 1986). A general principle of economic efficiency is that the marginal cost of producing a good should equal its marginal benefit to society. In monetary economics, this principle leads to the "Friedman rule" which has been reaffirmed repeatedly in a wide variety of models of the demand for money (Lucas 2000; Chari and Kehoe 1999; Lagos and Wright 2005). Applied to reserves, note that it costs nothing to add an extra unit to a bank's reserves balances at the Fed, and that the benefit, or opportunity cost, of holding reserves is the overnight federal funds rate at which banks could lend these funds to other banks. Therefore, the Friedman rule dictates that the central bank should pay an interest rate on overnight reserves equal to the overnight federal funds rate, thus satiating the market with as many reserves as it wants. This "floor policy" would make the interest rate on reserves the primary

instrument of monetary policy and, unlike the federal funds rate, it is perfectly set and controlled by the central bank (Goodfriend 2002; Woodford 2003). There is a strong case for requiring the central bank to not just pay interest on reserves, but also to always follow the Friedman rule via a floor policy.

Transparency, Commitments, and Accountability

Designing the central bank also involves choosing the rules of the game that it will play with private agents. These include what the central bank will reveal, what it will commit to do, and how it will be held accountable to its promises and goals. For example, central bank announcements can be valuable to economic agents because of the information about the economy that they provide, and because they can justify current policy and reveal likely future policies. In some circumstances, like in the "liquidity trap" setting when interest rates have been pushed to zero, forward guidance about keeping short-term interest rates low for a prolonged period of time in order to reduce long-term interest rates right away is one of the few effective tools left for the central bank to stimulate aggregate demand (Eggertsson and Woodford 2003).

Dimension 9: The Importance of Announcements and Commitments

Even when policymakers are benevolent in their intentions, the history of government includes many mistakes and blunders because of incompetence, shortsightedness, hubris, false models, or bad ideas. Milton Friedman (1968) strongly argued that rules for monetary policy are an effective way to prevent mismanagement. The difficulty with most strict instrument rules, such as Friedman's proposal for a constant growth rate of the money supply, is that our understanding of economics is far from complete. Knowledge is still evolving quickly, our data is imperfect, and our theories have uncovered few relationships that are invariant to the policies adopted. Therefore, situations typically arise rather quickly in which any rule becomes not slightly, but grossly, sub-optimal. By the time that Friedman's proposal for a constant growth rate of the money supply became popular in the early 1980s, this policy floundered as the velocity of money started fluctuating wildly. Even with a rule, there is still a significant role for what, for lack of a better word, may be called judgment.

Yet a remarkable result in economics shows that even if policymakers have the same goals and information as private agents, and even if they exercise their judgment to do what seems best, we may end up with clearly inferior outcomes (Kydland and Prescott 1977). The reason is that, even if the policymaker has no initial desire to mislead private agents, after they have made their choices, the incentives of the policymaker change, and it may then decide to implement a different policy from the one that was announced. If agents anticipate this behavior, society may end up worse off. Designing the central bank to tie the hands of policymakers along some dimensions may then improve welfare.

There is a long literature investigating different forms for implementing this commitment (for example, Stokey 2003; Alesina and Stella 2010). One design principle is that if there is a temptation—for instance to generate excess inflation in an attempt to maintain positive output gaps forever—then removing the temptation in the first place (if possible) eliminates the source of the time inconsistency (Blinder 1998). One design that has been adopted all over the world with some success gives central bank governors a long but nonrenewable term of office, which then limits the ability of politicians to remove them or exert pressure to temporarily lower unemployment to win elections (Crowe and Meade 2007).

This does not make the deeper problem disappear: after all, sometimes it will be socially optimal for inflation at a later time to be above what agents had been led to expect. Moreover, if this generates a temptation for the central bank to renege on previous commitments to private agents, it likewise generates a temptation for the government to dismiss a central banker who wished to follow the pre-existing understanding or to alter the terms of the contract it had offered the banker. The literature has suggested that if the central bank makes a public commitment and cares about its reputation in keeping to this commitment, it may be able to bring about a favorable equilibrium (Barro and Gordon 1983; Backus and Driffill 1985). Several countries have done so by adopting rules requiring that the central bank target inflation, and no country has so far abandoned such a rule. The key point here is not to make a case for inflation targeting in particular, but rather to argue for the importance of a commitment by the central bank to announce its projections for the variables in its objective function as transparently as possible.[6] Publishing periodic inflation reports, like the Bank of England or Norway's Norges Bank do, is a way for the central bank to justify its actions and commit to forecasts of its targets. Based on such reports, economic agents can infer whether the policymaker is sticking to its objectives or trying to mislead them. They can compare outcomes with previous announcements and adjust their future actions and expectations to punish policymakers that are perceived to be reneging on their commitment.

Dimension 10: Choosing the Extent of Transparency

While the US Federal Reserve does not publish an inflation report, the Federal Open Market Committee releases a statement and holds a press conference right after it makes decisions and, with varying but increasingly short delays, it makes available the votes, forecasts, and arguments made by each governor, releasing all transcripts after five years. How far can transparency go? Once information has been internally produced, revealing information has a cost that is close to zero on one side of the scale, and positive benefits on the other side of the scale arising from commitment, from improving public information about the economy, and

[6] Chari and Kehoe (2006) associate the adoption of clear rules with addressing the time-inconsistency problem, Svensson (2003) explains targeting rules, Giannoni and Woodford (2010) provide a very general theoretical treatment, and Bernanke and Mishkin (1997) early on defined inflation targeting as a broad framework where communication and transparency are central.

from providing forward guidance about future monetary policy (Woodford 2005; Blinder, Ehrmann, Fratzscher, De Haan, and Jansen 2008). Moreover, there is a prima facie argument for public institutions to be open in order to be democratically legitimate. The question should therefore be put backwards: is there any strong argument for the central bank *not* to reveal everything it knows?

It is arguably appropriate for the central bank to keep to itself the private information it receives from banks it regulates. It may also lead to a more productive internal discussion if the central bank does not reveal every step of its deliberative process too soon after monetary policy decisions. But both of these points are minor exceptions to the general rule of openness, and there is as much risk of these exceptions being violated as there is of them being overstretched.

Of greater concern is whether central bank announcements foster confusion rather than better understanding. A small literature uses models where agents have cognitive or informational limitations that can lead them to misinterpret public information. If the central bank reveals signals about the state variables that agents use to make decisions but it does so in a manner that buries the information in statistical noise, or if it announces the information too soon before it becomes relevant, or if it focuses on variables that are too far from the policy targets, then it is possible to lower the precision of private actions and achieve worse outcomes (Reis 2011; Eusepi and Preston 2010; Gaballo 2013). Moreover, public signals may lead agents to collect less private information, making the price system less efficient and inducing an overreaction of expectations to noisy public signals (Morris and Shin 2002, 2005; Amador and Weill 2010). But while the literature has developed theoretical arguments for why less information *might* raise welfare in a model, it has not convincingly shown that these effects are likely to be present (Roca 2010), quantitatively important (Svensson 2006), or empirically significant (Crowe 2010) in reality. Moreover, in these models, what is usually better than revealing less information is to optimize the form and timing of announcements. The work of national statistical agencies is subject to the same caveats, and they respond by working harder at being informative and clear, not by embracing obscurantism.

Dimension 11: Picking the Channel(s) of Communication

The Federal Reserve has a particular decentralized structure with seven members of the Board of Governors in the center and twelve Federal Reserve Bank presidents as independent poles. Having this many actors in monetary policy poses challenges for making public announcements. First, a decentralized structure makes it difficult to have model-based monetary policy. There is an economic model in Washington, DC, that is used to make staff forecasts, but the district presidents have no input into it. In turn, each of the district presidents has his or her own model and set of predictions. It is hard to explain monetary policy decisions, and especially to announce and commit to future policy and targets, when so many decision makers are partially revealing their views and plans (Ehrmann and Fratzscher 2007). Second, many voices raise the danger of confusing disagreement with uncertainty, in spite of the two being conceptually distinct and empirically only weakly

related (Mankiw, Reis, and Wolfers 2004; Zarnowitz and Lambros 1987). Third, the decentralized structure makes it harder for agents to coordinate on the public signals provided by policy. Some research has suggested that to aid coordination, the central bank could have fewer speeches, which would be more precise and targeted at different groups in the population (Chahrour 2013; Morris and Shin 2007; Myatt and Wallace, forthcoming).

While none of these problems can be completely solved, all of them are ameliorated with more information, including requiring each member of the Federal Open Market Committee to justify his or her views and to report the numerical forecast distributions that support these views. The literature offers few objections to giving the central bank a general mandate to be as transparent as possible while leaving policymakers some discretion on how to implement this mandate.

Dimension 12: The Accountability of the Central Bank

Transparency is a, or perhaps *the*, way of achieving accountability. If the central bank is open about its objectives, its procedure, and its views of the future, that will go almost all the way towards being accountable in its missions to society as a whole (Blinder 2004).

Political oversight goes hand in hand with accountability. The seven members of the Board of Governors of the Federal Reserve are appointed by the President, confirmed by Senate, and periodically answer to Congress. In that sense, both the executive and legislative powers, and the public that elected them, are represented. The overlapping terms for the governors ensure that different waves of those holding political power have an influence, which research has suggested reduces the likelihood of the central bank becoming captured by partisan governors (Waller 1989, 2000).

The regional structure of the Federal Reserve makes power more diffuse, so it is in principle harder for the central bank's actions to be taken over by one particular interest group (Friedman and Schwartz 1963). The 12 presidents of the regional Federal Reserve banks each answer to a board of nine members: three appointed by the Board of Governors, three from the local community, and three from the banks in their district. After the passage of the Dodd–Frank Act of 2010, banks no longer have a vote appointing the president. An interesting open question is whether banks from that district should be singled out, either positively in terms of having three reserved seats in the board, or negatively in terms of having no vote.

Conclusion

This paper has discussed 12 dimensions of central bank design. Table 1 summarizes the recommendations, together with the questions it left unanswered, and an assessment of the Federal Reserve System at present. Three broad issues have been pervasive throughout.

The first issue is central bank independence. While many have defended the virtues of central bank independence in general for preventing the tendency of

Table 1
Dimensions of Central Bank Design

Dimensions	*Suggestions*	*Open questions*	*Federal Reserve*
The strictness of the central bank's mandate	Clear on main goals, otherwise give discretion	Adopt numerical or qualitative targets?	Vague
The choice of long-run goals	Price-level target as nominal anchor	What measure of inflation to use? Include real target?	To provide a nominal anchor
The potential role of additional short-term goals	Dual mandate with clear weights	Tripartite mandate including financial stability?	Dual mandate, price and real stability
The choice of central banker(s)	Committee that shares goals but competes on ideas	Should it consider distributional effects of policy?	Peculiar regional structure
The role of the central bank as a dependable source of revenue	Central bank should not yield to Treasury's demands	How should monetary and fiscal policy interact?	The Fed is independent from the Treasury
The importance of fiscal backing for the central bank	Central bank with a deferred account on the Treasury	Sever the resource link between bank and the Treasury?	Untested until it has negative income
The set of assets held by the central bank	Treasuries at all maturities, other assets in crises but with some limits	Forbid ad hoc interventions that are not arms-length?	Wide in the past, narrower in the future
The payment of interest on reserves	Yes, definitely	Should it always equal the short-term market rate?	Friedman rule at present, future to be seen
The importance of announcements and commitments	Policymakers with long-term mandate and publish inflation reports	How to keep a reputation?	Increasing role through forward guidance
Choosing the extent of transparency	Be as transparent as possible	What is the best timing and form of communication	Rapidly improving, revealing more and sooner?
Picking the channel(s) of communication	All committee members should report their views	How to have model-based policy and diversity?	Rapidly improving, frequent and clear speeches
The accountability of the central bank	Be transparent, have overlapping terms of office	Should banks be singled out as stakeholders?	Strong political oversight, peculiar role of banks

democratic politicians towards ever-higher inflation, looking at more specific questions led to a more mixed message. Even if there is a case for central banks to independently conduct the operations of monetary policy, democratic principles would imply that society would still choose the goals of monetary policy. Committing to a stable long-run nominal anchor may reduce the costs of price uncertainty, but that is not the same as having a fanatic central banker committed to 2 percent

inflation at all times, and research shows that a flexible price-level target may be able to lower the variance of inflation and real activity. In turn, releasing the central bank from the duty to raise seignorage to make transfers to fiscal authorities does not imply that the central bank can assume large risks through unchecked credit policies. Moreover, even if central bankers are appointed to long terms that are independent from political pressure, so that they will not be tempted by the siren lure of unexpected easy money, the goal of avoiding monetary policies that are inconsistent over time also requires that the policymakers are politically accountable and transparent.

The second broad topic was the level of decentralization of the central bank, and in particular of the Federal Reserve. There are reasons to be skeptical of the ability of the Fed's regional structure to reconcile different business interests or to produce new information, and having so many voices raises difficulties for effective communication. At the same time, a decentralized structure makes different actors accountable and fosters the competition of ideas and perspectives. It is harder to argue persuasively that this decentralization should be tied to geography and very hard to justify the current structure of the Federal Reserve System as optimal if one were starting from scratch. The best structure to maximize advantages and minimize disadvantages remains an open question.

The final broad topic was the use of unconventional policy. During a financial crisis, possibly including being stuck in a liquidity trap, the economics literature has put forward arguments that support price-level targets, forward guidance in setting interest rates, paying interest on reserves, allowing the Fed's balance sheet to grow, or changing the maturity of the Fed's holdings of government securities. Yet there are strong objections to letting the Fed hold any type of assets, especially as the risks that comes with them exposes the central bank to potentially large losses of resources, as well as to pressure and scrutiny by those who benefit or lose from those purchases. Moreover, because controlling inflation requires fiscal backing from the Treasury, there must also be limits on the risks to the central bank's net income. More generally, institutional design rules that do not cover exceptional times are incomplete, and the analysis above suggested principles that apply during crises and normal times.

There are many other design issues that were not addressed, especially concerning financial regulation (as discussed by Gorton and Metrick, this issue; Blinder 2010). The broader message of this paper is that designing a central bank need no longer involve a resort to hunches, old aphorisms, or vague platitudes. Diverse tools and models, drawn from different branches of economics, can come together in informing this particular application of mechanism design.

■ *I am grateful for comments from Alan Blinder, Anil Kashyap, David Romer, and Mike Woodford, and especially from the editors of this journal. An earlier draft of paper was written for the National Bureau of Economic Research conference "The First 100 Years of the Federal Reserve," which took place in Cambridge, Massachusetts, on July 10, 2013.*

References

Adrian, Tobias, and Hyun Song Shin. 2010. "Financial Intermediaries and Monetary Economics." Chap. 12 in *Handbook of Monetary Economics,* Vol. 3A, edited by Benjamin M. Friedman and Michael Woodford, 601–650. Elsevier.

Akerlof, George A., William T. Dickens, and George L. Perry. 2000. "Near-Rational Wage and Price Setting and the Long-Run Phillips Curve." *Brookings Papers on Economic Activity* 31(1): 1–60.

Alesina, Alberto, and Andrea Stella. 2010. "The Politics of Monetary Policy." Chap. 18 in *Handbook of Monetary Economics,* Vol. 3B, edited by Benjamin M. Friedman and Michael Woodford, 1001–1054. Elsevier.

Alesina, Alberto, and Guido Tabellini. 2007. "Bureaucrats or Politicians? Part I: A Single Policy Task." *American Economic Review* 97(1): 169–179.

Amador, Manuel, and Pierre-Olivier Weill. 2010. "Learning from Prices: Public Communication and Welfare." *Journal of Political Economy* 118(5): 866–907.

Amano, Robert, Steve Ambler, and Peter Ireland. 2007. "Price-Level Targeting, Wage Indexation and Welfare." Unpublished paper, Bank of Canada.

Asdrubali, Pierfederico, Bent E. Sorensen, and Oved Yosha. 1996. "Channels of Interstate Risk Sharing: United States 1963–1990." *Quarterly Journal of Economics* 111(4): 1081–1110.

Backus, David, and John Driffill. 1985. "Inflation and Reputation." *American Economic Review* 75(3): 530–38.

Ball, Laurence. 1999. "Efficient Rules for Monetary Policy." *International Finance* 2(1): 63–83.

Ball, Laurence, N. Gregory Mankiw, and Ricardo Reis. 2005. "Monetary Policy for Inattentive Economies." *Journal of Monetary Economics* 52(4): 703–725.

Barro, Robert J., and David B. Gordon. 1983. "Rules, Discretion and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics* 12(1): 101–121.

Benati, Luca. 2009. "Long-Run Evidence on Money Growth and Inflation." European Central Bank Working Paper 1027, March.

Berentsen, Aleksander, Guido Menzio, and Randall Wright. 2011. "Inflation and Unemployment in the Long Run." *American Economic Review* 101(1): 371–98.

Bernanke, Ben S. 2012. "The Federal Reserve and the Financial Crisis." March. Lectures delivered at George Washington University.

Bernanke, Ben S., and Frederic S. Mishkin. 1997. "Inflation Targeting: A New Framework for Monetary Policy?" *Journal of Economic Perspectives* 11(2): 97–116.

Berriel, Tiago C. 2013. "Nominal Portfolio Choice and the Redistribution Effects of Inflation Surprises." April. Unpublished paper, FGV/EPGE Rio de Janeiro.

Blanchard, Olivier, and Jordi Galí. 2007. "Real Wage Rigidities and the New Keynesian Model." *Journal of Money, Credit and Banking* 39(s1): 35–65.

Blinder, Alan S. 1998. *Central Banking in Theory and Practice.* MIT Press.

Blinder, Alan S. 2004. *The Quiet Revolution: Central Banking Goes Modern*. Yale University Press.

Blinder, Alan S. 2006. "Monetary Policy Today: Sixteen Questions and about Twelve Answers." In *Central Banks in the 21st Century* (Proceedings of an international conference sponsored by the Banco de España, held June 8–9, 2006), edited by Fernández de Lis and F. Restoy, 31–72. Banco de España.

Blinder, Alan S. 2010."How Central Should the Central Bank Be?" *Journal of Economic Literature* 48(1): 123–33.

Blinder, Alan S., and Ricardo Reis. 2005. "Understanding the Greenspan Standard." In "The Greenspan Era: Lessons for the Future." Proceedings of symposium sponsored by The Federal Reserve Bank of Kansas City, August 25–27, 2005, Jackson Hole, Wyoming, pp. 11–96. Available at: http://www.kc.frb.org/publications/research/escp/escp-2005.cfm.

Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen. 2008. "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." *Journal of Economic Literature* 46(4): 910–45.

Boivin, Jean, Michael T. Kiley, and Frederic S. Mishkin. 2010. "How Has the Monetary Transmission Mechanism Evolved Over Time?" Chap. 8 in *Handbook of Monetary Economics*, Vol. 3A, edited by Benjamin M. Friedman and Michael Woodford, 369–422. Elsevier.

Brunnermeier, Markus K., and Yuliy Sannikov. 2013. "Redistributive Monetary Policy." In *The Changing Policy Landscape*, Proceedings of a symposium sponsored by the Federal Reserve Bank of Kansas City, held August 30–September 1, in Jackson Hole, Wyoming, pp. 331–384. Available at: http://www.kansascityfed.org/publications/research/escp/escp-2012.cfm.

Bruno, Michael, and William Easterly. 1998. "Inflation Crises and Long-Run Growth." *Journal of Monetary Economics* 41(1): 3–26.

Bullard, James, and Christopher J Waller. 2004. "Central Bank Design in General Equilibrium." *Journal of Money, Credit and Banking* 36(1): 95–113.

Carpenter, Seth B., Jane E. Ihrig, Elizabeth C. Klee, Daniel W. Quinn, and Alexander H. Boote. 2013. "The Federal Reserve's Balance Sheet and Earnings: A Primer and Projections." Finance and Economics Discussion Series paper 2013-03, US Federal Reserve Board.

Chahrour, Ryan. 2013. "Public Communication and Information Acquisition." Unpublished paper, Boston College.

Chari, V. V., and Patrick J. Kehoe. 1999. "Optimal Fiscal and Monetary Policy." Chap. 26 in *Handbook of Macroeconomics*, Vol. 1C, edited by J. B. Taylor and M. Woodford, 1671–1745. Elsevier.

Chari, V. V., and Patrick J. Kehoe. 2006. "Modern Macroeconomics in Practice: How Theory Is Shaping Policy." *Journal of Economic Perspectives* 20(4): 3–28.

Chari, V. V., and Patrick J. Kehoe. 2013. "Bailouts, Time Inconsistency, and Optimal Regulation." FRB Minneapolis Staff Report 481, April.

Chari, V. V., Lawrence J. Christiano, and Patrick J. Kehoe. 1991. "Optimal Fiscal and Monetary Policy: Some Recent Results." *Journal of Money, Credit and Banking* 23(3): 519–39.

Charness, Gary, and Matthias Sutter. 2012. "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives* 26(3): 157–76.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary Policy Shocks: What Have We Learned and to What End?" Chap. 3 in *Handbook of Macroeconomics*, Vol. 1A, edited by John B. Taylor and Michael Woodford. Elsevier.

Clarida, Richard, Jordi Galí, and Mark Gertler. 1999. "The Science of Monetary Policy: A New Keynesian Perspective." *Journal of Economic Literature* 37(4): 1661–1707.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2012. "Innocent Bystanders? Monetary Policy and Inequality in the U.S." NBER Working Paper 18710.

Crowe, Christopher. 2010. "Testing the Transparency Benefits of Inflation Targeting: Evidence from Private Sector Forecasts." *Journal of Monetary Economics* 57(2): 226–32.

Crowe, Christopher, and Ellen E. Meade. 2007. "The Evolution of Central Bank Governance around the World." *Journal of Economic Perspectives* 21(4): 69–90.

Cúrdia, Vasco, and Michael Woodford. 2010. "Credit Spreads and Monetary Policy." *Journal of Money, Credit and Banking* 42(s1): 3–35.

Doepke, Matthias, and Martin Schneider. 2006. "Inflation and the Redistribution of Nominal Wealth." *Journal of Political Economy* 114(6): 1069–97.

Eggertsson, Gauti B., and Michael Woodford. 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity* 34(1): 139–235.

Ehrmann, Michael, and Marcel Fratzscher. 2007. "Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness?" *Journal of Money, Credit and Banking* 39(2–3): 509–541.

Eichengreen, Barry. 1992. "Designing a Central Bank For Europe: A Cautionary Tale from the Early Years of the Federal Reserve System." Chap. 2 in *Establishing a Central Bank: Issues in Europe and Lessons from the U.S.*, edited by Matthew Canzoneri, Vittorio Grilli, and Paul Masson, 13–39. Cambridge University Press.

Eusepi, Stefano, and Bruce Preston. 2010, "Central Bank Communication and Expectations Stabilization." *American Economic Journal: Macroeconomics* 2(3): 235–71.

Farhi, Emmanuel, and Jean Tirole. 2012. "Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts." *American Economic Review* 102(1): 60–93.

Fischer, Stanley. 1981. "Towards an Understanding of the Costs of Inflation: II." *Carnegie-Rochester Conference Series on Public Policy* 15(1): 5–41.

Foote, Christopher, William Block, Keith Crane, and Simon Gray. 2004. "Economic Policy and Prospects in Iraq." *Journal of Economic Perspectives* 18(3): 47–70.

Friedman, Benjamin M. 2008. "What Objectives for Monetary Policy?" Chap. 3 in *Central Banks as Economic Institutions*, edited by Jean-Philippe Touffut. Cheltenham: Edward Elgar.

Friedman, Benjamin M., and Kenneth N. Kuttner. 2010. "Implementation of Monetary Policy: How Do Central Banks Set Interest Rates?" Chap. 24 in *Handbook of Monetary Economics*, Vol. 3B, edited by Friedman, Benjamin M. and Michael Woodford, 1345–1438. Elsevier.

Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58(1): 1–17.

Friedman, Milton. 1970. *The Counter-Revolution in Monetary Theory*, Vol. 33. Institute of Economic Affairs.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960.* National Bureau of Economic Research.

Gaballo, Gaetano. 2013. "Rational Inattention to News: The Perils of Forward Guidance." Banque de France Working Paper 416.

Gertler, Mark, and Nobuhiro Kiyotaki. 2010.

"Financial Intermediation and Credit Policy in Business Cycle Analysis." Chap. 11 in *Handbook of Monetary Economics*, Vol. 3A, edited by Benjamin M. Friedman and Michael Woodford, 547–99. Elsevier.

Giannoni, Marc P., and Michael Woodford. 2010. "Optimal Target Criteria for Stabilization Policy." NBER Working Paper 15757.

Goodfriend, Marvin. 1994. "Why We Need an Accord for Federal Reserve Credit Policy: A Note." *Journal of Money, Credit and Banking* 26(3): 572–80.

Goodfriend, Marvin. 1999. "The Role of a Regional Bank in a System of Central Banks." *Carnegie-Rochester Conference Series on Public Policy* 51(1): 51–71.

Goodfriend, Marvin. 2002. "Interest on Reserves and Monetary Policy." *Economic Policy Review,* Federal Reserve Bank of New York, 8(1): 77–84.

Goodfriend, Marvin. 2007. "How the World Achieved Consensus on Monetary Policy." *Journal of Economic Perspectives* 21(4): 47–68.

Goodfriend, Marvin. 2011. "Central Banking in the Credit Turmoil: An Assessment of Federal Reserve Practice." *Journal of Monetary Economics* 58(1): 1–12.

Greenlaw, David, James D. Hamilton, Peter Hooper, and Frederic S. Mishkin. 2013. "Crunch Time: Fiscal Crises and the Role of Monetary Policy." Unpublished paper, Columbia University and UC San Diego.

Hall, Robert E. 1984. "Monetary Strategy with an Elastic Price Standard." In "Price Stability and Public Policy" Proceedings of a symposium sponsored by the Federal Reserve Bank of Kansas City, held in Jackson Hole, Wyoming, August 2–3, 1984. Available at: http://www.kc.frb.org/publications/research/escp/escp-1984.cfm.

Hall, Robert E. 1986. "Optimal Monetary Institutions and Policy." In *Alternative Monetary Regimes*, edited by Colin Campbell and William Dougan, 224–239. Baltimore: Johns Hopkins University Press.

Hall, Robert E., and Ricardo Reis. 2013. "Maintaining Central-Bank Solvency under New-Style Central Banking." Unpublished paper, February. http://www.columbia.edu/~rr2572/papers/13-HallReis.pdf.

Hammes, David. 2001. "Locating Federal Reserve Districts and Headquarters Cities," *The Region*, Federal Reserve Bank of Mineapolis, September 1. http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=3434.

Hilscher, Jens, Alon Raviv, and Ricardo Reis. In Progress. "Measuring the Market Value of Central Bank Capital." Unpublished paper.

Jimenez, Gabriel, Steven Ongena, Jose-Luis Peydro, and Jesus Saurina. 2012. "Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications." *American Economic Review* 102(5): 2301–26.

Kashyap, Anil K., and Jeremy C. Stein. 2012. "The Optimal Conduct of Monetary Policy with Interest on Reserves." *American Economic Journal: Macroeconomics* 4(1): 266–82.

Kashyap, Anil K, Richard Berner, and Charles A. E. Goodhart. 2011. "The Macroprudential Toolkit." *IMF Economic Review* 59(2): 145–61.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity* 43(2): 215–287.

Kydland, Finn E., and Edward C. Prescott. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85(3): 473–92.

Lagos, Ricardo, and Randall Wright. 2005. "A Unified Framework for Monetary Theory and Policy Analysis." *Journal of Political Economy* 113(3): 463–84.

Lucas, Robert E. 2000. "Inflation and Welfare." *Econometrica* 68(2): 247–74.

Mankiw, N. Gregory, and Ricardo Reis. 2010. "Imperfect Information and Aggregate Supply." Chap. 5 in *Handbook of Monetary Economics*, Vol. 3A, edited by Benjamin Friedman and Michael Woodford, 183–230. Elsevier, North Holland.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual 2003,* vol. 18, pp. 209–248.

Meltzer, Allan H. 2003. *A History of the Federal Reserve,* Vol. 1: *1913–1951.* University of Chicago Press.

Meltzer, Allan H. 2009a. *A History of the Federal Reserve,* Vol. 2, Book 1: *1951–1969.* University of Chicago Press.

Meltzer, Allan H. 2009b. *A History of the Federal Reserve*, Vol. 2, Book 2: *1970–1986.* University of Chicago Press.

Morris, Stephen, and Hyun Song Shin. 2002. "Social Value of Public Information." *American Economic Review* 92(5): 1521–34.

Morris, Stephen, and Hyun Song Shin. 2005. "Central Bank Transparency and the Signal Value of Prices." *Brookings Papers on Economic Activity* 36(2): 1–66.

Morris, Stephen, and Hyun Song Shin. 2007. "Optimal Communication." *Journal of the European Economic Association* 5(2–3): 594–602.

Myatt, David, and Chris Wallace. Forthcoming. "On the Sources and Value of Information: Publicity and Performance in a Lucas–Phelps Island Economy." *Journal of Monetary Economics.*

Quah, Danny, and Shaun P. Vahey. 1995. "Measuring Core Inflation?" *Economic Journal* 105(432): 1130–44.

Reis, Ricardo. 2009. "Interpreting the Unconventional U.S. Monetary Policy of 2007–09." *Brookings Papers on Economic Activity* 40(2): 119–65.

Reis, Ricardo. 2011. "When Should Policymakers Make Announcements?" Unpublished paper, September.

Reis, Ricardo. 2013a. "Exit Strategies and the Federal Reserve." Chap. 7 in *No Way Out: Persistent Government Interventions in the Great Contraction*, edited by Vincent R. Reinhart. American Enterprise Institute Press.

Reis, Ricardo. 2013b. "The Mystique Surrounding the Central Bank's Balance Sheet, Applied to the European Crisis." *American Economics Review* 103(3): 135–40.

Reis, Ricardo, and Mark W. Watson. 2010. "Relative Goods' Prices, Pure Inflation and the Phillips Correlation." *American Economic Journal: Macroeconomics* 2(3): 128–57.

Roca, Mauro. 2010. "Transparency and Monetary Policy with Imperfect Common Knowledge." IMF Working Paper 10-91.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100(4): 1169–89.

Romer, Christina D., and David H. Romer. 1997. "Institutions for Monetary Stability." Chap. 8 in *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. National Bureau of Economic Research.

Romer, Christina D., and David H. Romer. 2004a. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94(4): 1055–84.

Romer, Christina D., and David H. Romer. 2004b. "Choosing the Federal Reserve Chair: Lessons from History." *Journal of Economic Perspectives* 18(1): 129–62.

Romer, Christina D., and David H. Romer. 2008. "The FOMC versus the Staff: Where Can Monetary Policymakers Add Value?" *American Economic Review* 98(2): 230–35.

Romer, Christina D., and David H. Romer. 2013. "The Most Dangerous Idea in Federal Reserve History: Monetary Policy Doesn't Matter." *American Economics Review* 103(3): 55–60.

Sims, Christopher A. 2001. "Fiscal Consequences for Mexico of Adopting the Dollar." *Journal of Money, Credit and Banking* 33(2): 597–616.

Sims, Christopher A. 2013. "Paper Money." *American Economic Review* 103(2): 563–84.

Stern, Gary H., and Ron J. Feldman. 2004. *Too Big to Fail: The Hazards of Bank Bailouts*. Brookings Institution Press.

Stokey, Nancy L. 2003. "'Rules vs. Discretion' After Twenty-Five Years." *NBER Macroeconomics Annual 2002*, vol. 17, pp. 9–64.

Svensson, Lars E. O. 1997. "Optimal Inflation Targets, Conservative Central Banks, and Linear Inflation Contracts." *American Economic Review* 87(1): 98–114.

Svensson, Lars E. O. 1999. "Price-Level Targeting versus Inflation Targeting: A Free Lunch?" *Journal of Money, Credit and Banking* 31(3, Part 1): 277–95.

Svensson, Lars E. O. 2003. "What Is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules." *Journal of Economic Literature* 41(2): 426–477.

Svensson, Lars E. O. 2006. "Social Value of Public Information: Comment: Morris and Shin (2002) Is Actually Pro-Transparency, Not Con." *American Economic Review* 96(1): 448–52.

Svensson, Lars E. O. 2010. "Inflation Targeting." Chap. 22 in *Handbook of Monetary Economics*, Vol. 3B, edited by Benjamin M. Friedman and Michael Woodford, 1237–1302. Elsevier.

Svensson, Lars E. O. 2013. "The Possible Unemployment Cost of Average Inflation below a Credible Target." Unpublished paper, Stockholm University. http://larseosvensson.se/files/papers/Phillipsa.pdf.

Vandenbussche, Jérôme. 2006."Elements of Optimal Monetary Policy Committee Design." IMF Working Paper No. 06/277, December.

von Hagen, Jürgen. 1997. "Monetary Policy and Institutions in the EMU." *Swedish Economic Policy Review* 4(1): 51–116.

Wallace, Neil. 1981. "A Modigliani–Miller Theorem for Open-Market Operations." *American Economic Review* 71(3): 267–74.

Waller, Christopher J. 1989. "Monetary Policy Games and Central Bank Politics." *Journal of Money, Credit and Banking* 21(4): 422–31.

Waller, Christopher J. 2000. "Policy Boards and Policy Smoothing." *Quarterly Journal of Economics* 115(1): 305–39.

Walsh, Carl E. 1995. "Optimal Contracts for Central Bankers." *American Economic Review* 85(1): 150–67.

Williamson, Stephen, and Randall Wright. 2010. "New Monetarist Economics: Models." Chap. 2 in *Handbook of Monetary Economics*, Vol. 3A, edited by Benjamin Friedman and Michael Woodford, 25–96. Elsevier.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

Woodford, Michael. 2005. "Central Bank

Communication and Policy Effectiveness." In *The Greenspan Era: Lessons for the Future,* 399–474. Proceedings of a Conference at the Federal Reserve Bank of Kansas City, August 25–27. Available at: http://www.kc.frb.org/publications/research/escp/escp-2005.cfm.

Woodford, Michael. 2007. "The Case for Forecast Targeting as a Monetary Policy Strategy." *Journal of Economic Perspectives* 21(4): 3–24.

Woodford, Michael. 2010. "Optimal Monetary Stabilization Policy." Chap. 14 in *Handbook of Monetary Economics*, Vol. 3, edited by Benjamin M. Friedman and Michael Woodford, pp. 723–828. Elsevier.

Woodford, Michael. 2012. "Inflation Targeting and Financial Stability." NBER Working Paper 17967.

Zarnowitz, Victor, and Louis A Lambros. 1987. "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy* 95(3): 591–621.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 45–64

The Federal Reserve and Panic Prevention: The Roles of Financial Regulation and Lender of Last Resort[†]

Gary Gorton and Andrew Metrick

Any market economy is susceptible to a fundamental mismatch that can lead to the negative externalities of liquidity demand, which include credit cycles, bank runs, and financial crises. Assets with liquidity are "safe" assets. More specifically, "liquidity" refers to the ease with which an asset can be sold quickly and without a loss of value, in the sense that substantial sales do not depress the price of this asset nor give rise to an adverse selection problem in which buyers fear that the asset being sold is of diminished quality. However, liquidity is hard to produce. Long-term investment is required for growth, but such investment is by its nature uncertain and costly to evaluate. On the other side, the ultimate suppliers of investment capital are subject to liquidity shocks: in particular, at times they will perceive higher risks and desire greater liquidity, which means holding short-term and very low-risk financial assets that can easily be sold, like US Treasury bills. In normal times, the maturity and information mismatch between the long-term investments and short-term liquidity needs are intermediated by the financial system through the creation of liquid "money-like" assets. In a simple example, a bank uses bank deposits to make long-term loans, while promising that the deposits will be available in the short run. However, a wide array of other short-term financial instruments are

■ *Gary Gorton is the Frederick Frank Class of 1954 Professor of Finance and Andrew Metrick is the Deputy Dean & Michael H. Jordan Professor of Finance and Management, both at the Yale School of Management, New Haven, Connecticut. Both authors are also Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are gary.gorton@yale.edu and andrew.metrick@yale.edu.*

[†] To access the disclosure statements, visit
http://dx.doi.org/10.1257/jep.27.4.45 doi=10.1257/jep.27.4.45

also backed by long-term assets, while allowing investors who desire liquidity to withdraw their funds, or more generally not renew their short-term investment, in a much shorter time horizon. During a financial crisis, the negative externalities of liquidity demand are manifested when investors race to withdraw their liquid assets; in "normal times," negative externalities occur when each additional liquid claim does not incorporate in its price its contribution to the risk of such a crisis.

To mitigate the risk of a liquidity-driven crisis, the United States has a financial sector safety net with two key pillars: the Federal Reserve as a lender-of-last-resort and the Federal Deposit Insurance Corporation (FDIC) as a guarantor of bank deposits. The existence of this safety net then alters the incentives of regulated financial institutions: in particular, they can take greater risks when their depositors and investors know that this safety net is in place. Thus, the existence of the safety net provides the rationale for close supervision and regulations that limit the scope, risk-taking, and leverage of these institutions. If the safety net is too large, then banks lack incentives to manage risks in a socially optimal way; if the safety net is too small, then failure of a large institution could have major spillovers to the whole financial system; and if only the largest institutions are thus given the most protection, then the private incentives will be for every institution to grow "too big to fail." This dynamic presents a complex problem for the Fed as the lender of last resort and regulator of the largest institutions.

This paper traces the Fed's attempts to address this problem from its founding. We will discuss how the effectiveness of the lender-of-last-resort function was eroded in the 1920s, which in turn contributed to the banking panics of the Great Depression and indeed has hampered its lender-of-last-resort efforts to the present day. We consider the regulatory changes of the New Deal, including deposit insurance and the centralization of Fed decision-making power in the Board of Governors, which by some combination of luck and design contributed to a quiet period of nearly 50 years in the US financial system. Indeed, during this time bank supervision was only peripheral to the Fed's priorities, which moved steadily towards a focus on price stability using interest-rate policy as its main instrument, and the Fed rarely needed to even think about the lender-of-last-resort function. The late 1970s saw the beginning of a transformation of the banking sector, with a rise of nonbank financial intermediaries and then regulatory adjustments so that banks could compete with these nonbank firms, which has continued to the present day. The financial crisis of 2007–2009 shook bank supervision efforts out of their slumber, made the lender-of-last-resort function central again, and led to a significant shift for the Fed back to its financial-stability roots. Indeed, the Fed's efforts in the recent financial crisis can largely be viewed as attempts to expand the lender-of-last-resort function beyond its traditional institutions and markets. We conclude by bringing the story to the present day with a discussion of the evolving role of the Federal Reserve in the context of the changes under the Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010.

The Establishment of the Federal Reserve System

Market economies will sometimes face a banking panic or financial crisis, which can be defined as an event in which the holders of short-term debt issued by intermediaries seek to withdraw cash en masse or refuse to renew their loans. A crisis is a *systemic* event; it involves the banking system, not this or that bank. Such a crisis is an information event (in the sense of Dang, Gorton, and Holmström 2013). That is, many holders of short-term debt previously viewed it as so safe that it was unnecessary to gather or process information about the debt, but these debt-holders then come to fear that the debt is not so safe and that they cannot distinguish good and bad collateral. In such a crisis situation, when what had seemed safe is no longer viewed that way, all banks are insolvent in the sense that they cannot honor their debt contracts without trying to sell assets—and if they try to sell assets, they will receive only a low fire-sale price, because the value of collateral has become so uncertain.

In a bank run, holders of banks' short-term debt come to doubt the collateral backing the debt. These doubts are not irrational. When no central bank is present, banking panics occur around the peak of the business cycle when holders of short-term bank debt receive news that indicates a recession is likely coming (Gorton 1988).[1] An unexpected deterioration in macroeconomic fundamentals causes a shift in expectations. In the ensuing recession, some banks will fail, but it is not known which banks. Depositors respond by withdrawing their cash from all banks. In the United States, such panics were common in the century before the start of the Federal Reserve.

The underlying dilemma in a bank run is that the depositors' doubts about the backing collateral can only be removed by showing them cash. But since the banks have lent the cash out, and the assets of the banking system cannot be sold (except possibly at low "fire sale" prices), there is no way for the banks to obtain cash except through a lender of last resort—an institution that lends against the impaired bank collateral. However, a lender of last resort can only prevent panics if it is sufficiently credible such that depositors believe it can essentially purchase the assets of the banking system. The Federal Reserve System was established for exactly this purpose.

At the time the Fed was established, the main perceived defect of the banking system was that currency was not "elastic"—that is, there was no way to obtain more currency to meet demands from depositors in times of bank runs, nor to meet seasonal demands. At that time, the main mechanism for responding to panic was the private bank clearing houses. Since being established in New York City in 1854, clearing houses had spread across the country and had become increasingly sophisticated in their responses to crises (for discussion of clearing houses, see Timberlake

[1] In the modern era with the presence of central banks, the links between financial crises and recessions are similar. For example, Demirgüç-Kunt and Detragiache (1998, p. 83) examine the period 1980–1994 and "find that low GDP growth, excessively high real interest rates, and high inflation significantly increase the likelihood of systemic problems in our sample."

1984; Gorton 1984, 1985; Gorton and Mullineaux 1987; Gorton and Huang 2006). Clearing houses, with one in each large city, were coalitions of member banks. Ostensibly set up to efficiently clear checks, they assumed a central bank–like role in crises, even though they were private associations.

A panic would trigger clearing house members to act as one large bank, issuing special liabilities—clearing house loan certificates—for which they were jointly responsible. At the outset of the crisis, the clearing house would prohibit the publication of bank-specific information, which was required during noncrisis times. Also, the amounts of clearing house loan certificates issued to individual member banks were kept secret, preventing those banks from being targeted for bank runs. Following the Panic of 1907, Congress passed the Aldrich–Vreeland Act, which among other provisions created a system for national banks to issue emergency "elastic" currency in a panic.

However, these responses of the clearing house member banks were only triggered by the panic itself. The ability of the clearing houses to issue loan certificates and Aldrich–Vreeland emergency currency did not prevent panics and their associated real effects. William Ridgely (1908, p. 173), the US Comptroller of the Currency from 1901 to 1908, put the issue this way: "The real need is for something that will prevent panics, not for something that will relieve them; and the only way to attain this is through the agency of a Governmental bank."

Thus, the idea behind the establishment of the Federal Reserve System was that it could do something that the clearing houses and the Aldrich–Vreeland Act could not do. It could establish a credible emergency mechanism *in advance*. When the Federal Reserve System was founded, the main focus was on the potential benefits of a "bills market"—that is, a market for bankers' acceptances, which are a documented promise by a bank to make a payment at a future time. The Federal Reserve would participate in this market by purchasing bankers' acceptances. In addition, banks would be able to use their holdings of commercial paper and other marketable securities as collateral to borrow at the discount window—thus in effect exchanging private debt for currency.

Moreover, being a (quasi-)government entity, the Federal Reserve System could be expected to be solvent and would always be able to lend to banks. By contrast, the coalitions of clearing house banks might not be solvent, so expectations that the clearing house would act did not fully deter panics. Indeed, currency premia on the certified checks, which were joint clearing house liabilities, were positive during crisis periods (in other words, it took more than $1 of certified checks to buy $1 of currency), reflecting uncertainty about clearing house solvency. The Aldrich–Vreeland emergency currency was issued with bank loans as collateral, not US Treasury bonds. Again, there was uncertainty about the outcomes.

There is an important difference between providing the reassurance that can prevent bank runs and responding to a crisis once it has happened. Once a financial event is seen to be systemic and the lender of last resort begins lending, these actions take time and the process of exchanging private bank assets for government assets (whether money or Treasury debt) can be costly and painful.

It was widely believed that the discounting authority of the Federal Reserve would *prevent* banking panics. Banks needing cash could take bankers' acceptances (that is, their promise to pay at a near-term date) which were discounted from par to the Fed's discount window, where the Fed would buy it at a further discount—"rediscounting" it. Representative Carter Glass (1927, p. 387), who sponsored the Federal Reserve Act in the House of Representatives, wrote that the most important accomplishments of the legislation were to remove "seasonals" in interest rates and to prevent panics. Senator Robert Owen (1919, p. 99), sponsor of the bill in the Senate, said that the Federal Reserve Act "gives assurance to the business men of the country that they never need fear a currency famine. It assures them absolutely against the danger of financial panic . . ." Congressman Michael Phelan of Massachusetts, Chairman of the House Committee on Banking and Currency, argued (as quoted in Hackley 1973, p. 10): "In times of stress, when a bank needs cash, it can obtain it by a simple process of rediscounting paper with the Federal reserve [sic] banks. Many a bank will thus be enabled to get relief in time of serious need." Businessmen and regulators agreed. Magnus Alexander, the president of the National Industrial Conference Board announced (quoted in Angly 1931, p. 12) that "there is no reason why there should be any more panics." The Comptroller of the Currency (1915, p. 10) announced that, with the new Federal Reserve Act, "financial and commercial crises, or 'panics,' . . . with their attendant misfortunes and prostrations, seem to be mathematically impossible." The Federal Reserve System's (1914, p. 17) first Annual Report states that "its duty is not to await emergencies but by anticipation to do what it can to prevent them."

The 1920s

The establishment of the Federal Reserve System did change the expectations of depositors about systemic banking crises.[2] Gorton (1988) creates a leading indicator of recessions for the earlier US "National Banking Era" from the Civil War up to 1913, and finds that panics arose when the unexpected component of this leading indicator of recession exceeded a threshold. During the National Banking Era, no panic occurred without this threshold being exceeded, and there are no cases where it was exceeded without a panic. This model predicts that there should have been a panic in June 1920 (and another panic in December 1929). Thus, the 1920–21 recession can be viewed as the first test of the ability of the Federal Reserve to prevent bank runs.

As dated by the National Bureau of Economic Research, there was a business cycle peak in January 1920 and a trough in July 1921. Banks started to fail in 1920;

[2] There is some evidence that seasonal swings in short-term interest rates were eliminated, although the point is controversial. For a sampling of the evidence that the Fed did eliminate seasonal swings, see Miron (1986) and Mankiw, Miron, and Weil (1987). For the alternative view, see Shiller (1980), Clark (1986), Fishe and Wohar (1990), and Fishe (1991).

505 banks failed in 1921, and the number of failures continued to rise, averaging 680 per year from 1923 to 1929. The peak was 950 in 1926 (Alston, Grove, and Wheelock 1994). Hamilton (1985, p. 585) observes that the failed banks were overwhelmingly small banks in small rural communities: "National banks were only 13 percent of the failures and only 17 percent were members of the Federal Reserve System." In other words, for the most part the banks that failed did not have access to the Federal Reserve discount window.

Though many small banks failed, there was no panic. As many contemporary commentators noted, depositors did not run on banks. For example, Henry Parker Willis (1923, p. 1406, emphasis added), who received a PhD in economics from the University of Chicago and was later the first Secretary of the Federal Reserve System, wrote:

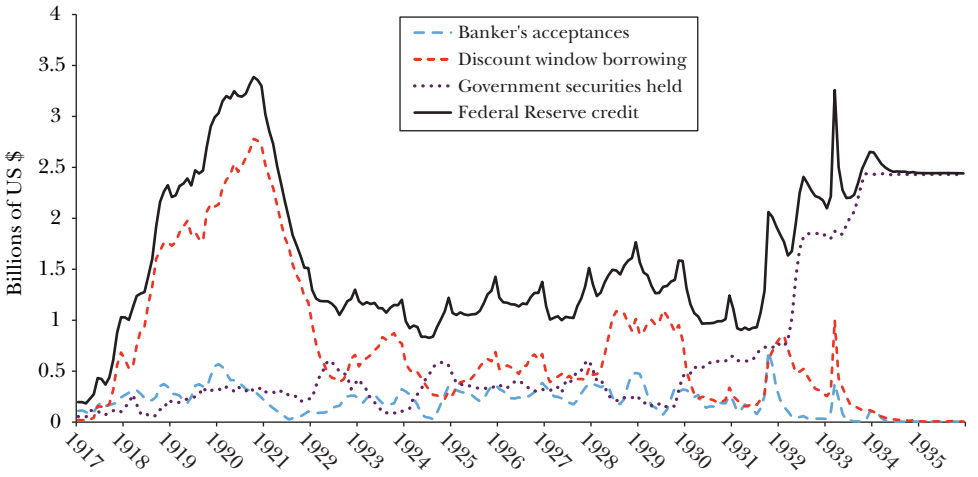
> In previous panics or periods of stringency, difficulty had grown out of the fact that doubts arose concerning the ability of given institutions to meet their obligations, owing to the fact that their loans were frozen or that public confidence had resulted in withdrawing an undue amount of cash from them. On such occasions relief was obtained by the banks banding together for the purpose of supporting any of their number which had sound assets. In the depression of 1920–1921, the federal reserve system [sic] was in the position of a clearing house association, *already organized in advance* and able to assist the community . . .

Perhaps predictably, the Federal Reserve Annual Report (1921, p. 99) took a similar view that the creation of the Federal Reserve had prevented a panic:

> Other nations, such as Great Britain and France, with their great central banking institutions, have always had their years of prosperity and their periods of depression, although they have been free from the money panics which we formerly had in this country as a result of our inadequate banking system and which we would, no doubt, have had in the most aggravated degree a year or so ago but for the efficiency and stabilizing influence of the Federal Reserve System.

If bank depositors did not run because they expected banks to have access to the discount window, then it might not be necessary for banks to have actually borrowed from the discount window. But in fact, national banks did use the discount window, as shown in Figure 1. Tallman (2010, p. 104) also notes this use of the discount window over the years 1914–27. In 1921, discounts and advances as a proportion of Federal Reserve credit was at its peak of 82 percent with about 60 percent of member banks borrowing. "It was not uncommon, evidently, for hundreds of banks to be continuously borrowing amounts in excess of their capital and surplus" (Shull 1971, p. 37). Notably, there was no evidence that borrowers from the discount window experienced any particular stigma in credit markets.

Figure 1
Federal Reserve Credit Extended, 1917–1935



Source: Tallman (2010); used with permission.

One reason that banks borrowed so much from the discount window was that the discount rate was below the market interest rate. During World War I, the Fed felt that low discount rates were important. "The Board did not believe, during the war period, that marked advances in rates would be advisable in view of the obvious necessity of avoiding any policy likely to disturb the financial operations of the Treasury" (Harding 1925, p. 147). During the steep 1920–21 recession, the low discount rate may have been fortuitous. As an Assistant Secretary of the Treasury wrote (Leffingwell 1921, p. 35), "by permitting rates to remain below the open market rates and credit to be expanded during the period of deflation of prices, it has prevented the present business depression from degenerating into an old-fashioned panic." But over time, of course, freely available discount lending at below-market interest rates was bound to bring tensions.

Indeed, unbeknownst to the wider world, Fed policy on discount window lending was fundamentally altered in the mid-1920s. As Shull (1993, p. 20, with quotations from Keynes, 1930, pp. 239–40) explains: "A set of non-price rationing rules, limiting use of the discount window to short-term borrowing for unanticipated outflows of funds, were developed; banks were encouraged to be 'reluctant to borrow;' i.e., the Fed "turned to 'gadgets' and conventions . . . without any overt alteration of the law." Creating a reluctance to borrow can informally come about through possible implicit threats to examine the borrowing bank more frequently and intensively, ostensibly to determine whether such borrowing is warranted.

Why was the policy on discount lending changed? There seem to be several reasons. First, it became clear that hundreds of banks were borrowing from the Fed for extended periods of time. Shull (1971, p. 35) reports that as of August 31, 1925,

588 banks had been borrowing continuously for at least a year; 239 had been borrowing since the start of the recession in 1920; and 122 had been borrowing continuously since before 1920. In addition, "259 national member banks had failed since 1920, and a guess was made that at least 80 per cent had been habitual borrowers prior to their failure." Thus, the Federal Reserve Annual Report of 1926 (p. 4) stated that "the funds of the Federal Reserve banks are ordinarily intended to be used in meeting temporary requirements of members, and continuous borrowing by a member bank as a general practice would not be consistent with the intent of the Federal Reserve Act."

In addition, by the latter part of the 1920s, the Fed became concerned with trying to distinguish between "speculative security loans" and loans for "legitimate business." In other words, was discount window credit being used to pump up stock market values (Anderson 1966)? Was it leading to high growth in real estate prices, labeled a "bubble" by some (White 2009)? The Fed sought to restrain credit growth through moral suasion that would deter member banks from borrowing for speculative purposes, while at the same time trying to maintain a preferential discount rate for "legitimate" borrowing (Friedman and Schwartz 1963, p. 225–26). But the Fed decided that attempting to influence the economy via the discount window was not going to work. In short, the purpose of the discount window changed. It would no longer serve to provide an "elastic currency." While contemporary observers noted that there had been no banking panics in the 1920s, there appears to have been no understanding of the details of how freely available lending through the discount window had avoided the panic. The Fed's new policy of creating a "reluctance to borrow" based on nonpecuniary measures, and an emphasis that such lending should be only temporary, meant that a bank that did borrow from the discount window must be in trouble. This was the creation of "stigma," which has complicated lender-of-last-resort policy ever since.

The Great Depression

Explaining the timing and causes of the banking panics of the Great Depression has been difficult and many researchers have offered explanations.[3] There is a reason that researchers have found this confusing: at the time, bank depositors were also confused. They had been told repeatedly that banking panics would not occur under the Federal Reserve System—and in fact, no panics had occurred in the 1920s. Depositors, however, were unaware of the shift in Fed policy with regard to the discount window, so depositors reasonably assumed that banks would again avail themselves of the discount window as needed. But by the late 1920s, banks had been repeatedly told not to use the discount window, and when the 1930s arrived,

[3] This literature is very large and we do not survey it here. As a starting point, see Friedman and Schwartz (1963), Wicker (1996), and Meltzer (2003). Richardson (2007) relates this literature to new archival data on bank failures and suspensions (which are not the same thing).

they were quite hesitant to do so. As shown in Figure 1, discount window borrowing from 1929 to 1931 was much lower than in the 1920s, and after peaking in 1932, it declines slightly. Apparently, banks feared the stigma the Fed policies had created in the mid-1920s on discount lending.

When the Great Depression started in 1929, there were no bank runs. As mentioned earlier, Gorton's (1988) calculations looking at how unexpected movements in leading indicators had predicted financial crises in the pre-Fed era suggested that, in the Great Depression, there should have been bank runs starting in December 1929. Similarly, Wicker (1980, p. 573) noted: "Historically, banking panics in the United States usually developed shortly after a downturn in economic activity. The banking crisis in November–December 1930, however, was unlike previous banking collapses: there was little or no discernible impact on the central money market, and the panic lagged the downturn by eighteen months."

Bank runs did not happen in the Great Depression until late in 1930. As Richardson (2007, p. 40) notes: "Before October 1930, the pattern of [bank] failures resembled the pattern that prevailed during the 1920s. Small, rural banks with large loan losses failed at a steady rate. In November 1930, the collapse of correspondent networks triggered banking panics. Runs rose in number and severity after prominent financial conglomerates in New York and Los Angeles closed amid scandals covered prominently in the national press." There is some dispute over which bank collapse loomed largest. Friedman and Schwartz (1963) argue that the failure of the Bank of United States on December 11, 1930, was especially important—in part because of the bank's name. Wicker (1980, p. 581; 1996) disputes the importance of that bank failure, and instead cites the collapse of Caldwell and Company in mid-November as the trigger of the panic. Caldwell was large; it controlled a large chain of banks in the South.

A second wave of bank runs began in March 1931. There were runs, for example, on Chicago-area banks that were followed by a 40 percent increase in postal savings deposits (Wicker 1996, p. 85; for additional discussion, see Calomiris and Mason 1997). Finally, there was the Panic of 1933, actually in the last quarter of 1932 and early 1933, which led to President Roosevelt declaring a four-day "bank holiday" in March 1933, during which banks and the stock exchange were closed and forbidden to do any business without special government permission.

During this time, although the Federal Reserve was not engaging in much discount lending, the Reconstruction Finance Corporation, established in January 1932 under President Hoover, had started lending to banks in February 1932. The Reconstruction Finance Corporation action was needed because the Fed took no "positive action to intervene directly to keep open troubled banks. No direct assistance was offered other than to discount eligible paper of the [Federal Reserve] member banks" (Wicker 1996, p. 85). There were 17,000 banks in existence just prior to Roosevelt's March 1933 banking holiday. Only 12,000 survived, and half of those were borrowing some or as much as all of their capital from the Reconstruction Finance Corporation (Todd 1992). Ironically, the chairman of the Reconstruction Finance Corporation was Eugene Mayer, who was also chairman of the Fed.

At first, there was apparently no stigma attached to borrowing from the Reconstruction Finance Corporation until the clerk of the House of Representatives revealed the names of borrowers in July 1932 (Butkiewicz 1995, 1999; see also Friedman and Schwartz 1963, p. 331). Figure 2 illustrates the scale of loans from the Reconstruction Finance Corporation to banks as well as to other institutions like state and local governments, railroads, and mortgage institutions. Prior to the revelation of borrower names beginning in July 1932, total Reconstruction Finance Corporation borrowing had reached approximately $1 billion, with about half of this total going to banks. Following the name revelation, net bank borrowing flattened out and was below $500 million four years later, even though nonbank borrowing—where stigma is far less of an issue—rose to more than $2 billion of the total.

The bank runs of the Great Depression were haphazard, chaotic, and spread out in time, unlike those of the pre-Fed period. Given that there was no bank run in 1929 at the onset of the Depression, the timing suggests that when depositors eventually saw the failures of large banks in the 1930s, they realized that the discount window mechanism was not working and the bank runs started. What happened? Friedman and Schwartz (1963, pp. 318–19) write: "The aversion to borrowing by banks, which the Reserve System had tried to strengthen during the twenties, was still greater at a time when depositors were fearful for the safety of every bank and were scrutinizing balance sheets with great care to see which banks were likely to be the next to go . . ." Wheelock (1990, p. 424) provides some evidence for this:

> This study also finds evidence of a downward shift in borrowed reserve demand during the Depression. Financial crises made banks cautious and less willing to borrow reserves. The Fed's failure to recognize this change in bank willingness to borrow contributed to its failure to interpret monetary conditions accurately. Fed officials continued to believe that low levels of bank borrowing signaled easy money.

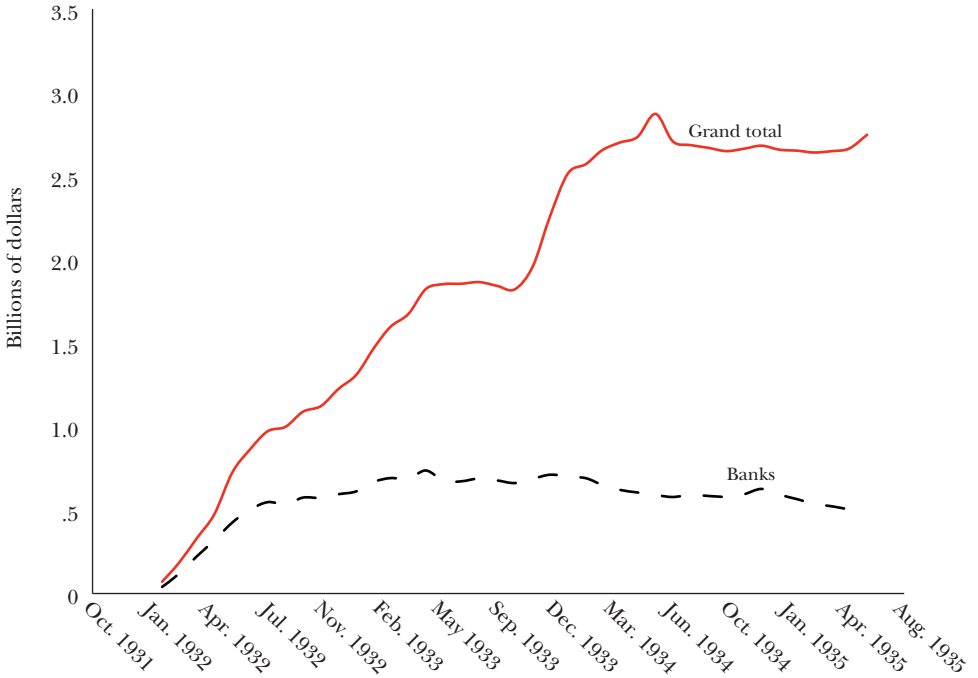
The problem was that the expectations of depositors that banks could and would avail themselves of the discount window when in trouble were not (widely) realized. Large banks failed and depositors then ran on the banks.[4]

New Deal Legislation and the Quiet Period: 1933–1978

The financial legislation of the New Deal period transformed the financial regulatory system and the role of the Federal Reserve within it; in addition, it

[4] We are not making any claims here about the effectiveness of the Fed as a lender of last resort when banks *actually did borrow*. For example, Richardson and Troost (2009) contrast the policies of two regional Federal Reserve Banks (St. Louis and Atlanta) with regard to their responses to bank troubles in Mississippi during the Great Depression. Atlanta aggressively assisted banks and the bank failure rate was lower than in the part of Mississippi in the St. Louis district. The interesting question here is how Atlanta managed to overcome (or avoid) the stigma that depressed borrowing in other districts.

Figure 2
Reconstruction Finance Corporation Loans Outstanding



Source: Federal Reserve, Flow of Funds.
Note: Figure 2 illustrates the scale of loans from the Reconstruction Finance Corporation to banks as well as to other institutions like state and local governments, railroads, and mortgage institutions.

represented the last major set of changes in financial regulation until the 1970s.[5] The Banking Acts of 1933 and 1935 amended the Federal Reserve Act to establish the Federal Deposit Insurance Corporation. The advent of deposit insurance rendered moot—for a time—the mistake of developing the policy of "reluctance to borrow," and there was no discussion or realization of the problem that had been created by the discount–rate policies of the 1920s. Over the subsequent 75 years, the original insurance cap of $2,500 per bank account would be raised many times, finally reaching $250,000 in the aftermath of the recent financial crisis.

The Banking Acts also had a profound influence on the power and structure of the Fed. The balance of power between the Board and the regional Reserve Banks was tipped in favor of the center, with a Board-dominated Federal Open Market Committee established in 1935. Far more obscure at the time was a small

[5] We do not attempt anything close to a review of all financial regulation during this time period. For a comprehensive treatment of regulatory and competitive changes in the key 1979–1994 period, see Berger, Kashyap, and Scalise (1995). For a discussion of changes since the 1990s leading to the rising share of nonbank financial intermediaries, see Gorton and Metrick (2010).

amendment to Section 13 of the Federal Reserve Act, granting the Fed the power to greatly expand its lending programs under "unusual and exigent circumstances." These powers were invoked often in the recent crisis, as discussed later in this paper.[6]

The Banking Act of 1933 is often known by the last names of its sponsors, Glass and Steagall, and by the provision of the law that enforced the separation of deposit-taking and securities underwriting. This separation of banking and securities was coincident with significant new financial regulation, beginning with the Securities Act of 1933, which focused on the primary sale of securities, and the Exchange Act of 1934, which created the Securities and Exchange Commission and focused on the secondary trading markets. The SEC was granted further powers to regulate market intermediaries in the Investment Company Act of 1940 (for mutual funds and other investment companies) and in the Investment Adviser Act of 1940 (which today covers hedge funds and private equity funds, in addition to traditional advisers).

After the New Deal legislation, the most important piece of financial regulation to affect the Fed during this time period was the Bank Holding Company Act of 1956, in which the Fed was given oversight responsibility over holding companies that included commercial banks in their structure, with rules codified about the separation of banking and nonbanking activities. Importantly, this responsibility gave the Fed insight and access to the largest commercial banks, all of which (over time) became part of bank holding companies. The role of bank holding companies in the overall financial system has increased steadily, so that today they cover the vast majority of assets in the US banking system.

The Transformation of Banking: 1979–2006

Into the 1970s, banking in the United States was still a relatively simple business, at least compared with today, with this simplicity supported by ceilings on the interest rates that could be paid on time deposits ("Regulation Q"), a prohibition of paying interest on demand deposits, and by restrictions on both inter- and intrastate branching of banks. The story of banking since the 1970s is largely about attempts to work around regulations and the resulting growth in nonbank alternatives in the far more complex financial system of today. Liquid safe assets—assets that can safely store value for a short period of time with almost no risk such as money market mutual funds, and sale and repurchase agreements—began to be produced in large volumes. In Gorton, Lewellen, and Metrick (2012), we show that the net effect of these changes is that bank deposits' share of the "safe" financial assets in the United States fell from 80 percent in 1952 to less than 30 percent by 2007.

[6] The Fed's emergency-lending power in Section 13(3) was first granted by the Emergency Relief and Construction Act of 1932, which later received amendments in the Banking Act of 1935 and in Federal Deposit Insurance Corporation Improvement Act of 1991. As discussed later, these amendments proved crucial for the lending powers used in the recent crisis (Mehra 2011).

One controversial element of bank regulation and supervision has played a large role in recent Federal Reserve history: the setting of capital standards. "Capital" in this context is defined in its narrowest sense as the common-equity component on the right-hand-side of the balance sheet, with various broader definitions including other forms of equity and the present-value of different kinds of safe revenue claims. For our purposes here, we will just refer to all of these definitions as "capital," unless there is an important reason to be more specific.[7]

The benefit of banks having higher capital should be that each individual bank has a lower probability of distress. Given the access to the government safety net of deposit insurance and the lender of last resort, banks may not fully internalize the social cost of failure. In addition, even in the absence of such access, banks would not internalize the spillover effects of their own failure on other financial institutions. For these reasons, the government has an interest in lowering the probability of bank failure by requiring higher levels of capital than may seem privately optimal to banks.

An international consortium of regulators began work on a set of standards that could be applied across the major economies; this process culminated in the "Basel I" accords of 1988, implemented in 1990 in the United States. In the 1990s, the Basel requirements were revised and updated for riskiness of bank assets, resulting ultimately in the Basel II accords of 1997. As of 2006, most of the developed world had fairly complex implementations of capital standards for banks, with the Fed as the primary regulator for the largest financial holding companies in the United States. Nevertheless, regulatory capital proved to be a slow-moving measure of bank health, and in no country did it provide clear warnings of the coming crisis. In the aftermath of the crisis, the Fed was a main driver of the next round of "Basel III" accords, although the Basel III standards have not yet been implemented in the United States. Whether raising bank capital requirements is desirable has been the subject of great debate.[8]

[7] Our discussion of capital rules and the Basel process focuses on the role played by the Federal Reserve and the implications for the growth of the shadow banking system. For a more comprehensive treatment, Goodhart (2011) is a definitive history of the Basel process up through 1997, and Hanson, Kashyap, and Stein (2011) is an accessible survey of the intellectual debate about capital standards in the post-crisis world.

[8] DeAngelo and Stulz (2013) point out that if banks' liabilities, short-term liquid debt, are useful because of their liquidity, they have a "convenience yield" (part of the return the holder gets is the benefits of liquidity) and then banks optimally have high leverage. Kashyap, Stein, and Hanson (2010) point out that even small increases in the cost of the capital could be sufficient to drive significant flows from banks into nonbank financial institutions. For the most forceful argument in favor of the Modigliani–Miller interpretation that raising additional capital would not be costly for banks, see Admati and Hellwig (2013). Other recent perspectives on this debate include Baker and Wurgler (2013) and Gorton and Winton (2002).

The Financial Crisis of 2007–2009

When the financial crisis began in 2007, the Federal Reserve faced two major challenges in its function as lender of last resort. First, the stigma of the discount window, originally created by the policies of the 1920s, was still causing a reluctance to borrow by member banks. Second, the sharp growth of a financial sector outside of member banks—in the so-called "shadow banking" sector where institutions like money market mutual funds take deposits and funds are invested in bonds and other financial assets—left a large portion of the financial system without access to the discount window. Most of the Fed's actions during the crisis can be viewed as attempts to deal with these challenges.

Policies both formal (raising the discount rate) and informal (implicit threats to conduct more extensive and frequent bank examinations) continued to discourage borrowing from the discount window from the 1920s through the rest of the twentieth century. Despite an additional change in August 2007 that decreased the discount-window premium by 50 basis points and increased the eligible term for discount window loans, banks were still reluctant to borrow throughout 2007. In an interesting parallel to the role of the Reconstruction Finance Corporation during the Great Depression, many banks found an alternative source of back-up liquidity to escape the stigma of the discount window—in this case the Federal Home Loan Banks. Ashcraft, Beck, and Frame (2010) describe how the FHLB system became a "lender of next-to-last resort" with over $1 trillion in loans at the peak of the crisis.

In December 2007, the Fed created the Term Auction Facility in a major attempt to overcome the reluctance of banks to borrow at the discount window. In the Term Auction Facility (TAF), the Fed created regular auctions of pre-set total quantities of loans for set terms (either 28 or 84 days), and the same institutions eligible to use the discount window were able to submit bids for what they would pay to borrow these funds. The rules for these loans were similar (although not identical) to those for the discount window. The institutions that received the loans were not publicly revealed, and the market apparently believed that some combination of the stigma and risk of possible disclosure of these loans was significantly lower than those from the discount window. According to Almantier, Ghysels, Sarkar, and Shrader (2011), TAF credit outstanding peaked at over $300 billion, nearly three times the peak for discount window credit. This occurred although interest rates for borrowing through the Term Auction Facility were higher on average than rates at the discount window, by an average of 37 basis points overall and more than 150 basis points after the Lehman bankruptcy in September 2008. Banks were apparently willing to pay a premium to avoid the stigma of borrowing at the discount window.

Continued pressure in short-term funding markets led to the near-bankruptcy and fire sale of Bear Stearns to JPMorgan in March 2008. As Bear Stearns was not a depository institution and thus did not have access to the discount window, the eventual Fed guarantee that enabled the JPMorgan sale required use of the 13(3) authority granted in the 1930s, its first invocation during the crisis. The

Fed also responded by expanding the discount window—historically reserved for depository institutions—to include a broader group of primary dealers including Lehman Brothers, Merrill Lynch, and Goldman Sachs. A further expansion of the lender-of-last-resort function, which also required the use of Section 13(3), authority came through the Term-Securities Lending Facility (TSLF), which allowed primary dealers to effectively exchange illiquid securities for government bonds. The Term Securities Lending Facility was successful in reducing stress in the sale and repurchase or "repo" markets (Fleming et al. 2010; Hrung and Seligman 2011), in which one firm sells securities to another firm and then agrees to repurchase them at a slightly higher price in the near future—thus in effect receiving a short-term loan (Gorton and Metrick 2012).

However, the Federal Reserve was only getting started in expanding its role as lender of last resort for other parts of the shadow banking system. In fall 2008, 13(3) authority was used to create an alphabet soup of facilities, each targeted to extend the lender-of-last-resort function to another part of the shadow banking system. The Term Asset-Backed Securities Loan Facility (TALF) allowed borrowers to post various asset-backed securities as collateral for term loans; the Commercial Paper Funding Facility (CPFF) created facilities to buy commercial paper directly from issuers; the Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility (AMLF) purchased asset-backed commercial paper from money-market mutual funds; and the Money-Market Investor Funding Facility (MMIFF) lent to money-market mutual funds on a broad range of collateral, effectively acting as a discount window for these funds. The Fed also used its 13(3) authority to create special-purpose vehicles to support lending programs to the insurance company AIG in September 2008.

Overall, the Fed made significant use of its 13(3) powers during the crisis, expanding its role as a lender of last resort well beyond the depository institutions typically served by the discount window. To go with the lender-of-last-resort function, the Fed marked the end of the panic phase of the financial crisis with the Supervisory Capital Assessment Program (SCAP), also known as the "stress tests," carried out in spring 2009. The stress tests expanded the standard supervisory reviews to include considering the stresses that might arise in specific look-ahead scenarios, an element of what is now called "macroprudential" regulation (that is, policies aimed at protection of the entire financial system) in which potential reactions across financial firms and markets are considered, not just whether individual companies seem to be holding sufficient capital.

The Dodd–Frank Act and the Fed's Role Today

The Dodd–Frank Act of 2010 targeted several of the most glaring holes in the pre-existing financial regulatory structure, with significant implications for the Federal Reserve's role as supervisor and as lender of last resort. As of mid-2013, many important components of the legislation are still in the rule-writing stage, and thus any assessment of the law's effect is necessarily preliminary. The Dodd–Frank Act unambiguously

expanded the Fed's role as a supervisor of financial institutions. However, the legislation was drafted and passed during a time when the Fed was under tremendous political and media pressure for its actions during the financial crisis, and this pressure led to some restrictions on the Fed's discretionary power as a lender of last resort.

From a supervisory viewpoint, the 2010 legislation created the Financial Stability Oversight Council, a new coordinating body that has the power to designate some financial institutions (including nonbanks) as being systemically important, with these institutions then subject to oversight and (additional) regulation by the Fed. Such designations effectively make the Fed a primary regulator for all large financial institutions, no matter what their main function. Furthermore, the Fed now has an explicit mandate to set higher capital standards and to give extra scrutiny to these largest firms.

One motivation of the Dodd–Frank Act was to end public bailouts of the largest institutions. Such a promise is complex and somewhat at odds with the lender-of-last-resort function. Specifically, the 13(3) powers that the Federal Reserve used during the crisis have been restricted by requiring more cooperation with the Treasury, more disclosure to Congress, and less flexibility to design programs to aid specific borrowers. In addition to the restrictions on the Fed's 13(3) powers, other restrictions were made on Treasury's emergency use of rescue powers such as those used for money-market funds, and the ability of the Federal Deposit Insurance Corporation to broadly guarantee bank assets without an act of Congress. Taken together, Dodd–Frank significantly reduced the flexibility of the executive branch and the Federal Reserve to act quickly during a financial crisis, while expanding their ability to act pre-emptively before one.

The Dodd–Frank Act did little to address the vulnerabilities in the shadow banking system at the heart of the panic during the crisis. For instance, repurchase agreements serve as a market for short-term loans and can be a source of troubles in a crisis when such loans are not rolled over as expected; yet reform of repurchase agreements was left entirely out of the legislation, with no clear jurisdiction for any agency to act. Reform of money market mutual funds was left to the existing statutory powers of the Securities and Exchange Commission, and it is has proved difficult (so far) to make significant changes to the status quo. Financial securitization received some new rules under which those who originally make loans need to retain some of the risk, rather than completely passing it on to others, but larger-scale reforms were not included. The Financial Stability Oversight Council has some flexibility to address all of these shadow-banking issues in the future, but the necessary powers are still untested. Overall, the Fed and other regulators still have significant limitations for liquidity provision and oversight for many of the shadow banking markets in which financial runs occurred in 2007–2008.

Conclusion

The Federal Reserve plays a central role in financial regulation, with responsibility as both a lender of last resort and as a supervisor for the largest institutions.

The discount window was originally intended to provide this lender-of-last-resort function through the provision of contingent liquidity to banks; that is, there would always be a credible supplier of liquidity should the state of the world be one in which depositors would otherwise run on the banks. If this institution was credible, then depositors would never run. But, in the 1920s the main concern of the Fed was to discourage discount window borrowing. The intellectual and policy history of the discount window following the Great Depression is one of discouraging its use with virtually no thought about its role in preventing crises. In an ideal world, all depositors (wholesale and retail) would be confident that the Fed would lend freely in a systemic crisis, while letting all institutions fail outside of a crisis. In the less-than-ideal real world, there are many challenges to this balance. The Fed is now in the position of having to try to reestablish its credibility to meet bank runs. We conclude with a statement of three of these challenges.

First, in the recent financial crisis, the Fed extended its lender-of-last-resort function beyond traditional banks, recognizing the broad expansion of the financial system. These programs have been discontinued. To prevent market panics, do these programs need to be in existence all the time, just like the discount window for traditional banks?

Second, given the continued reluctance to borrow from the discount window, and the new informational requirements for other emergency lending programs, what are the Fed's best options to reduce the stigma for its lender-of-last-resort function so that it has the tools to prevent liquidity runs before they start? How can the expectations of market participants be changed to believe that these programs are sufficient to prevent runs?

Third, following the passage of the Dodd–Frank Act, the Fed has a greatly expanded responsibility for supervision of the largest financial institutions and for the monitoring of financial stability. What is the optimal way to perform these functions to prevent future liquidity crises?

■ *Thanks to David Autor, Doug Diamond, Chang-Tai Hseih, Anil Kashyap, Ulrike Malmendier, Christina Romer, David Romer, and Timothy Taylor for helpful comments, and to Ellis Tallman for sharing Figure 1, which appears in Tallman (2010).*

References

Admati, Anat, and Martin Hellwig. 2013. *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It.* Princeton University Press.

Almantier, Olivier, Eric Ghysels, Asani Sarkar, and Jeffrey Shrader, Jr. 2011. "Stigma in Financial Markets: Evidence from Liquidity Auctions and

Discount Window Borrowing during the Crisis." Federal Reserve Bank of New York Staff Report, no. 483. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1754558.

Alston, Lee J., Wayne A. Grove, and David C. Wheelock. 1994. "Why Do Banks Fail? Evidence from the 1920s." *Explorations in Economic History* 31(4): 409–431.

Anderson, Clay J. 1966. "Evolution of the Role and Functioning of the Discount Mechanism." Study in the series "Fundamental Reappraisal of the Discount Mechanism." Federal Reserve Bank of Philadelphia. Available at FRASER, Federal Reserve Archive: http://fraser.stlouisfed.org/publication/?pid=706.

Angly, Edward. 1931. *Oh, Yeah?* New York: Viking.

Ashcraft, Adam B., Morten L. Bech, and W. Scott Frame. 2010. "The Federal Home Loan Bank System: The Lender of Next-to-Last Resort?" *Journal of Money, Credit and Banking* 42(4): 551–83.

Baker, Malcom, and Jeffrey Wurgler. 2013. "Do Strict Capital Requirements Raise the Cost of Capital? Bank Regulation and the Low Risk Anomaly." NYU Working Paper no. FIN-13-003.

Bordo, Michael, and David Wheelock. 2011. "The Promise and Performance of the Federal Reserve as Lender of Last Resort, 1914–1933." NBER Working Paper 16763.

Butkiewicz, James. 1995. "The Impact of a Lender of Last Resort during the Great Depression: The Case of the Reconstruction Finance Corporation." *Explorations in Economic History* 32(2): 197–216.

Butkiewicz, James. 1999. "The Reconstruction Finance Corporation, the Gold Standard, and the Banking Panic of 1933." *Southern Economic Journal* 66(2): 271–293.

Calomiris, Charles W. 1994. "Is the Discount Window Necessary? A Penn–Central Perspective." *Federal Reserve Bank of St. Louis Review* 76(3): 31–56.

Calomiris, Charles W., and Joseph R. Mason. 1997. "Contagion and Bank Failures during the Great Depression: The June 1932 Chicago Banking Panic." *American Economic Review* 87(5): 863–83.

Chandler, Lester. 1967. "Selective Credit Control." Study in the series "Fundamental Reappraisal of the Discount Mechanism." Board of Governors of the Federal Reserve System. Available at FRASER, Federal Reserve Archive: http://fraser.stlouisfed.org/publication/?pid=706.

Clark, Truman A. 1986. "Interest Rate Seasonals and the Federal Reserve." *Journal of Political Economy* 94(1): 76–125.

Comptroller of the Currency. 1915. *Annual Report, 1914.*

Dang, Tri Vi, Gary Gorton, and Bengt Holmström. 2013. "Ignorance, Debt and Financial Crises." Unpublished paper.

DeAngelo, Harry, and Rene M. Stulz. 2013. "Why High Leverage is Optimal for Banks." ECGI Finance Working Paper no. 356/2013, European Corporate Governance Institute.

Demirgüç-Kunt, Asli, and Enrica Detragiache. 1998. "The Determinants of Banking Crises in Developing and Developed Countries." *IMF Staff Papers* 45(1): 81–109.

Federal Reserve System. 1921, 1923, 1926. Annual Reports.

Fishe, Raymond P. H. 1991. "The Federal Reserve Amendments of 1917: The Beginning of a Seasonal Note Issue Policy." *Journal of Money, Credit and Banking* 23(3): 308–326.

Fishe, Raymond P. H., and Mark Wohar. 1990. "The Adjustment of Expectations to a Change in Regime: Comment." *American Economic Review* 80(4): 968–76.

Fleming, Michael J., Warren B. Hrung, and Frank M. Keane. 2010. "Repo Market Effects of the Term Securities Lending Facility." Federal Reserve Bank of New York Staff Reports 426.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States.* Princeton University Press.

Glass, Carter. 1927. *An Adventure in Constructive Finance.* New York: Doubleday, Page and Company.

Goodhart, Charles. 2011. *The Basel Committee on Banking Supervision: A History of the Early Years 1974–1997.* Cambridge University Press.

Gorton, Gary. 1984. "Private Bank Clearinghouses and the Origins of Central Banking." Federal Reserve Bank of Philadelphia *Business Review*, January/February, 3–12.

Gorton, Gary. 1985. "Clearinghouses and the Origin of Central Banking in the United States." *Journal of Economic History* 45(2): 277–83.

Gorton, Gary. 1988. "Banking Panics and Business Cycles." *Oxford Economic Papers* 40(4): 751–81.

Gorton, Gary, and Lixin Huang. 2006. "Bank Panics and the Endogeneity of Central Banking." *Journal of Monetary Economics* 53(7): 1613–29.

Gorton, Gary, Stefan Lewellen, and Andrew Metrick. 2012. "The Safe-Asset Share." *American Economic Review* 102(3): 101–106.

Gorton, Gary, and Andrew Metrick. 2010. "Regulating the Shadow Banking System." *Brookings Papers on Economic Activity*, Fall, 41(2): 261–97.

Gorton, Gary, and Andrew Metrick. 2012. "Securitized Banking and the Run on Repo." *Journal of Financial Economics* 104(3): 425–51.

Gorton, Gary, and Don Mullineaux. 1987. "The Joint Production of Confidence: Endogenous Regulation and Nineteenth Century

Commercial-Bank Clearinghouses." *Journal of Money, Credit and Banking* 19(4): 458–68.

Gorton, Gary B., and Andrew Winton. 2000. "Liquidity Provision, Bank Capital, and the Macroeconomy." Available at SSRN: http://ssrn.com/abstract=253849 or http://dx.doi.org/10.2139/ssrn.253849.

Hackley, Howard H. 1973. *Lending Functions of the Federal Reserve Banks: A History*. Board of Governors of the Federal Reserve System.

Hamilton, David. 1985. "The Causes of the Banking Panic of 1930: Another View." *Journal of Southern History* 51(4): 581–608.

Hamlin, Charles. 1914. "The Federal Reserve System as Established and in Operation." An Address delivered before the New York Chamber of Commerce, December 3.

Hanson, Samuel, Anil Kashyap, and Jeremy Stein. 2011. "A Macroprudential Approach to Financial Regulation." *Journal of Economic Perspectives* 25(1): 3–28.

Harding, Warren P. G. 1921. "Working Back to Normal." An Address delivered at the Annual Dinner of Group VIII of the New York State Bankers' Association, New York City, January 17.

Harding, Warren P. G. 1925. *The Formative Period of the Federal Reserve System*. Boston and New York: Houghton Mifflin Company.

Hrung, Warren B., and Jason S. Seligman. 2011. "Responses to the Financial Crisis, Treasury Debt, and the Impact on Short-Term Money Markets." Federal Reserve Bank of New York Staff Report 481.

Kashyap, Anil, Jeremy C. Stein, and Samuel G. Hanson. 2010. "An Analysis of the Impact of 'Substantially Heightened' Capital Requirements on Large Financial Institutions." Unpublished paper.

Keynes, John Maynard. 1930 [1960]. *A Treatise on Money*, Vol. 2. London: Macmillan and Co.

Laughlin, J. Laurence. 1933. *The Federal Reserve Act: Its Origin and Prospects*. New York: Macmillan and Co.

Leffingwell, Russell C. 1921. "The Discount Policy of the Federal Reserve Banks: Discussion." *American Economic Review* 11(1): 30–36.

Mankiw, N. Gregory, Jeffrey A. Miron, and David N. Weil. 1987. "The Adjustment of Expectations to a Change in Regime: A Study of the Founding of the Federal Reserve." *American Economic Review* 77(3): 358–74.

Mehra, Alexander. 2010/2011. "Legal Authority in Unusual and Exigent Circumstances: The Federal Reserve and the Financial Crisis." *University of Pennsylvania Journal of Business Law* 13(1): 221–72. (Published in 2010 and updated in 2011 after publication.)

Meltzer, Allan H. 2003. *A History of the Federal Reserve*, Vol. 1: *1913–1951*. University of Chicago Press.

Miron, Jeffrey A. 1986. "Financial Panics, the Seasonality of the Nominal Interest Rate, and the Founding of the Fed." *American Economic Review* 76(1): 125–40.

Mitchell, Wesley C. 1922. "The Crisis of 1920 and the Problem of Controlling Business Cycles." *American Economic Review* 12(1): 20–32.

Owen, Robert L. 1919. *The Federal Reserve Act*. New York; Century Co.

Richardson, Gary. 2007. "The Collapse of the United States Banking System during the Great Depression, 1929 to 1933: New Archival Evidence." *Australasian Accounting Business and Finance Journal* 1(1): 39–50.

Richardson, Gary, and William Troost. 2009. "Monetary Intervention Mitigated Banking Panics during the Great Depression: Quasi-Experimental Evidence from a Federal Reserve District Border, 1929–1933." *Journal of Political Economy* 117(6): 1031–73.

Ridgely, William B. 1908. "The Nation's Currency and Banking System." *North American Review* 627(February): 166–75.

Shiller, Robert J. 1980. "Can the Federal Reserve Control Real Interest Rates?" Chap. 4 in *Rational Expectations and Economic Policy*, edited by Stanley Fischer. National Bureau of Economic Research and University of Chicago Press.

Shull, Bernard. 1971. "Report on Research Undertaken in Connection with a System Study." In *Reappraisal of the Federal Reserve Discount Mechanism*, Vol. 1, p. 249–70. Board of Governors of the Federal Reserve System.

Shull, Bernard. 1993. "The Limits of Prudential Supervision: Economic Problems, Institutional Failure and Competence." Levy Economics Institute Working Paper no. 88.

Silber, William. 2007. *When Washington Shut Down Wall Street: The Great Financial Crisis of 1914 and the Origins of America's Monetary Supremacy*. Princeton University Press.

Tallman, Ellis W. 2010. Comment on 'The Promise and Performance of the Federal Reserve as Lender of Last Resort, 1914–1933,' by Michael D. Bordo and David C. Wheelock, presented at the conference "A Return to Jekyll Island: The Origins, History, and Future of the Federal Reserve," Federal Reserve Bank of Atlanta and Rutgers University, November 5–6, 2010, Jekyll Island Club Hotel, Jekyll Island, Georgia.

Timberlake, Richard H. 1984. "The Central Banking Role of Clearinghouse Associations." *Journal of Money, Credit and Banking* 16(1): 1–15.

Todd, Walker. 1992. "History of and Rationales for the Reconstruction Finance Corporation."

Federal Reserve Bank of Cleveland *Economic Review*‹QIV, 22–35.

West, Robert Craig. 1974. *Banking Reform and the Federal Reserve, 1863–1923*. Ithaca and London: Cornell University Press.

Wheelock, David. 1990. "Member Bank Borrowing and the Fed's Contractionary Monetary Policy during the Great Depression." *Journal of Money, Credit and Banking* 22(4): 409–26.

White, Eugene. 2009. "Lessons from the Great American Real Estate Boom and Bust of the 1920s." NBER Working Paper 15573.

Wicker, Elmus. 1980. "A Reconsideration of the Causes of the Banking Panic of 1930." *Journal of Economic History* 40(3): 571–83.

Wicker, Elmus. 1996. *The Banking Panics of the Great Depression*. Cambridge University Press.

Willis, Henry Parker. 1923. *The Federal Reserve System: Legislation, Organization and Operation*. New York: The Roald Press Company.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 65–86

Shifts in US Federal Reserve Goals and Tactics for Monetary Policy: A Role for Penitence?[†]

Julio J. Rotemberg

During its first century, the Federal Reserve has made a substantial number of changes in the conduct of monetary policy. Figure 1 plots a short-term interest rate, a measure of inflation, and the dates of recessions (shaded areas), which allows one to separate the 100-year history of policy making at the Fed into distinct periods. I focus on four of them.

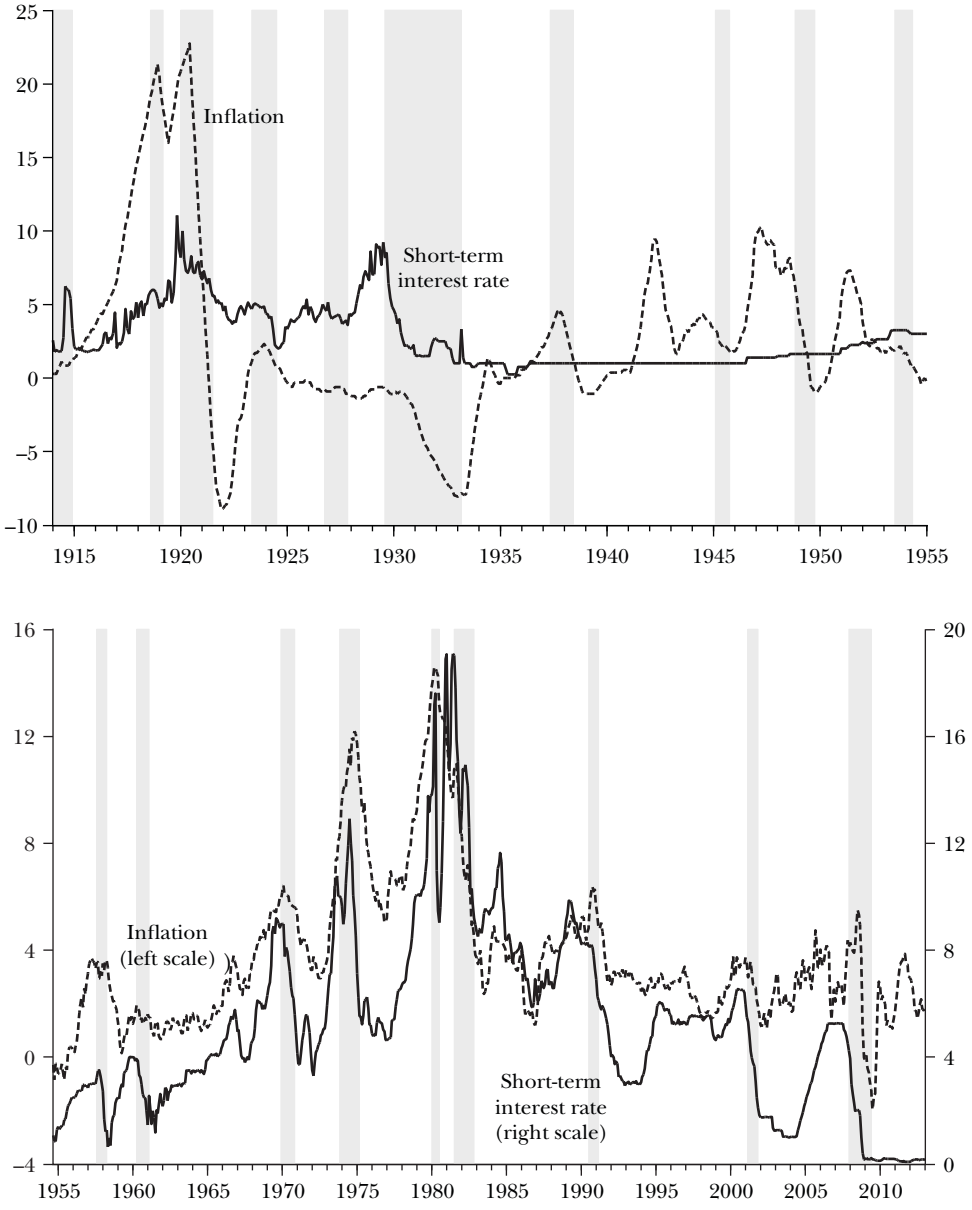
During the first period, starting in the mid-1920s, the Federal Reserve official policy was to support high-quality bank lending, but not speculative lending. This goal was set aside once in 1927, in an episode that many observers then blamed for the economic collapse that followed the financial crash of 1929. The Fed was then reluctant to increase the funds available to banks through the early 1930s, even as the Great Depression ravaged the economy. The Fed's concern with the volume and quality of lending in the setting of monetary policy did eventually wither. However, this only seems to have happened after the publication of Friedman and Schwartz (1963), a revisionist history of the Great Depression that blamed its depth on the Fed's inappropriate focus on "productive lending."

In the second period, after the experience of post-World War II inflation, the Federal Reserve in the 1950s was highly concerned with inflation and was willing to raise interest rates and bring on recessions to nip even modest inflation rates in the bud. This brought withering criticism for the Federal Reserve on the grounds that the recessions of 1957 and 1960 had been unnecessary. By the mid-1960s, some Fed officials seem to have developed an aversion to creating recessions as a method of

■ *Julio J. Rotemberg is William Ziegler Professor of Business Administration, Harvard Business School, Boston, Massachusetts. His email address is jrotemberg@hbs.edu.*

[†] To access the disclosure statement, visit
http://dx.doi.org/10.1257/jep.27.4.65 doi=10.1257/jep.27.4.65

Figure 1
Interest Rate Policy, Inflation, and NBER Recessions



Notes: Due to data availability, the variables (and sources) are not the same for these two panels. In the 1955–2013 panel, the short-term interest rate is the federal funds rate from the Federal Reserve Board while the inflation rate is the growth rate in the Consumer Price Index (CPI) (from the Bureau of Labor Statistics) over the past 12 months. The data used to construct the 1914–1955 panel are drawn from the NBER Macrohistory Database. The short-term interest rate is the call money rate while inflation is given by the 12-month growth rate in the NBER's estimate of CPI inflation. Periods of NBER recessions are indicated by shading.

fighting inflation, and this aversion may have contributed to the Great Inflation of the late 1960s and 1970s.

The third period I focus on is Paul Volcker's pursuit of disinflation from 1979 to 1982. This involved a change in operating procedures that yielded unparalleled interest volatility. This too seems responsive to a criticism, in this case that the Fed's focus on interest rates as an intermediate target led to large departures from the Fed's announced paths for the growth of monetary aggregates.

Finally, the fourth period from 1982–2007 was a time of renewed inflation intolerance known as the "Great Moderation." This period shows that, although the Volcker-led deflation of the late 1970s and early 1980s was widely viewed as a success, the Fed continued to change its approach to monetary policy. The federal funds rate became more stable, for example, though this change was much more gradual than the change in 1979.

A theme that emerges in these episodes is the tendency of the Fed to alter its methods and its objectives drastically when critics successfully argue that "bad outcomes" are a product of Fed "mistakes." The Fed then acts as if it were penitent, in that it becomes averse to this now vilified pattern of behavior. My discussion draws on Romer and Romer (2002), in that they too emphasize the role of policymakers' ideas in the determination of Fed policy. However, many of the changes in ideas emphasized in the existing literature on the Fed are unrelated to the penitence scheme I propose here.

The Two Abandonments of the Quality of Bank Lending as an Objective

The *Tenth Annual Report of the Federal Reserve Board* (1924) is often taken as a landmark statement of its policy intentions in the 1920s. The report specified that the Fed should extend credit only for "productive" and not for "speculative" purposes (p. 33). At a minimum, this implied that loans made by the Fed to individual banks needed to be collateralized with loans that those banks had made for industry, agriculture, and so on. The Report worried, however, that "paper offered by a member bank when it rediscounts with a Federal Reserve bank may disclose the purpose for which the loan evidenced by that paper was made, but it does not disclose what use is to be made by the proceeds of the rediscount" (p. 35). Regional Federal Reserve banks were thus supposed to keep tabs on the overall lending portfolio of the individual banks borrowing from them. In addition, the Fed was supposed to use a "quantitative" criterion to limit "the volume of credit within the field of its appropriate uses to such amounts as may be economically justified—that is justified by a commensurate increase in the Nation's aggregate productivity" (p. 33).

In late 1925, this approach led the Fed to tighten monetary policy on the grounds that loans for purchases of securities had been rising (Wicker 1966, p. 98). This tightening does not seem to have generated much controversy, the resulting recession was mild, and there were no loud complaints afterwards.

However, in 1927 the Fed pursued an expansionary policy that temporarily ignored the objective of lending only for the purposes of productive credit. This expansionary turn was championed by Benjamin Strong, the president of the New York Federal Reserve, who was motivated at least in part by a desire to lower US interest rates so that England would find it easier to get back on the gold standard (Wicker 1966, p. 112). On the other side, Governor of the Federal Reserve System "Adolph Miller bitterly opposed [this] . . . on the ground that purchases of securities would fan the flames of stock market speculation" (p. 106). James McDougal and George Norris, the heads of the Federal Reserve Banks of Chicago and Philadelphia, actually wanted to raise rates at the time (Meltzer 2003, p. 226). McDougal famously resisted lowering his own discount rate in line with the requirements of Strong's policy (the Federal Reserve Board ultimately succeeded in reasserting its oversight over regional Federal Reserve Bank discount rates).

The Fed quickly changed gears and started raising discount rates in early 1928 (Friedman and Schwartz 1963, p. 289; Hamilton 1987), and the tight policy was continued, and even somewhat strengthened in 1929. A key reason was that the Fed was unhappy with the substantial increase in speculative lending that took place in 1928 while the stock market was booming. As the Fed said in its 1929 Annual Report, "The problem was to find suitable means by which the growing volume of security credit could be brought under orderly restraint without occasioning avoidable pressure on commercial credit and business." The Board asked regional banks to limit the credit they extended to banks that engaged in speculative lending. Friedman and Schwartz (1963, p. 257) report that several Regional Banks, including the New York Federal Reserve, resisted this pressure for "direct action." Instead, George Harrison of the New York Federal Reserve wished to curb speculation by raising rates further.

Several Federal Reserve officials blamed the open market operations of 1927 for the dramatic contraction that took place between September 1929 and September 1930. For example, Governor Miller's congressional testimony of January 1931 depicted the breakdown of the autumn of 1929 as an "inevitable" consequence of the increase in asset prices and linked these directly to the 1927 monetary expansion. After noting that the Fed had purchased a great many government securities in 1927, he said: "Coupled with the heavy purchases of [bankers'] acceptances it was the greatest and boldest operation ever undertaken by the Federal Reserve system, and, in my judgment, resulted in one of the most costly errors committed by it or any other banking system in the last 75 years" (US Senate, 1931, p. 134). Treasury Secretary Glass, who had a direct role in the Federal Reserve at the time, was also convinced that the 1929–30 collapse was due to the abandonment of the doctrine that lending should only be directed to "productive uses" (Meltzer 2003, p. 470).

According to Friedman and Schwartz (1963), the depth of the subsequent Great Depression was due to the timidity of the Fed's response. The Fed did lower interest rates in 1929 and 1930, and while it loaned less to banks, it engaged in modest open market purchases so that the money supply (as measured by M1) fell only modestly. But when bank runs became widespread, the Fed generally refused

to lend to banks subject to runs. Moreover, the Fed resisted large-scale open-market purchases to offset the declines in banking, even as the money supply dropped substantially. Under pressure from Congress, such a program was started in April 1932, though it quickly ended in August.

Meltzer (2003, pp. 327–328, pp. 341, 364) and Romer and Romer (2013) point out that several Fed officials argued that, because banks were holding excess reserves, monetary conditions were easy and attempts to loosen monetary policy further would be ineffective. Meltzer (2003) and Romer and Romer (2013) suggests that this explains the Fed inaction at the time, but this explanation seems incomplete as an explanation of the Fed's behavior because some Fed members including Chairman Meyer favored increasing purchases even in 1933. Meyer's lack of success presumably owes something to people who saw expansionary policy not as irrelevant, but as actually detrimental. Negative views of this sort were expressed by Federal Reserve Bank of Richmond President George Seay, who "believed that the dangers of a further accumulation of reserves were greater than those of disposing of some securities" (Open Market Policy Conference Meeting, January 4, 1933). As excess reserves increased further in the 1930s, this concern became more widespread and reserve requirements were doubled between 1935 and 1937 (Meltzer 2003, p. 509).

A common explanation for the Fed's unwillingness to be more expansionary in this period is that it stuck to the principles of its *Tenth Annual Report* (1924) and to the procedures it had adopted in its wake (Calomiris and Wheelock 1998; Meltzer 2003, p. 400). As Friedman and Schwartz (1963, p. 411) argued, however, the expansionary policy of 1927 seems to represent a break from these principles and procedures. Given that this break was later condemned, it seems possible that penitence for departing from these principles in 1927 played a role in the 1930s. If the Fed now viewed the 1927 open market purchases as a mistake because they increased the liquidity of banks without a clear sense that this would be used for productive lending, penitence would be consistent with the Fed's aversion to excess reserves during the 1930s.

Of course, other factors contributed to the Fed's relatively tight stance. The 1931 increase in discount rates was clearly designed to stem gold outflows, for instance, so faithfulness to the ideals of the gold standard must have mattered too (Eichengreen 1992). However, Hsieh and Romer (2006) argue that even before the gold inflows that followed the devaluation of 1933, the Fed had ample room for more expansionary policies.

The level and quality of bank loans continued to play a role in Federal Open Market Committee (FOMC) discussions for some time. In 1953, for example, New York Federal Reserve President Allen Sproul told the FOMC that "bank credit, except for consumer credit and perhaps mortgage credit, has not moved out of line with a balanced situation" so that the evolution of several classes of bank loans was still followed closely. This changed after Friedman and Schwartz (1963) published their landmark study showing that the depth of the Great Depression was attributable to the Fed's concern for "productive lending" and its lack of attention to monetary aggregates. Even as late as 1964, Friedman complained that independent central

banks inevitably fell under the influence of bankers and thus "put altogether too much emphasis on the credit effects of their policies and too little emphasis on the monetary effects" (US House of Representatives, 1964, p. 73). Instead, Friedman argued: "Monetary policy ought to be concerned with the quantity of money and not with the credit market" (p. 74).

Eventually, this perspective became dominant and, consistent with penitence for its pattern of behavior during the Great Depression, members of the Federal Open Market Committee stopped focusing on the asset side of bank balance sheets. In the detailed memoranda of the first three meetings of the FOMC in 1970, for example, there is no substantive discussion concerning the composition of bank lending. The aggregate behavior of the banking sector, and total bank credit in particular, were still discussed, though some members explicitly said that they thought monetary aggregates were more relevant.

One has to wait until after the financial crisis of 2007 to see a resurgence of the argument that the Federal Reserve should pay attention to the quality of loans being made by financial institutions. The lead-up to the Great Recession featured a substantial number of mortgages that ended up in default. The dynamics of the financial crisis also suggest (as in the formal model of Shleifer and Vishny 1992) that economic downturns can force banks to sell certain assets at fire sales prices. As noted by Stein (2012), this means that an increase in one bank's risky lending imposes an externality on other banks because it reduces the fire sale prices at which these other banks can dispose of their own assets. This externality suggests that the main institution charged with macroeconomic stabilization should pay some attention to the quality of loans being made and to how they would fare in a downturn.

Friedman and Schwartz's (1963) analysis of the Great Depression also seems responsible for Ben Bernanke's (2002) apology on the Fed's behalf for its Depression-era policies. Consistent with a degree of penitence for these policies, the Fed responded to the 2007 financial crisis with heroic efforts to prevent bankruptcies among liquidity providers and with dramatic increases in excess reserves. Such policies were the opposite of the Fed's passivity in the face of bank failures and its reluctance to allow excess reserves to rise during the Great Depression.

The Abandonment of Inflation Intolerance

The Eroding Anti-Inflation Stance of William McChesney Martin

During World War II and the rest of the 1940s, the Fed maintained the low interest rates desired by the rest of the US government. But after seeing inflation rise again in 1950–51, the Fed became less submissive and negotiations led to the Treasury–Federal Reserve Accord of 1951. In these negotiations, William McChesney Martin represented the Treasury. Once the negotiations were concluded, Martin was appointed Fed chairman so that, while the 1951 Accord recognized the Fed's independence, the Fed was widely expected to abide by President Truman's wishes for continued low interest rates. Instead, Martin's inaugural statement painted

inflation as more threatening "than the spectacular aggressions of enemies outside our borders," and the Fed immediately raised rates. Hetzel and Leach (2001), who describe how the Fed managed to reassert its independent basis of power in 1951, demonstrate that the Truman administration regarded these policies as a betrayal.

Martin's hawkish stance on inflation remained in evidence for some time. At the FOMC meeting of July 30, 1957, for example, not all participants viewed "inflationary pressure" as the paramount problem. Those that did proposed raising the discount rate further from 3 to 3.5 percent, even though interest rates had been rising since early 1955. Martin noted that a discount rate increase might "create . . . difficulties . . . from the standpoint of relations with the Treasury." He nonetheless added that "as far as he was concerned personally, he would want to assume the risk of being charged with precipitating a downturn rather than to take any action except one that was believed to be correct" (Minutes, July 30, 1957, p. 37–38). Discount rates were raised shortly after this meeting and, according to the dating by the National Bureau of Economic Research, a recession began in August 1957.

Barely a year later, even though the August 1958 level of the Consumer Price Index was actually 0.5 percent lower than in February 1958, some participants at the FOMC meeting of August 19, 1958, worried about the presence of an "inflation psychology." Aside from a modest rise in long-term interest rates, the main source of this concern appears to have been the rapid growth of bank credit and money. Again, Martin agreed that "the System was dealing with . . . an inflationary psychosis as well as inflationary psychology." Noting that the Treasury had not always done its part in fighting inflation, Martin added "that the System had to stand up and be counted in these things" (Minutes, August 19, 1958, p. 54). The Fed then embarked on a series of interest rate increases in 1958–59, and a new recession started in April 1960.

In this second case, the 12-month rate of inflation as measured by the Consumer Price Index never rose above 2 percent. The Fed was roundly and widely criticized by economists, with many examples on display during the Congressional hearings conducted by Wright Patman on the occasion of the Fed's 50th anniversary (US House of Representatives 1964). Paul Samuelson complained about the "disastrously biased tight-money capers of 1956–60" (p. 50). Dudley Johnson opined "that we have been paying a very dear price in terms of foregone production and unemployment to fight a nonexistent inflation," while Harry Johnson concurred saying that "in peacetime they have displayed a pronounced tendency to allow deflationary policies on the average" (p. 47). Milton Friedman testified, "Contrary to widely held views, the major mistakes of this kind in peacetime have all been in a deflationary direction" (p. 24).

While Martin refused to take responsibility for the downturns that Federal Reserve policy was widely perceived to have generated, he may nonetheless have been affected by this criticism, and this may explain why his commitment to fight inflation weakened. In September 1967, the Consumer Price Index had risen by 2.6 percent in the last year, the unemployment rate was considerably lower than

in August 1958, and the Fed had been lowering interest rates since November 1966. Some members of the Federal Open Market Committee had been expressing concern about inflation for several months. Martin recognized that "the simple logic of the economic situation implied the desirability of changing monetary policy" and then added, "[b]ut the overriding need at this point was to get some restraint from fiscal policy through a tax increase, and in his judgment that would be less likely if Congress came to believe that adequate restraint was being exercised by monetary policy" (FOMC Minutes, September 12, 1967, p. 78). As Bremner (2004, p. 237) notes in his biography, it was extraordinary for Martin to trust Congress to take the initiative against inflation. Nonetheless, monetary easing continued. In August 1968, when the 12-month inflation rate had climbed to 4.5 percent, Martin said that "the objective should be disinflation without recession" (FOMC Minutes, August 13, 1968, p. 81). The birth of the Great Inflation may thus be partly explicable by penitence over causing recessions earlier.

The Federal Open Market Committee did set a course for tighter monetary policy starting with the December 1968 meeting (Romer and Romer 1989). While Martin was absent from this meeting, he endorsed tight policy from then on. In the January 14, 1969, FOMC meeting, in particular, he said that "he thought monetary policy was now on the right track" and that, in his judgment, "it would be better to risk overstaying, rather than understaying, a policy of restraint" (Minutes, January 14, 1969, p. 73). The rate of money growth fell substantially. In December 1969, Milton Friedman (1969, p. 75) called this policy "unduly restrictive" and predicted it would lead to a recession. Indeed, a recession would soon start in November 1969. A short while later, Friedman (1970, p. 68) expressed satisfaction that his "close friend and former teacher Arthur Burns" would become chairman of the Federal Reserve, and urged the Fed to "shift promptly to a less restrictive policy."

The Flourishing of Inflation under Arthur Burns

Like many contemporaries, Arthur Burns was openly critical of the Fed actions that preceded the 1960 recession. Before taking office, he had written: "The abrupt shift in policy proved more restrictive than government officials planned or expected. Largely as a result of their actions, the economic expansion that started in April 1958 came to a premature end" (Burns 1969, pp. 284–85). Consistent with this, he was averse to creating recessions and told the Federal Open Market Committee in 1973 that "it was attempting to achieve an objective that had never been accomplished before—that of keeping the economy from developing an inflationary boom but without releasing forces of a new recession" (Memoranda of Discussion, March 20, 1973, p. 108).

Burns agreed with Friedman that the Fed needed to reduce the volatility of its own actions if it wanted to avoid unnecessary recessions. Friedman had testified, "The chief defect in Federal Reserve policy has been a tendency to go too far in one direction or the other, and then to be slow to recognize its mistake and correct it" (US House of Representatives 1964, p. 27). Echoing this sentiment,

Burns (1969, p. 284–85) had written before becoming chairman "we need to make necessary shifts of economic policy more promptly, so that they may be gradual instead of abrupt."

Once Burns joined the Fed, his conviction that smooth changes in monetary policy were desirable appears to have had two implications. First, he seemed unwilling to react sharply to the inflation facing him. As he put it in his July 1974 testimony, "From a purely theoretical point of view, it would have been possible for monetary policy to offset the influence that lax fiscal policies and the special factors have exerted on the general level of prices. . . . But an effort to use harsh policies of monetary restraint to offset the exceptionally powerful inflationary forces of recent years would have caused serious financial disorder and dislocation" (US House of Representatives 1974, p. 257).

Second, Burns repeatedly expressed his intention to extinguish inflation over a number of years. His July 1974 testimony, for example, also said that "we shall need to stay with a moderately restrictive monetary policy long enough to let the fires of inflation burn themselves out. . . . We are determined to reduce, over time, the rate of monetary and credit expansion to a pace consistent with a stable price level" (US House of Representatives 1974, p. 253, 258). Similarly, in July 1977, Burns said: "We've enunciated a policy and repeated it on every occasion, namely, that we will gradually move our longer-range [money supply] targets down so that, several years from now, the monetary basis for general price stability may be restored. We've been proceeding slowly, perhaps too slowly, but that is a debatable point" (FOMC Transcript, July 19, 1977, p. 32).

However, certain apparent inconsistencies in Burns's statements have allowed him to be characterized differently. In particular, Nelson (2005), DiCecio and Nelson (2013), and Romer and Romer (2013) have attributed Burns's general failure to act against inflation to his conviction that the Fed was somewhat impotent. In a statement reflecting this conviction, Burns declared at the Federal Open Market Committee meeting of April 7, 1970, that "the inflation that was occurring—and that was now being accentuated, how far he could not say—was of the cost-push variety. That type of inflation, he believed, could not be dealt with successfully from the monetary side and it would be a great mistake to try to do so" (FOMC Memoranda of Discussion, April 7, 1970, p. 49). Some members of the FOMC strongly disagreed with this position.

Nonetheless, Burns continued to make statements of this sort, particularly in connection with his advocacy of administrative controls to prevent excessive increases in wages and prices. His July 1971 testimony, for example, stated: "In my judgment, and in the judgment of the Board as a whole, the present inflation in the midst of substantial unemployment poses a problem that traditional monetary and fiscal remedies cannot solve as quickly as the national interest demands. That is what has led me, on various occasions, to urge additional governmental actions involving wages and prices" (*Federal Reserve Bulletin*, August 1971, p. 662). According to Wells (1994, p. 72), this testimony was instrumental in pressuring a reluctant President Nixon to impose wage and price controls less than a month later.

When these wage and price controls were eventually lifted, inflation rose considerably, and the Fed became sufficiently concerned to raise interest rates to the point of causing the 1974 recession. Indeed, interest rates were increased even as this recession was in progress. As noted by Wells (1994, p. 136), Burns's July 1974 testimony alludes to the costs that a fight against inflation would impose, and this suggests he was aware at the time that he had temporarily departed from gradualism. In any event, the ensuing disinflation brought Burns a great deal of notoriety and prestige (Wells 1994, p. 178).

Alternative Sources of the Great Inflation

The Great Inflation of the 1970s has been attributed to a number of additional forces. Fed officials may, for example, have felt that they could not be tough on inflation for fear of the reactions in Congress and the Executive Branch (Burns 1979). What is certain is that Nixon pressured Burns to maintain a high rate of money growth on the eve of the 1972 election. On the other side, it is difficult to provide concrete evidence that political pressure for looser monetary policy had much effect (Mayer 1999, p. 64–82); after all, politicians sometimes were extremely critical of the Fed for having *caused inflation*.[1]

Another view emphasizes the influence of the idea that a long-run downwards-sloping Phillips curve existed, so that higher inflation would bring down unemployment (Taylor 1992, p. 13; DeLong 1997). Analyses based on this idea were common among members of the Council of Economic Advisors in the 1960s (Romer and Romer 2002, p. 20). However, as far as I know, no one has found a Fed official arguing for higher inflation on the grounds that this would lower long-term unemployment. Indeed, several Fed officials went out of their way to distance themselves from this idea. For example, Martin testified in January 1963 that he thought the Phillips Curve was a "fallacy" (*Federal Reserve Bulletin*, February 1963, p. 124). Indeed, he suggested that the long-run relation between inflation and unemployment was actually upwards sloping when he said that low rates of unemployment "have been facilitated, and indeed made possible, by the absence of inflationary expectations on the part of both labor and management" (*Federal Reserve Bulletin*, December 1965, p. 1,678). Similarly, in the hearings conducted by Congressman Wright Patman in 1974 to pin the blame for inflation on the Fed (and thereby absolve budget deficits), Burns said the "so-called tradeoff between inflation and unemployment" was "quite misleading" (US House of Representatives 1974, p. 252). DiCecio and Nelson (2013) offer extensive additional evidence that Burns did not think a rise in inflation would lower unemployment.

[1] In a very interesting article, Weise (2012) shows that Federal Open Market Committee discussions were more likely to mention politicians who desired looser conditions in meetings in which the committee chose to loosen monetary policy. Note, however, that this correlation may reflect less the effect of outside pressure than the desire to present all the arguments that come to mind in favor of one's desired course of action.

Another literature seeking to explain the Great Inflation relies on imperfect information and learning by the Fed. In Sargent (1999), Primiceri (2006), and Carboni and Ellison (2009), the Fed acts as a rational decision maker that estimates the parameters governing the cost of disinflation, and inflation stops when these estimated parameters fall inside a particular region of parameter space. In Orphanides (2003) and Orphanides and Williams (2013), the Fed learns instead about the level of output or unemployment that is likely to trigger inflation. These last two variables do appear in Federal Open Market Committee discussions, whereas these discussions do not appear to involve the parameters governing the costs of disinflation (or even the relationship between future inflation and the variables chosen by the Fed). On the other hand, Bullard and Eusepi (2005) suggest that the post-1980 disinflation cannot be rationalized by a theoretical model in which the Fed learns about the level of output that triggers inflation.

Finally, an important contemporary explanation of the Great Inflation was that it was due to the use of a faulty operating procedure. The problem, according to Milton Friedman, was that the Fed targeted interest rates rather than targeting money growth directly. There clearly was some truth in this description of the Fed. In the Federal Open Market Committee meeting of March 16, 1976, for example, then-President of the New York Federal Reserve Paul Volcker said that "he favored . . . keeping the [federal funds] rate at about its current 4-3/4 per cent level or a little higher" and that "he would not want to see the funds rate move above 5 per cent at any time in the near future." He would, thus "set relatively wide ranges for the aggregates for the March–April period—say, 3 to 8 per cent for M1 and 6 to 11 per cent for M2" (FOMC Memoranda of Discussion, p. 64).[2]

At the time, the Fed operated in a context in which money growth rates were very much in the public eye. Many individuals had testified at Congressman Wright Patman's 1974 hearings that inflation was due to excessive money growth, and this had led Congress to pass a resolution in March 1975 requiring the Fed to publish its projections for money growth. Friedman (1975a) applauded this change on the ground that "the requirement that it state [money growth targets] publicly in advance and justify failure to achieve them makes it far more likely that they will be achieved." However, the Fed consistently overshot its M2 upper limit from the fourth quarter of 1976 until the third quarter of 1977, at which point it consistently started overshooting its M1 upper limit. This pattern was intensely criticized as fueling inflation, which duly rose in the period.

Friedman's (1975b) argued that the "anachronistic procedure" of targeting interest rates led to "self-reinforcing" errors in money growth rates. A mistake in which the Fed set the federal funds rate at a level that was too low would lead to high money growth rates and high inflation, which would itself tend to raise other market interest rates, thereby necessitating an even higher federal funds rate. An

[2] M1 and M2 are measures of the total money supply. While their definitions changed somewhat over time, M1 always included currency in circulation and most checking accounts, while it always excluded savings deposits and small time deposits, both of which were always included in M2.

alternative suggestion would have been to encourage the Fed to change its interest rate objectives more vigorously.[3] However, Friedman argued that the Federal Open Market Committee should target the growth of reserves (or the monetary base) and let all interest rates be determined by the market.

The Abandonment of Interest Rate Stability

Upon becoming Fed Chairman in August 1979, Paul Volcker was not a gradualist. He seemed quite willing to bring about an immediate recession to lower inflation. At the Federal Open Market Committee meeting of March 18, 1980, Governor Frederick Schultz said: "I doubt that we can get out of this situation without a recession, and I think the unkindest thing we can do is to drag this on." Volcker followed this with: "I share the thoughts that some people have expressed, most recently Governor Schultz, that we better get this over with in terms of minimizing the total pain over a period of time" (Transcript, p. 35–36).

This sentiment may not have been shared by the entire Federal Open Market Committee. Indeed, at the September 18, 1979, FOMC meeting, Governors Charles Partee, Emmett Rice, and Nancy Teeters, as well as Boston Federal Reserve President Frank Morris and Philadelphia Federal Reserve President David Eastburn were sufficiently concerned about the possibility of a recession that they were reluctant to raise interest rates (Transcript, September 18, 1979, pp. 19, 24, 26, and 28). As suggested by Lindsey, Orphanides, and Rasche (2005), their reluctance may have led Volcker to suggest the widely publicized change in procedures that the FOMC discussed and adopted on October 6, 1979. This section discusses both the effects of these new procedures and the possible reasons leading FOMC members to use them.

The Effects of the October 1979 Procedures

At the Federal Open Market Committee meeting of October 6, 1979, the Committee started instructing its trading operation to assume a particular level of bank borrowing from the Fed and, on this basis, set a target for nonborrowed bank reserves that would keep the growth of money aggregates within the ranges that had been announced previously.[4] At the same time, the Committee widened considerably the range of values that interest rates were allowed to take.

[3] Mayer (1999, p. 45) and Clarida, Galí, and Gertler (2000) emphasize the weakness of the response of the federal funds rate to inflation and output in this period, and this is related to Friedman's complaint in his November 1975 statement that the Fed did not lower interest rates rapidly enough during the 1974 recession (US Senate, 1975, p. 38).

[4] The procedures that the Fed adopted were not identical to those recommended by its critics. The focus on nonborrowed as opposed to total reserves or the monetary base was deemed by Allan Meltzer to lead to excessively volatile money growth (Rasche, Meltzer, Sternlight, and Axilrod 1982). Moreover, according to Friedman (1982), the requirement that banks hold reserves on the basis of their past (rather than their current) deposits also complicated the control of money. One reason the Fed may have settled on nonborrowed rather than total reserves might have been to stabilize interest rates somewhat.

Figure 1 shows that interest rates did become substantially more volatile after these procedures were instituted. The average of the absolute value of monthly changes in the federal funds rate from October 1979 to November 1980 was 145 basis points. For the twelve monthly changes from September 1978 to October 1979, it had been only 42 basis points. More generally, the volatility of interest rates immediately after October 1979 was both historically unprecedented and contrary to a key goal of the founding of the Federal Reserve (Strong 1922 [1989]). Consistent with that goal, the creation of the Fed had stabilized seasonal fluctuations in interest rates (Mankiw, Miron, and Weil 1987).

The procedures also had a mixed record in terms of keeping money growth rates within their announced ranges. In the period between October 6, 1979, and the Federal Open Market Committee meeting of January 8, 1980, money growth rates were close enough to their targets that Governor Partee considered the procedures to have been a "successful experiment" in the latter meeting (Transcript, p. 14). On the other hand, monthly money growth rates proved quite volatile under the new procedures (McCallum 1985). The standard deviation of monthly M1 growth rates was 9.3 percent from November 1979 to November 1981, whereas it had been only 4.6 percent from September 1977 to September 1979 inclusive.[5] Not surprisingly, Volcker complained that "we got criticized by the bankers when they were here the other day for having too much volatility in the money supply growth and too much volatility in interest rates" (Transcript, September 16, 1980, p. 9).

Moreover, there were long periods in which money growth exceeded its official target. In particular, the growth in M1 equaled 11 percent in the 11 months from May 1980 to April 1981, and this led the Fed to be severely criticized by some Reagan administration officials (Greider 1987, p. 378). One potential reason for this growth was that money market mutual funds and checking accounts that paid interest (NOW accounts) grew in this period. Financial innovation of this sort led Governor Morris to exclaim, "we simply don't have any basis for measuring what transactions balances are anymore" (FOMC Transcript, July 7, 1981, p. 24).

These failures to meet money targets should not be taken to mean that the procedures failed to have an effect on policy. Perhaps the most telling evidence that they mattered is that Volcker complained about their role in the October 5, 1982, meeting in which these procedures were at least partially jettisoned. Volcker was unhappy at the interest rate that had resulted from the previous meeting's decision concerning nonborrowed reserves and said: "What we did last time was unacceptable to me. I just want to make that plain. I think we made a mistake last time . . . [I]t's unfortunate that we ended up at this meeting with the federal funds rate and private rates about 1 percentage point higher than they were at the time of the last

[5] These figures and those below are based on current measures of seasonally adjusted M1. In December 1980, before all these data became available, two Federal Reserve economists presented a paper at the AEA annual meetings saying that money growth over longer periods of time was close to its targets under the new procedures (Axilrod and Lindsey 1981).

meeting because we had a high M1 figure in September. That was the only reason it happened" (FOMC Transcript, October 5, 1982, p. 32).

Rationales for the October 1979 Procedures

I start with the rationales that were given when the October 1979 procedures were first instituted and then discuss the reasons why they remained in place even after they had quite clearly failed to stabilize money growth. Volcker seemed an unlikely champion for these new procedures because he had stated, for example in a 1978 *Journal of Monetary Economics* article, that the demand for money was sufficiently unstable in both the short run and the long run that fixing money growth rates would lead to undesirable movements in interest rates. In Volcker (1978), he also seemed somewhat uncertain of the Fed's ability to hit its money growth targets by setting the level of reserves.

Nonetheless, Volcker gave an argument for these procedures in October 1979, namely that their announcement would lower inflation expectations. As he explained in Greider (1987, p. 111), "What I hoped was that there would be a strong reaction in the markets. . . . The sign of psychological success was whether long-term rates would stabilize and start coming down." This did not happen right away; long-term rates rose alongside short-term rates immediately after the October 1979 announcement.

Meltzer (2009, pp. 1040, 1064, 1075, and 1093) suggests that, more generally, the 1979 procedures had only a modest effect on inflation expectations, and that these fell mainly when economic activity slowed. After the procedures had been operating for a year, Volcker himself seemed to doubt that they mattered for inflation expectations. In December 1980, he said "If we, in effect, go to the brink or let some of these things happen that we have not allowed to happen during the entire postwar period, people are not expecting that and they are not going to be very happy if and when it happens. And I'm not at all sure that we can change inflationary expectations without it happening" (FOMC Transcript, December 19, 1980, p. 62).

Governor Partee's initial support may have been based in part on his view during the September 1979 meeting: "I think it's important, very important, that we try to keep the aggregates within the ranges that we specify" (FOMC Transcript, September 18, 1979, p. 26). Partee recalled a different reason for his approval in his interview in Greider (1987, p. 112). There, he declared that the new procedures dealt with the Fed's past tendency of sticking "stubbornly with a strong position too long and causing more damage to the economy than it had intended" and that in recessions, particularly in the 1974–75 recession "there [was] also a hesitancy to reduce interest rates once they have been raised." As it happens, this hesitancy to lower rates may have had some benefits. While interest rates rose substantially when the new procedures were instituted, the decline in rates when the 1980 recession started was so dramatic that the recession was over almost immediately, and the reduction in inflation to acceptable levels had to wait until the arrival of the 1981–82 recession. Partee did not mention any concern he might have had with "sticking stubbornly" to the 1979 procedures themselves if velocity shifted. Such velocity shifts did, in fact, eventually lead to difficulties with the procedures.

As the procedures were being abandoned, two arguments for keeping them became prominent. The first was that the procedures provided "political shelter" for raising rates to fight inflation (FOMC Transcript, February 8–9, 1983, p. 24, 26, 29, and 30). The procedures may have diminished the criticism of the Fed, but they certainly did not eliminate it. Indeed, the high rates of interest of 1982 had led to a strong movement in Congress to reduce the Fed's independence (Greider 1987, p. 474).

A second argument for keeping the procedures intact was made at the October 5, 1982, meeting in which the Federal Open Market Committee decided to announce that it would pay less attention to M1. Federal Reserve of St. Louis President Lawrence Roos, an ardent supporter of monetary targets, argued that reducing the official importance of the growth rate of M1 would imperil the Fed's credibility and would be "misconstrued by the markets" (FOMC Transcript, October 5, 1982, p. 48). In fact, the reduction in short-term interest rates that followed this meeting was accompanied by a reduction in long-term rates.

It would seem, then, that the arguments that were given for initiating and maintaining these procedures were not very strong. This suggests another possibility, namely that these procedures embodied a form of penitence for the pre-1979 procedures, which critics had successfully associated with the Great Inflation. Roos emphasized this association at the October 5, 1982, Federal Open Market Committee meeting when he argued that the high interest rates that prevailed at the time were the ultimate consequence of "irresponsible monetary policies throughout the world" and to "a well-meant effort on the part of the Federal Open Market Committee . . . to try to do just what we're doing today, and that is to lean against interest rate movements. I think that contributed in a major way to inflation" (Transcript, October 5, 1982, p. 48).

Extreme concern with the possibility of uncontrolled money growth if interest rates were stabilized even at very high levels was also on display at the earlier meeting of July 1, 1982, when Partee noted that he seemed "to have shocked quite a number of people with my suggestion that we ought to put a cap on the funds rate." He had proposed that the federal funds rate should not be allowed to rise above 15 percent. Since the rate that day was equal to 14.73 percent, this cap was perceived as being potentially binding. At the same time, the unemployment rate was 9.8 percent and the growth rate in the Consumer Price Index over the last 12 months had been 6.5 percent, so a 15 percent federal funds rate would have been likely to be associated with a high real interest rate. This led Partee to argue that this "would give us an upper limit that is not unreasonable."

Nonetheless, Partee was asked by Governor Henry Wallich, in apparent disbelief, "But if it got there, we would provide unlimited reserves?" and by Roos "how would that differ from the pre-1979 practices of our Committee?" When Partee answered it would be "similar on the top side," Federal Reserve Bank of Atlanta President William Ford said "Are you implying that there wasn't a change in October '79? If I understood you, you said it would be similar to pre-October '79—that there is precedent for it" (FOMC Transcript, July 1, 1982, p. 55). One reason for the aversion to returning to the pre-October 1979 may have been that, as Volcker and others

suggested, some members of the Federal Open Market Committee may have been afraid of losing their "self-discipline" if they were not constrained by "rules" (FOMC Transcript, December 21, 1982, p. 29, 38, and 43).

The "Great Moderation"

Once the procedures of targeting monetary aggregates were abandoned, interest rates came down and the Volcker disinflation was widely seen as a success. What followed was a period of low inflation and stable output growth that came to be referred to as the Great Moderation. This period involved a variety of gradual changes both in the way that policy was discussed inside the Federal Open Market Committee and in the way the Fed communicated with the public. This raises the possibility that the Fed's capacity to adapt its approach to changing circumstances is enhanced when it can claim credit for some successes.

The Fed changed its approach incrementally along several dimensions, beginning with the way the Federal Open Market Committee dealt with interest rates. At the December 1982 meeting of the FOMC, Paul Volcker made it clear that he wanted interest rates to be more stable than in the past (Transcript, December 21, 1982, p. 42). However, discussions at the FOMC meetings continued to emphasize the quantity of discount window borrowing for a considerable period after October 1982. Also, discount window borrowing remained central in the policy options laid out in the "Bluebook" that members received before the meeting. Different options involved different assumptions regarding the amount that banks would borrow from the Federal Reserve. On the grounds that it was trying to stabilize total money growth, the Federal Reserve System would supply fewer nonborrowed reserves if it assumed that the amount borrowed was larger—and nonborrowed reserves were the intermediate target for managing the aggregate money supply under the October 1979 procedures. On its own, a smaller supply of nonborrowed reserves would be expected to raise overnight federal funds interest rates. These higher market interest rates would create an incentive for banks to borrow from the Fed (at an unchanged discount rate), so that actual borrowing could be expected to be higher as well. To some extent, then, a higher assumed level of bank borrowing would tend to raise actual borrowing.

One has to wait until October 1989 to find a Bluebook that lays out policy alternatives in terms of levels of the federal funds rate and *expected* levels of borrowings rather than doing the reverse (that is, offering alternative assumptions about borrowing combined with implications for *expected* federal funds rates). Even at the October 1989 meeting, some members preferred to discuss policy in terms of borrowing. As time went on, this focus ceased, and the federal funds rate became the focus of discussion.[6] This is not to say that Fed chairmen were not targeting

[6] Thornton's (2006) quantitative evidence confirms this gradualism. He shows that the average distance of the federal funds rate from his constructed target was smaller after 1989, when it was still somewhat larger than it had been before 1979.

the federal funds rate much earlier. Indeed, as Thornton (2006) documents, some Federal Open Market Committee members openly suspected Volcker of doing so as early as 1983.

Unlike what happened in October 1979, the public was not told that a change in the conduct of monetary policy had taken place. No target for the federal funds rate was announced throughout the 1980s or into the early 1990s. Rather, just as had been true since 1983, the press releases continued to suggest that the federal funds rate would remain within a 4 percent range until the next meeting. Meanwhile, the Fed continued to publish its expected ranges for the growth in monetary aggregates, though it softened its commitment to these ranges.

Even in February 1993, many members of the Federal Open Market Committee expressed apprehension about releasing their federal funds target (Transcript, February 2, 1993, p. 62–67). But by then, movements in velocity of M1 and M2 had become so large that the Fed's plans regarding the growth in these aggregates were not very informative. After this point, its statements started explaining the federal funds rate changes that the FOMC had instituted in the past. Still, as late as March 1997, when FOMC members voted to raise the federal funds rate from 5.25 to 5.5 percent, the public minutes only commented on the past rate of 5.25 percent. This lack of transparency would finally end in August 1997, when the intended federal funds rates started to be published in the official minutes, although this was accompanied by a statement that the operating procedures of the Fed would not change. After this, the Fed gradually expanded the amount of information it released about its intentions concerning future policy (Woodford 2005). The Fed managed to stop supplying any monetary targets whatsoever when the legislation requiring these expired in 2000.

One of the most striking aspects of US monetary policy in this period is that the simple "rule" proposed by Taylor (1993)—in which the suggested federal funds rate is a function of inflation (as measured by the Consumer Price Index) over the last year and of the distance between current real GDP and trend GDP—leads to a federal funds rate that is remarkably close to the actual one for the period 1987 to 2000. This too was the result of a gradual evolution. Even though the correspondence is weaker before 1987, the relatively fast rise in the federal funds rate in 1983 and early 1984, as well as its subsequent decline were consistent with the Taylor rule. As Kahn (2012) demonstrates, discussion of the implications of variants of the Taylor rule for the federal funds rate quickly became part of the fabric of meetings of the Federal Open Market Committee. Nonetheless, the FOMC drifted towards applying the coefficients of the Taylor rule to their anticipations of future values of inflation rather than to the past values (FOMC Transcript, January 27, 2004, p. 76).

Conclusion

This paper has suggested that some of the changes in the Fed's approach to monetary policy are consistent with a form of penitence, where this penitence is the

end result of a three-step process. First, there are some deplorable economic results such as those in the initial 1930 downturn, the full Great Depression, the recessions of 1957 and 1960, or the Great Inflation. Second, critics attribute these results to patterns of Fed behavior that are interpreted as having been mistaken. Third, the Fed acts as if it implicitly accepted one of these criticisms and becomes averse to the criticized pattern of behavior. It is possible to view this form of penitence as helping the Fed perfect its approach to monetary policy. Particularly if one agrees with the critics, this penitence would represent a form of learning: it leads the Fed not to repeat mistakes.

Without further evidence, however, it seems premature to view this form of penitence as involving an accumulation of knowledge of the form one typically associates with learning. To see this, it is sufficient to imagine a two-state system that toggles from one state to the other whenever something bad happens outside the system. Such a system responds to poor outcomes, but is essentially devoid of historical information at all times.

The Fed has access to a rich menu of policy approaches, and one role of outsiders is to help devise new ones. Still, there are two aspects of the Fed's evolution that seem somewhat similar to the two-state system I just described, and which raise concerns over the extent to which the Fed's response to bad outcomes involves the accretion of knowledge. First, many of the changes in Fed behavior that follow such outcomes seem later to be reversed. In particular, the Fed both gained and lost its aversion to stabilizing interest rates, as well as its aversion to inducing recessions in response to inflation. Second, some knowledge seems to be lost when the Fed develops a new aversion. Entire topics can practically vanish from the discussions of the Federal Open Market Committee. As an example, the FOMC meeting of January 26, 1960, contained a remark by President of the Richmond Federal Reserve Hugh Leach in which he based his assessment of the tightness of monetary policy on the evolution of "loans to build up inventories" (Minutes, p. 20). Information of this sort stopped being incorporated into policy discussions when the Fed reduced its attention to the asset composition of bank balance sheets.

Even if one believes that the changes in approach triggered by poor outcomes have led to only limited accretions in Fed knowledge, the Fed may have accumulated a great deal of information at other times. During the Great Moderation, for example, the Fed appears to have gradually learned to stabilize interest rates to an ever-greater extent.

So how might the Fed's knowledge and approach evolve in response to the financial crisis of 2007? As was the case with previous bad outcomes, critics who blame this crisis on Fed mistakes do not speak with a single voice. Fleckenstein (2008) argues that the Fed started being prone to generate asset bubbles by having low interest rates as far back as 1987, when it lowered rates in response to a stock market crash. By contrast, Taylor (2012) applauds the Fed's approach from 1987 to 2003, and singles out for criticism the post-2003 period in which the Fed set interest rates below those implied by a backward-looking Taylor rule.

If such criticisms became accepted by the Fed to some extent, they could lead to dramatic changes in the Fed's approach by creating new aversions. The

Fed could, for example, seek to tamp down any potential increase in asset prices that it regarded as a "bubble," though it seems likely that such an approach would quickly lead the Fed to be criticized for causing unnecessary losses in output. Acceptance by the Fed that it had mistakenly kept interest rates too low starting around 2003 could result in different aversions. If a consensus developed that the Fed's mistake was to abandon a Taylor rule based on past values for one based on Fed projections, the Fed could become averse to using its forecasts in setting policy, at least for a time.

Another move that could come to be seen as an error is the Fed's policy of announcing its expectations concerning future policy actions. At the December 9, 2003, meeting of the Federal Open Market Committee, Governor Donald Kohn said "policy is quite easy, quite stimulative" and nonetheless recommended that the Fed "continue to take [its] risks on the easy side of policy." At the same time, he worried about the FOMC's "flexibility" to raise rates given that its August 2003 statement had said "that policy accommodation can be maintained for a considerable period" (FOMC Transcript, December 9, 2003, p. 67). This raises the possibility that Kohn felt trapped into keeping interest rates low to honor the Fed's implicit promise to do so. Thus, there is the possibility that the Fed's use of "forward guidance" concerning its future policies could come to be seen as a mistake. Consistent with penitence, the Fed might decide in the future to steer clear of communicating in a way that seeks to affect expectations of future policy.

Papers in the volume *Is Inflation Targeting Dead?* (Reichlin and Baldwin 2013) propose more gradual changes that would not require the development of an aversion to past Fed practices. As discussed earlier, one possibility along these lines would be to return partially to the pre-1963 view that monetary policy ought to respond to the quality of assets held by institutions with monetary liabilities (Stein 2013). More gradual changes may prove less prone to reversals, and this would constitute an advantage. To institute such gradual changes, more radical changes may need to be held at bay. To successfully counter arguments for more radical change it might help to understand how, in the past, critics often succeeded in championing the abandonment of practices that, eventually, came to be seen as beneficial once again.

■ *This is a revised version of a paper prepared for a symposium on "The First Hundred Years of the Federal Reserve: The Policy Record, Lessons Learned, and Prospects for the Future," held at the National Bureau of Economic Research on July 10, 2013. I wish to thank Chang-Tai Hsieh, Robin Greenwood, Anil Kashyap, Ulrike Malmendier, Edward Nelson, Robert Pindyck, Christina Romer, David Romer, and Timothy Taylor for comments and, especially, Rawi Abdelal for several conversations and for his help in crystallizing the argument of this paper.*

References

Axilrod, Stephen H., and David E. Lindsey. 1981. "Federal Reserve System Implementation of Monetary Policy: Analytical Foundations of the New Approach." *American Economic Review* 71(2): 246–52.

Bernanke, Ben. 2002. "On Milton Friedman's Ninetieth Birthday." Speech at the Conference to Honor Milton Friedman, University of Chicago, Chicago, IL, November 8. http://www.federalreserve.gov/BOARDDOCS/SPEECHES/2002/20021108/.

Bremner, Robert P. 2004. *Chairman of the Fed: William McChesney Martin and the Creation of the American Financial System*. New Haven: Yale University Press.

Bullard, James, and Stefano Eusepi. 2005. "Did the Great Inflation Occur Despite Policymaker Commitment to a Taylor Rule?" *Review of Economic Dynamics* 8(2): 324–59.

Burns, Arthur F. 1969. *The Business Cycle in a Changing World*. New York: NBER and Columbia University Press.

Burns, Arthur F. 1979. "The Anguish of Central Banking." The 1979 Per Jacobson Lecture, Belgrade, Yuogslavia, September 30.

Calomiris, Charles W., and David C. Wheelock. 1998. "Was the Great Depression a Watershed for Monetary Policy?" In *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*, edited by Michael D. Bordo, Claudia Goldin, and Eugene N. White, 23–65. University of Chicago Press.

Carboni, Giacomo, and Martin Ellison. 2009. "The Great Inflation and the Greenbook." *Journal of Monetary Economics* 56(6): 831–41.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115(1): 147–80.

DeLong, J. Bradford. 1997. "America's Peacetime Inflation: the 1970s." In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer, 247–76. University of Chicago Press.

DiCecio, Riccardo, and Edward Nelson. 2013. "The Great Inflation in the United States and the United Kingdom: Reconciling Policy Decisions and Data Outcomes." In *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides, 393–438. University of Chicago Press.

Eichengreen, Barry. 1992. *Golden Fetters: The Gold Standard and the Great Depression, 1919–1939*. New York: Oxford University Press.

Federal Open Market Committee (FOMC). Various years. Historical Minutes, Memoranda of Discussions, and Transcripts. Available at: http://www.federalreserve.gov/monetarypolicy/fomc_historical.htm.

Federal Reserve Board. 1924. *Tenth Annual Report of the Federal Reserve Board*. Washington, DC: Government Printing Office.

Federal Reserve Board. Various years. *Federal Reserve Bulletin*.

Fleckenstein, William. 2008. *Greenspan's Bubbles: The Age of Ignorance at the Federal Reserve*. With Frederick Sheehan. New York: McGraw Hill.

Friedman, Milton. 1969. "Monetary Overkill." *Newsweek*, August 18.

Friedman, Milton. 1970. "A New Chairman at the Fed." *Newsweek*, February 2.

Friedman, Milton. 1975a. "Congress and the Federal Reserve." *Newsweek*, June 2.

Friedman, Milton. 1975b. "How to Hit the Money Target." *Newsweek*, December 8.

Friedman, Milton. 1982. "Monetary Policy: Theory and Practice." *Journal of Money, Credit and Banking* 14(1): 98–118.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, NJ: Princeton University Press.

Greider, William. 1987. *Secrets of the Temple: How the Federal Reserve Runs the Country*. New York: Simon & Schuster.

Hamilton, James D. 1987. "Monetary Factors in the Great Depression." *Journal of Monetary Economics* 19(2): 145–69.

Hetzel, Robert L., and Ralph F. Leach. 2001. "The Treasury–Fed Accord: A New Narrative Account." Federal Reserve Bank of Richmond *Economic Quarterly* 87(1): 33–55.

Hsieh, Chang-Tai, and Christina D. Romer. 2006. "Was the Federal Reserve Constrained by the Gold Standard during the Great Depression? Evidence from the 1932 Open Market Purchase Program." *Journal of Economic History* 66(1): 140–76.

Kahn, George A. 2012. "The Taylor Rule and the Practice of Central Banking." In *The Taylor Rule and the Transformation of Monetary Policy*, edited George A. Kahn, Evan F. Koenig, and Robert Leeson, 63–101. Stanford: Hoover Institution Press.

Lindsey, David E., Athanasios Orphanides, and Robert H. Rasche. 2005. "The Reform of October 1979: How It Happened and Why." Federal Reserve Bank of St. Louis *Review*, 87(2, Part 2): 187–235.

Mankiw, N. Gregory, Jeffrey A. Miron, and David N. Weil. 1987. "The Adjustment of Expectations to a Change in Regime: A Study of the Founding of the Federal Reserve." *American Economic Review* 77(3): 358–74.

Mayer, Thomas. 1999. *Monetary Policy and the Great Inflation in the United States: The Federal Reserve and the Failure of Macroeconomic Policy, 1965–1979.* Northampton, MA: Edward Elgar.

McCallum, Bennett T. 1985. "On the Consequences and Criticism of Monetary Targeting." *Journal of Money, Credit, and Banking* 17(4): 570–97.

Meltzer, Allan H. 2003. *A History of the Federal Reserve,* Vol. I: *1913–51.* University of Chicago Press.

Meltzer, Allan H. 2009. *A History of the Federal Reserve,* Vol. II; Book 1: *1951–1969*, and Book 2: *1970–86.* University of Chicago Press.

Nelson, Edward. 2005. "The Great Inflation of the Seventies: What Really Happened?" *Advances in Macroeconomics* 5(1): Article 3.

Open Market Policy Conference (OMPC). 1933. Meeting, January 4. Available at: FRASER: Federal Reserve Archive: http://fraser.stlouisfed.org/topics/?tid=93.

Orphanides, Athanasios. 2003. "The Quest for Prosperity without Inflation." *Journal of Monetary Economics* 50(3): 633–63.

Orphanides, Athanasios, and John C. Williams. 2013. "Monetary Policy Mistakes and the Evolution of Inflation Expectations." In *The Great Inflation: The Rebirth of Modern Central Banking,* edited by Michael D. Bordo and Athanasios Orphanides, 255–300. University of Chicago Press.

Primiceri, Giorgio E. 2006. "Why Inflation Rose and Fell: Policy-Makers' Beliefs and U.S. Postwar Stabilization Policy." *Quarterly Journal of Economics* 121(3): 867–901.

Rasche, Robert H., Alan H. Meltzer, Peter D. Sternlight, and Stephen H. Axilrod. 1982. "Is the Federal Reserve's Monetary Control Policy Misdirected?" *Journal of Money, Credit and Banking* 14(1): 119–47.

Reichlin, Lucrezia, and Richard Baldwin, eds. 2013. *Is Inflation Targeting Dead? Central Banking after the Crisis.* London: Centre for Economic Policy Research.

Romer, Christina D., and David H. Romer. 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual 1989*, vol. 4, pp. 121–70.

Romer, Christina D., and David H. Romer. 2002. "The Evolution of Economic Understanding and Postwar Stabilization Policy." In *Rethinking Stabilization Policy*, proceedings of a symposium sponsored by the Federal Reserve Bank of Kansas City, in Jackson Hole, Wyoming, August 29–31. Available at: http://kansascityfed.org/publications/research/escp/escp-2002.cfm.

Romer, Christina D., and David H. Romer. 2013. "The Most Dangerous Idea in Federal Reserve History: Monetary Policy Doesn't Matter." *American Economic Review* 103(3): 55–60.

Sargent, Thomas J. 1999. *The Conquest of American Inflation.* Princeton, NJ: Princeton University Press.

Shleifer, Andrei, and Robert W. Vishny. 1992. "Liquidation Values and Debt Capacity: A Market Equilibrium Approach." *Journal of Finance* 47(4): 1343–66.

Stein, Jeremy C. 2012. "Monetary Policy as Financial-Stability Regulation." *Quarterly Journal of Economics* 127(1): 57–95.

Stein, Jeremy C. 2013. "Overheating in Credit Markets: Origins, Measurement, and Policy Responses." Speech presented at the Federal Reserve Bank of St. Louis, February 7.

Strong, Benjamin. 1922 [1989] "Federal Reserve Control of Credit." Address delivered on November 28, 1922 at Harvard College, published in *Federal Reserve Bank of New York Quarterly Review*, Special 75th Anniversary Issue, 1989, pp. 6–14.

Taylor, John B. 1992. "The Great Inflation, the Great Disinflation, and Policies for Future Price Stability." In *Inflation, Disinflation and Monetary Policy*, edited by Adrian Blundell-Wignall, 9–34. Australia: Ambassador Press.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Series on Public Policy* 39(1): 195–214.

Taylor, John B. 2012. "Monetary Policy Rules Work and Discretion Doesn't: A Tale of Two Eras." *Journal of Money, Credit and Banking* 44(6): 1017–32.

Thornton, Daniel L. 2006. "When Did the FOMC Begin Targeting the Federal Funds Rate? What the Verbatim Transcripts Tell Us." *Journal of Money, Credit, and Banking* 38(8): 2039–72.

US House of Representatives, Committee on Banking and Currency, Subcommittee on Domestic Finance. 1964. *The Federal Reserve System after Fifty Years.* Proposals for Improvement of the Federal Reserve and Staff Report on Domestic Finance of the Committee on Banking and Currency, 88th Congress, 2nd session. Washington DC: US Government Printing Office.

US House of Representatives, Committee on Banking and Currency. 1974. *Federal Reserve Policy and Inflation and High Interest Rates.* Hearings before the Committee on Banking and Currency, House of Representatives, 93rd Congress. Washington DC: US Government Printing Office.

US Senate. 1931. *Operations of the National and Federal Reserve Banking Systems.* Hearings Before a

Subcommittee of the Committee on Banking and Currency, US Senate, 71st Congress, 3rd Session. Washington DC: US Government Printing Office.

US Senate, Committee on Banking, Housing and Urban Affairs. 1975. *Second Meeting on the Conduct of Monetary Policy*. Hearings before the Committee on Banking, Housing and Urban Affairs, United States Senate, 95th Congress, 2nd Session. Washington DC: US Government Printing Office.

Volcker, Paul A. 1978. "The Role of Monetary Targets in the Age of Inflation." *Journal of Monetary Economics* 4(2): 329–39.

Weise, Charles L. 2012. "Political Pressures on Monetary Policy during the US Great Inflation." *American Economic Journal: Macroeconomics* 4(2): 33–64.

Wells, Wyatt C. 1994. *Economist in an Uncertain World: Arthur F. Burns and the Federal Reserve, 1970–1978*. New York: Columbia University Press.

Wicker, Elmus R. 1966. *Federal Reserve Monetary Policy 1917–1933*. New York: Random House.

Woodford, Michael. 2005. "Central-Bank Communication and Policy Effectiveness." In *The Greenspan Era: Lessons for the Future*. A Symposium Organized by the Federal Reserve Bank of Kansas City, Jackson Hole, WY, August 25–27, pp. 399–474.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 87–104

Does the Federal Reserve Care about the Rest of the World?

Barry Eichengreen

Many economists are accustomed to thinking about Federal Reserve policy in terms of the institution's "dual mandate," which refers to price stability and high employment, and in which the exchange rate and other international variables matter only insofar as they influence inflation and the output gap—which is to say, not very much.

In fact, this conventional view is heavily shaped by the distinctive and peculiar circumstances of the last three decades, when the influence of international considerations on Fed policy has been limited. In this paper, I will discuss how the Federal Reserve paid significant attention to international considerations in its first two decades, followed by relative inattention to such factors in the two-plus decades that followed, then back to renewed attention to international aspects of monetary policy in the 1960s, before the recent period of benign neglect of the international dimension. This longer perspective is a reminder that, just because the Fed has not attached priority to international aspects of monetary policy in the recent past, we have no guarantee that it will not do so in the future. Indeed, I will argue in the conclusion that in the next few decades, international aspects are likely to play a larger role in Federal Reserve policy making.[1]

[1] The treatment here is necessarily summary in form. For readers interested in additional rigor and detail on the intersection of international issues and monetary policy, some starting points would be Galí and Gertler (2010), which is an especially useful compendium of recent scholarship; Friedman and Schwartz's (1963) *Monetary History of the United States*, which touches more than incidentally on the role

■ *Barry Eichengreen is the George C. Pardee and Helen N. Pardee Professor of Economics and Political Science, University of California, Berkeley, California. His email address is eichengr@econ.berkeley.edu.*

http://dx.doi.org/10.1257/jep.27.4.87 doi=10.1257/jep.27.4.87

International from the Start

The founding of the Federal Reserve in 1913 is commonly portrayed in terms of domestic financial stability considerations. Prior to 1914, financial crises were frequent. Interest rates spiked in the planting and harvest seasons, giving rise to financial stringency and instability. There was dissatisfaction with how market participants had managed the most recent crisis in 1907. The Fed was therefore created "to furnish an elastic currency . . . and for other purposes" in the words of the Federal Reserve Act of 1913. Importantly, the Act did nothing to change the international dimension of American monetary policy. The dollar was still convertible exclusively into gold at $20.67 a troy ounce, as it had been since the Gold Standard Act of 1900. Federal Reserve Banks were now obliged to hold gold in the amount of 40 percent of their notes (and gold and other eligible assets equal to 35 percent of deposits and reserves) and to pay out gold at this price.

But this is only part of the story. Political agreement to create a new central banking institution required building a coalition. In addition to those desiring a more elastic currency, there were exporters, importers, and financiers interested in establishing a market in dollar-denominated trade credits and enhancing the currency's role as a vehicle for international investment (Broz 1997). Attaining these goals required creating a central bank to provide liquidity to international markets. Before World War I, the US dollar and American financial firms played little role in financing international trade, including even the trade of US importers and exporters. A US coffee roaster seeking to import beans from Brazil would request a letter of credit from its bank and that bank in turn would arrange a letter of credit denominated in pounds sterling with its London correspondent because that was the only instrument that the Brazilian exporter would accept. Taking payment in US dollars was unattractive, given the volatility of US markets. In addition, because US banks were prohibited from branching abroad, exporters to US markets faced practical difficulties in converting US dollar payments back into their local currencies.

Thus, the New York financial community found itself unable to compete with London for an important source of business. US importers and exporters faced a competitive disadvantage from having to pay two commissions, one to their local bank and one to its London correspondent, in order to arrange trade credit. Paul Warburg, the German-born financier who was heavily involved in drawing up the blueprint that became the Federal Reserve Act, was familiar from his career in the import–export and banking business in Hamburg and London with the advantages that European economies derived from possessing local markets in "trade acceptances"—the contemporary name for what we now call trade credits.

---

of international factors in Fed decision making; and my personal favorite, William Adams Brown's (1940) *The International Gold Standard Reinterpreted*, which devotes successive chapters to the United States and its central bank.

These are essentially explicit documented promises by the purchaser of a good produced in another country to pay for a purchase at a future time.

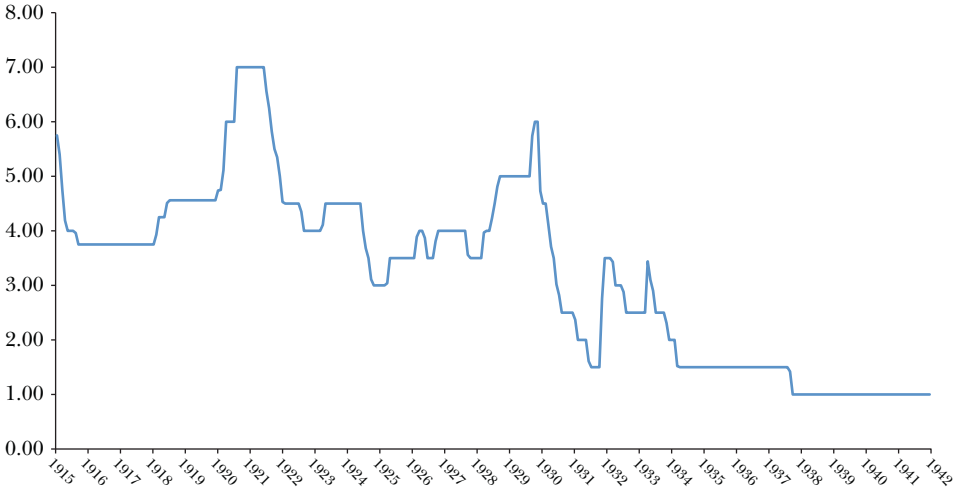
Agreement to establish the Fed required building an alliance between those with financial stability concerns and these other groups seeking to remedy the problems created by the absence of a trade-acceptance market. The Federal Reserve Act thus authorized US banks to branch abroad to originate foreign business. Moreover, one of the first initiatives of the new central bank was to take steps to foster a market in acceptances (LaRoche 1993; Eichengreen and Flandreau 2012). A key challenge in creating a new financial market is the problem of developing liquidity. Without a minimum level of transactions, the market will lack liquidity; conversely, if the market lacks liquidity, no one will transact. This was the chicken-and-egg problem that the Fed, seeking to foster a market in acceptances, faced in the 1920s. It responded by stepping in as buyer and liquidity provider of last resort, purchasing US dollar trade acceptances when private demand was lacking. For much of the 1920s, the Fed was the dominant purchaser. Its efforts succeeded in that New York and the dollar matched, and in some years surpassed, London and sterling as a source of credit for global trade. This was a startling change from before 1914.

In addition to underscoring the early Fed's international orientation, this episode had two further features relevant to modern central banking. First, a policy of direct Federal Reserve purchases in credit markets with liquidity problems—while a controversial aspect of recent policy—is not at all unprecedented. Second, the success of the 1920s was fleeting. When international trade declined in the 1930s, the market in trade acceptances declined even more rapidly. The other investors who the Fed had sought to attract by providing liquidity and stabilizing pricing never entered the market in large numbers. It is tempting to speculate that the Fed's overwhelming buy-side dominance crowded out potential entrants. When the central bank curtailed its involvement in the trade acceptance market of the 1930s, at a time when it had bigger fish to fry, the market collapsed.

The Federal Reserve Bank of New York was the single most active participant in the US dollar acceptance market, which is no surprise given that the bulk of acceptance business was transacted in New York. The New York Fed similarly took the lead on the new central bank's other international policy initiative, namely, reconstruction and maintenance of the international gold standard. Its leadership reflected the views of Benjamin Strong, the influential Governor of the Federal Reserve Bank of New York. Strong saw exchange rate stability—an important corollary of the gold standard—as critical for the expansion of international trade, and international trade as a key to US prosperity. Great Britain had traditionally been at the center of the gold standard system; only if it restored gold convertibility in the mid-1920s were other countries apt to follow.

Thus, in the spring and summer of 1924, the New York bank cut its discount rate (the interest rate it charged when advancing money to exporters and their banks by purchasing trade acceptances) by a cumulative 150 basis points, as shown

Figure 1
Discount Rate, Federal Reserve Bank of New York



Note: Data from NBER Macrohistory Database (series 13009).

in Figure 1, with the objective of making the pound sterling a relatively more attractive investment, inducing capital to flow from New York to London, and helping the Bank of England push sterling up to the pre-World War I exchange rate against the dollar, where it could then be stabilized (Clarke 1967).[2] To ensure a wider impact on financial markets, the New York Fed also purchased US Treasury securities, in the course of so doing, helping to establish the efficacy of open market operations as an instrument of monetary control. After importing gold for 51 consecutive months from December 1920 to April 1925, the United States now exported it instead. In January 1925, the Federal Reserve agreed to advance the British Treasury an additional $200 million in gold while encouraging a banking syndicate led by J.P. Morgan to provide a $100 million line of credit.

This policy was not an act of altruism: Strong firmly believed that helping Britain back onto the gold standard, by paving the way for a broader stabilization of exchange rates, would lend stimulus to US exports and economic growth. But that view was not universally shared within the Federal Reserve System. Strong's initiative was criticized, for example, by Adolph Miller, founding Governor of the Federal Reserve System and previously professor at the University of California at Berkeley. Miller argued that the resulting monetary policy was inappropriately loose for domestic circumstances. Along with others like then-Secretary of Commerce Herbert Hoover, Miller warned that low interest rates were fueling unsustainable real estate bubbles across Florida and

[2] At this time, individual Federal Reserve Banks were free to change their own discount rates, subject to the review and determination of the Federal Reserve Board.

from Detroit to Chicago. This was the first full-blown controversy within the Federal Reserve over the relative importance of domestic and international objectives.

The second controversy arose in 1927, when Strong again proposed cutting interest rates, this time in order to help Britain stay on the gold standard. Miller would have objected, but he was on summer holiday in California. When he returned, he mounted a strenuous attack on the policy as inappropriate for an economy already recovering from a brief recession. Monetary historians have been similarly critical (as described in Meltzer 2003), suggesting that a policy looser than appropriate from a domestic standpoint helped to fuel the commercial real estate boom and Wall Street run-up of the late 1920s, both of which came down with a crash. It would have been better, they conclude, for the Fed to keep its eye on the domestic ball.

The traditional constraint in which a central bank needs to hold interest rates high to attract capital inflows and defend the exchange rate then reemerged with a vengeance late in 1931. In a shock to financial markets, Britain departed from the gold standard on September 21, 1931. The dollar weakened against the continental European currencies, and gold losses mounted rapidly. In part, this reflected worries about US competitiveness as it became clear that some two dozen other countries were preparing to follow Britain in devaluing their currencies. Even more important was psychological contagion—the wake-up-call effect: if one reserve-currency country could depreciate its currency, it was no longer inconceivable that the other, the United States, might follow.

At this point the Fed made its priorities unambiguously clear. On October 8, 1931, the directors of the New York Fed voted to raise its discount rate by 100 basis points and then a week later by another 100 basis points (see Figure 1). Other Reserve Banks followed. These higher interest rates were designed to attract financial capital from abroad, or at least discourage it from leaving, thereby supporting the US dollar. The wisdom of the decision can be questioned, coming as it did in the midst of the Great Depression when domestic conditions warranted lower interest rates. But it clearly privileged exchange rate stability over price stability, financial stability, and economic stability.

The final attack on the dollar came in February–March 1933 in the interregnum between the Hoover and Roosevelt administrations. Worries that the new president might devalue—something that only he, together with the Congress, and not the Federal Reserve could decide—encouraged capital flight from the United States to France, Switzerland, and Belgium, countries seen as having stronger currencies (Wigmore 1987). The decision in February 1933 to let Henry Ford's Union Guardian Trust Company go under—the equivalent for its time of the September 2008 bankruptcy of Lehman Brothers—then ignited a nationwide banking panic (Kennedy 1973; Awalt 1969). At this point there was essentially no choice but to embargo gold exports, close the banks, and regroup. On his first day in office, President Franklin Roosevelt invoked the Trading with the Enemy Act for the necessary authority.[3]

[3] One cannot help but be reminded of UK Chancellor of the Exchequer Gordon Brown invoking the UK Anti-Terrorism Act to prevent the repatriation of Icelandic assets in 2008.

At this point, the first era in which international considerations played a prominent role in US monetary policy drew to a close. President Roosevelt took the next step in April 1933, making clear that there would be no early return to the gold standard. Currency would no longer be exchanged for gold, and individuals were required to turn in their gold and to receive currency in exchange. The Congress then passed a joint resolution canceling all contracts, public and private, that called for payment in gold. Starting in October, FDR used the authority of the Reconstruction Finance Corporation, an emergency agency created in 1932, to purchase gold with US dollars. Pushing up the dollar price of gold was equivalent to pushing down the exchange rate of the dollar against the currencies of other countries still on the gold standard. In effect, the Executive Branch had taken emergency control of monetary policy. Finally, in January 1934 the president agreed to stabilize the price of gold at $35 per ounce. This was not a formal return to the gold standard, since freedom for individuals to hold gold and the reintroduction of gold clauses into private and government contracts were not part of the bargain. Nonetheless, changes in the gold stock again translated into changes in the monetary base (the money supply narrowly defined)—with some important exceptions detailed below—because the government again bought gold for currency when it flowed in from abroad, to keep the US dollar's price from falling.

These steps inaugurated a new era in which international considerations played little role in US monetary policy. There was no need for high interest rates to stem capital outflows. The new higher dollar price of gold attracted the yellow metal toward the United States; more generally, devaluation enhanced the country's international competitiveness. In addition, as the outlines of World War II became visible, foreign capital fled in growing volumes to American shores. Foreign economic and financial conditions mattered less for the US economy than in the 1920s, now that global trade had collapsed due to the imposition of higher tariffs, and long-term international investment had been discouraged by exchange controls and sovereign debt defaults.

On the Horns of the Triffin Dilemma

After President Roosevelt stabilized the price of gold at $35 an ounce, the Federal Reserve and the US Treasury jousted over who would control monetary policy. Standard practice was for Treasury to purchase all gold inflows using funds in its account at the Federal Reserve, print "gold certificates" in the same dollar amount, and deposit those certificates with the Fed, where they could back an increase in the supply of money and credit. But on some occasions, notably in 1937, when it grew worried about inflation, Treasury would place its gold certificates in an "inactive" account where they could not be used to expand the money supply. This practice was known as "gold sterilization." The Fed might still seek to influence the money supply by changing the discount rate or reserve requirements for the banks. But Treasury's sterilization policy could frustrate or, as in 1937, undesirably amplify the effects (Irwin 2012).

Confronted with a US Treasury that was seeking to carry out monetary policy in this way, the Fed sought to regain the ability to influence money and credit markets. Its campaign was unsuccessful: with the outbreak of World War II and the very large government deficits of that time, the Fed was drafted into keeping federal borrowing costs low by holding interest rates on Treasury bills at 0.375 percent and Treasury bonds at 2.5 percent. The practice continued, despite growing Fed resistance, for two years following World War II. Current controversies over how a change in Federal Reserve policy would affect federal government borrowing costs in some ways echo this earlier situation.

This long period of fiscal dominance over monetary policy eventually came to an end with the Treasury–Fed Accord of 1951, which officially ended the expectation that the Fed would purchase US Treasury securities in whatever quantities were necessary to keep interest rates at these low levels. Recent scholarship portrays monetary policy over the balance of the 1950s in a relatively favorable light (for example, Romer and Romer 2002a). More pertinent from the standpoint of this paper, monetary policy at this time focused on keeping inflation low and, to a lesser extent, on responding to temporary deviations from full employment. The Minutes of the Federal Open Market Committee in this period (available online at http://fraser.stlouisfed.org/publication/?pid=677) offer little emphasis or even mention of the US dollar exchange rate, the US balance of payments, or the effect of US monetary policy on the rest of the world. Of course, the discussions do mention exports and imports, since these variables were seen by members of the committee as containing information useful for forecasting the future paths of inflation and what would later be called the "output gap." But beyond such comments, international factors do not appear to have impinged on the committee's deliberations.

There was a commitment after the Bretton Woods agreement of 1944 to continue stabilizing the price of gold at $35 an ounce and pay out gold on demand to official foreign creditors, but no matter. This was the period of the "dollar shortage." The term refers to the difficulty experienced by other countries in acquiring, whether through exporting or foreign borrowing, the dollars they needed in order to finance merchandise imports from the United States (Kindleberger 1950). At this time, gold held by the US monetary authorities far exceeded the foreign liabilities of the Federal Reserve, US commercial banks, and the US government combined. In this sense, the orientation of monetary policy was not constrained by international implications.

This situation began to change around 1960, when US foreign monetary liabilities first threatened to exceed US gold reserves. Investors worried that John F. Kennedy, if elected president, might follow in Roosevelt's footsteps and devalue the dollar to get the economy "moving again" (to quote Kennedy's campaign literature). Robert Triffin (1960) published the first of a series of books in which he warned that if the dollar remained the only source of global liquidity (other than gold, which still lurked behind the scenes at the agreed-upon price of $35 per ounce), a crisis of confidence in the greenback would ultimately develop. Triffin's

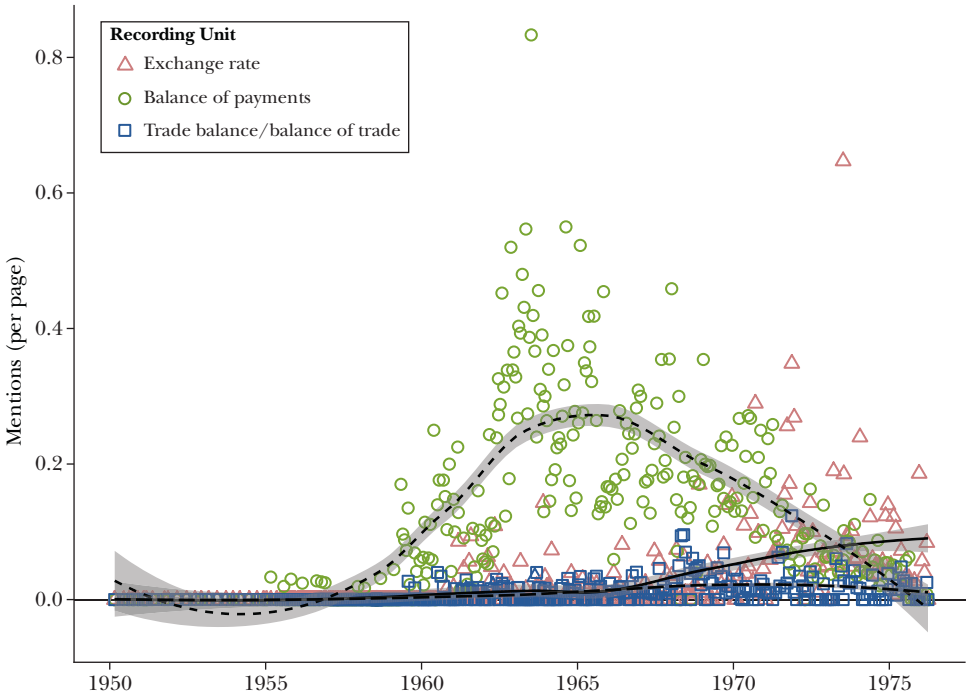
dilemma was based on the insight that the expansion of the global economy would bring growing demands for international liquidity. If dollar-denominated claims, and specifically US Treasury bonds, were the primary source of such liquidity (on the margin), then US foreign liabilities would eventually come to exceed US gold reserves. When this occurred, it would call into question the ability of the US authorities to convert these liabilities into gold at the fixed price of $35 an ounce, creating a crisis of confidence. Alternatively, if the authorities took steps to limit US current account deficits and foreign lending, the rest of the world would be starved of liquidity in an ongoing "dollar shortage," and international transactions generally would suffer.

The expectation that President Kennedy would devalue the US dollar proved erroneous, but the other worries were not without foundation. Inflation accelerated in the later 1960s and grew more erratic. The goals of Federal Reserve policy shifted from an overarching emphasis on inflation to greater attention to unemployment and economic growth. Romer and Romer (2002b) argue that this period saw a revolution in ideas in which policymakers forgot much of what they had learned about the natural rate of unemployment. Instead, they overestimated potential output and succumbed to the temptation to use monetary policy to target real variables. Federal Reserve Chairman William McChesney Martin believed that the Fed had an obligation to help keep federal debt service at manageable levels, which constrained monetary policy in the direction of lower interest rates as budget deficits grew.

Even if the Fed was concerned about its gold losses and other international variables, it might nonetheless be hard to detect that concern amongst these other changes. However, in Bordo and Eichengreen (2008), my coauthor and I argue that the Fed paid considerable attention to balance-of-payments considerations in the first half of the 1960s, tightening when it grew worried by the pace of gold outflows. In addition to his concern with debt service, Fed Chairman Martin was a firm believer in maintenance of the gold peg. Already in 1960, the Fed abandoned its traditional "bills-only policy" (the policy of buying only short-term Treasury debt) in order to let short-term interest rates rise, attract capital flows, and strengthen the balance of payments (Solomon 1977, p. 36). The Minutes of the Federal Open Market Committee (FOMC) regularly refer to balance-of-payments considerations. Many of these statements, in an echo of the 1920s, came from the President of the Federal Reserve Bank of New York, Alfred Hayes (Meltzer 2013). A count of references in the minutes and memoranda of the FOMC, as available from 1950 through March 1976, normalized by pages, shows mentions (and presumably concern) relating to the balance of payments to be mounting in the first half of the 1960s and peaking around mid-decade, as shown in Figure 2.[4]

[4] Where dots are missing in Figure 2 (as in the first half of the 1950s), there were zero mentions. Normalizing by pages adjusts for the fact that the minutes and memoranda tended to grow longer over time, although raw counts show basically the same picture. In principle, it should be possible to extend this

Figure 2
References to Balance of Payments and Related Terms in the Minutes



Notes: Mentions for each term are taken from minutes and memoranda of discussion for all meetings and telephone conferences of the Federal Open Market Committee from 1950 through March 1976. Data are fit with a 2nd degree local polynomial LOESS regression with span parameter $\alpha = 0.75$ (indicating that 75 of the data are used to estimate each local regression) and the bands represent 95 percent confidence intervals.

Of course, mentioning international issues in the minutes is not the same as taking action with these issues in mind, and so in Bordo and Eichengreen (2008), we attempt to identify the role of those mentions (and that concern) in the policy decisions of the Federal Open Market Committee. We identify seven occasions when policy action was motivated primarily by international considerations, and 23 when it was motivated by a combination of domestic and international factors. That is, only

analysis beyond 1976 when transcripts of Federal Open Market Committee meetings become available. However, the transcripts are sharply discontinuous with the minutes in terms of comprehensiveness; in addition, pagination and font size are quite different, and the pagination and format of the transcripts themselves are not constant over time. This makes trends in both raw and per-page counts more difficult to interpret. I have resisted the temptation. The outlier in mid-1963 is from a meeting in a period of heightened concern about dollar stability (Eichengreen 2000). The Federal Reserve system had recently drawn its full $150 million swap line with the Bundesbank, and dollar weakness had been a prominent topic at the most recent monthly meeting of the Bank for International Settlements, where the Fed had been represented by Charles Coombs of the Federal Reserve Bank of New York.

rarely were international factors an overriding consideration in Fed decisions, but they generally combined with domestic factors to prompt policy action. The total of 30 such instances suggests that international considerations were not inconsequential. The majority of these instances were in the period through 1965 (although international considerations were also invoked during crisis episodes in 1967–68 and 1971). These international concerns help to explain why monetary policy in the first half of the 1960s was tighter than would be expected on the basis of inflation and the output gap alone (Bordo and Eichengreen 2008; Taylor 1999).[5]

What changed between the first and second halves of the 1960s that made international considerations less salient for Federal Reserve policymakers? In the first half of the decade, balance-of-payments management had effectively been a shared responsibility of the US Treasury and the Fed. But as the Treasury pushed for adoption of an "interest equalization tax"—a tax on purchases of foreign securities by US investors—and took a variety of other policy measures designed to strengthen the balance of payments, it assumed primary responsibility for balance-of-payments management. In addition, measures like the interest equalization tax had effects tantamount to capital-flow taxes and controls. They gave the central bank some room for looser monetary policy with less concern that it would encourage an outflow of investment capital. Whether these steps were good policy is debatable (Meltzer 1991 offers a critical assessment). The acceleration of inflation and mounting political pressure on Chairman Arthur Burns for loose monetary policy starting in 1970 suggests that there would have been benefits to the Federal Reserve if it had continued to perceive itself as under the external constraint under which it operated in the first half of the 1960s.

After Bretton Woods

The 1970s was a decade of mixed signals and uncertainty about Federal Reserve policy. Unconstrained by the exchange rate, or for that matter much else, monetary policy drifted. This changed in 1977 when Congress amended the Federal Reserve Act to include the "dual mandate" that monetary policy should consider both stable prices and maximum employment, and in 1979 when Paul Volcker succeeded G. William Miller as Chairman of the Federal Reserve and made inflation control a priority. The touchstones of policy became deviations of inflation from low single digits and fluctuations in the output gap.[6] One finds periodic mention of international considerations in the minutes and transcripts of the Federal Open Market Committee of this time, but it is clear that these variables

[5] Romer and Romer (2002b, p. 57) similarly make the point that balance-of-payments considerations prevented the Fed from being as expansionary as it would have otherwise wanted in the first half of the 1960s.

[6] Indeed, the observation of how the federal funds interest rate changed with changes in inflation and unemployment after 1983 is what led the eponymous Professor Taylor to develop his rule.

mattered principally insofar as they were relevant to the future evolution of inflation and the output gap.

A combination of factors explains why international factors were less influential in the late 1970s and early 1980s than in the 1920s, 1930s, and 1960s. After the final collapse of the Bretton Woods agreement in 1973, there was no longer an exchange rate peg or statutory gold price to defend. The 1977 legislation gave the Fed a mandate to pursue price stability and full employment but said nothing about the exchange rate, balance of payments, or international financial stability. The experience of the 1970s had taught that a Federal Reserve that failed to achieve price stability would lack the credibility to successfully pursue other economic and financial goals. The US economy was large and closed enough that the Fed could afford to act to a first approximation like the central bank of a closed economy. The US share of world GDP peaked in 1985 at 33 percent, at a time when the Soviet economy was in decline and China's growth spurt had just begun.[7] The US trade/GDP ratio was rising, but more slowly than in the subsequent quarter century. The explosive growth of international capital flows and deepening of international financial linkages was yet to come.

International considerations did in some circumstances play a role in monetary policy during this time. Volcker's inflation-control strategy itself had an international dimension; the fact that higher interest rates meant a stronger dollar made for sharper disinflation through the channel of lower import prices (as argued by Sachs 1985).[8] The Fed's decision to back away from a very tight monetary policy in 1982 may have been influenced by the outbreak of the Latin American debt crisis and the threat this posed to the solvency of major US banks (which is not to deny that there was also an influence toward looser monetary policy from the US recession in 1981–82). Central bank governors as well as finance ministers were party to the Plaza Agreement of 1985 between the United States, the United Kingdom, France, Germany, and Japan to attempt to stem the continued rise of the dollar exchange rate. This led to coordinated foreign exchange market intervention and, in March 1986, coordinated interest rate reductions. Starting in 1986, central bank governors of the so-called G-7 countries (the United States, United Kingdom, France, Germany, Italy, Canada, and Japan) met regularly, together with their finance ministry counterparts, on the sidelines of the spring and fall meetings of the IMF and World Bank. In addition, senior central bank officials met bimonthly at the Bank of International Settlements. These meetings facilitated information exchange. They also facilitated coordinated foreign-exchange-market intervention, frequently before the mid-1990s and sporadically thereafter: in

[7] One should be cautious about these comparisons, because they depend on the exchange rate used to value transactions in dollars, and 1985 was when the dollar exchange rate was at a local peak. At purchasing power parity, the US share of the global economy was more like 23 percent in 1985 and reaches another local peak in 1999. Economists will of course debate which valuation method is more relevant when considering the conduct of monetary policy.

[8] Nelson (2005) argues that the Federal Open Market Committee had something similar in mind when it tightened in 1978.

June 1998 when the yen depreciated in the wake of the Asian crisis; in September 2000 when the euro weakened reflecting uncertainty about the policies of Europe's new central bank; and in March 2011 in response to the rise of the yen induced by the Fukushima earthquake and the liquidation of foreign assets by Japanese insurance companies.[9]

The global economic and financial crisis that unfurled in 2007 is another reminder that the Fed cannot afford to neglect the impact of its policies on conditions abroad or the implications of conditions abroad for its policies. On October 8, 2008, in the wake of the Lehman Brothers failure, the Federal Reserve coordinated a reduction in the federal funds rate with the lending rates of the European Central Bank, Bank of England, Bank of Canada, Swiss National Bank, and Swedish Riksbank. Irwin (2013), with a little exaggeration, calls this the "first globally coordinated easing in history." Unusually, the Fed issued a joint statement together with these other central banks announcing the action, which I interpret as an acknowledgment that coordinating policy with foreign central banks might produce better outcomes under the circumstances.

In addition, the Fed arranged dollar swap lines with 14 foreign central banks starting in December 2007, when the subprime crisis intensified. In these arrangements, the Federal Reserve stands ready to swap US dollars with other central banks for the currencies of that bank. The Fed renewed five of those dollar swap lines, notably that with the European Central Bank, in May 2010. These swap facilities were designed to alleviate financial problems abroad and limit the blowback to US markets if foreign banks, unable to secure US dollar funding, were forced to liquidate their holdings of US financial securities. These swap lines were an acknowledgement that what happens abroad doesn't always stay abroad and, while not modifying monetary policy to take this fact into account, that the Fed must develop ancillary policy instruments to address strains in foreign markets for US dollars. The Board of Governors (2013), in justifying the practice to a skeptical Congress, noted that foreign currency swap lines might also be helpful for addressing financial strains should US institutions experience a shortage of foreign currency-denominated liquidity, although the most recent swaps were not activated for this purpose.

Back to the Future

The questions are whether international considerations will have a more powerful impact on the US economy in the future and how, if at all, the Federal Reserve should modify the formulation and conduct of monetary policy to take

[9] To be clear, these intervention operations are decided in consultation by the US Treasury and the Federal Reserve. Such operations are typically sterilized with the goal, sometimes questioned by academics, of moving the dollar exchange rate without also moving the monetary base. On the effectiveness of sterilized intervention, a useful starting point is Rogoff (1984).

this into account. I posit three trends that will heighten the impact of international variables on the US economy.

First, I assume that the United States will continue to grow more open to international trade and financial transactions. Admittedly, this assumption is contestable. While technological progress works inexorably to reduce the costs of international transactions, and—I would argue—also reduces the cost of international relative to domestic transactions, international openness depends not just on technology but also on policy, which has been known to push in the other direction. But bear with me.

Second, I assume that emerging and frontier markets will continue to grow more rapidly than mature economies, so that the US economy will come to account for a progressively declining share of the global economy. Again, continuing catch-up and convergence are not inevitable; their progress will depend on policy choices. But recent experience makes this assumption a reasonable starting point.

Third, I assume that the US dollar will lose its monopoly as a funding currency for international banks and as the all but exclusive vehicle and currency of denomination for international transactions (for discussion, see Shin 2011). This is not to suggest, as does the film *Looper*, that we will wake up tomorrow and discover that all transactions are conducted in renminbi.[10] Movement toward other funding and vehicle currencies will be gradual, so that the dollar ends up sharing its international role with other national units. But there is no fundamental reason why the United States should be the only country with deep and liquid financial markets open to the rest of the world. Moreover, the US economy alone will not be able to provide safe and liquid assets on the scale required by an expanding global economy. It follows that US banks and firms will rely more on foreign currency funding and liquidity in the future than the recent past (Eichengreen 2011).

Taken together, these three trends suggest that shocks to the exchange rate and balance of payments will have a larger impact on the US economy in the future, and that the implications may extend beyond those for inflation and the output gap. For example, US dollar appreciation which creates competitiveness problems for the traded-goods sector will be more of a problem as a larger share of US output and employment is exposed to international competition. If dollar appreciation causes US firms to exit the market and they face fixed costs of reentering (as in Baldwin and Krugman 1989), then transitory currency swings may have permanent welfare-reducing effects. This is one explanation for why other open economies adjust their policies in response to movements in the exchange rate. It is an explanation for the aversion to freely floating exchange rates, known as "fear of floating," in emerging markets (Calvo and Reinhart 2002).

[10] *Looper*, as film buffs know, is set in 2044. In the original script, the protagonist planned to move to Paris "in the future." When the director found filming in Paris prohibitively expensive, the future was shifted to Shanghai, the Chinese distributor having offered to pay for a crew to film there—see http://www.imdb.com/title/tt1276104/trivia.

A country on the receiving end of large capital inflows, in addition to experiencing a rising exchange rate, is likely to feel unwelcome pressure on housing and financial markets. Capital inflows into the US economy in the period leading up to the subprime crisis are an illustration of the problems that can arise. The effect on local markets will be larger the smaller the economy is relative to the rest of the world and the more open it is to global markets. In mid-2013, New Zealand is a case in point of a country that is dealing with these kinds of housing and asset market concerns due to exchange-rate and capital-flow pressures (Wheeler 2013).[11] More generally, a variety of small open economies, and a number of middle-sized countries like Brazil, have complained about the adverse impact of policies abroad on their economies, operating through these channels, and have adjusted monetary and other policies to address them: for example, Brazilian Finance Minister Guido Mantega has registered strong complaints along these lines in a series of press interviews and speeches starting in September 2010.

Finally, as global markets grow relative to the US market, and as international finance is provided in a wider range of currencies, US banks and firms will rely more on foreign currency funding. As they accumulate liabilities denominated in foreign currency, the Federal Reserve may then feel a growing reluctance to let the dollar exchange rate move for fear of the destabilizing balance sheet effects (specifically, that banks and firms with foreign-currency-denominated debts will be unable to service them using their domestic-currency earnings). Those adverse balance sheet effects are another explanation for why smaller, more open economies where such currency mismatches are prevalent find it hard to commit to regimes of flexible inflation targeting that imply benign neglect of the exchange rate.

Assume, as a result of the changes posited above, that the effect of the exchange rate and capital flows grows more important. Does it follow that the Fed will have to modify the formulation and conduct of monetary policy to take them into account?

The answer, as with many economic questions, is "yes and no." For example, if a higher level of reliance on foreign currency funding causes exchange rate movements to have more destabilizing balance sheet effects, then the first-best response is not to use monetary policy to prevent those movements, but instead to strengthen prudential supervision and regulation of banks and governance of corporations to prevent excessive exposure to this form of balance sheet risk from arising in the first place. Mishkin and Savastano (2001) is an early statement of the tradeoff between, on the one hand, the strength of supervision and regulation of balance-sheet mismatches, and on the other hand, policies of benign neglect of the exchange rate.

Similarly, if capital inflows place worrisome upward pressure on housing and other asset markets, then the first-best solution is to strengthen lending standards,

[11] It is tempting to point to Ireland and Spain before 2009 as additional examples, but their cases are special for obvious (euro-related) reasons.

raise margin requirements, and otherwise address problems in housing and asset markets directly. The second-best set of responses in this case would be to address the capital inflows that are the proximate source of the problem by applying taxes to such inflows and, among other steps, tightening fiscal policy, which should put downward pressure on interest rates and on the exchange rate. Trying to address problems in asset markets caused by inflows of foreign capital by making adjustments in monetary policy would be no more than a third-best policy choice.

As yet another example, if the issue is the risk of permanent damage to traded-goods sectors because temporary exchange rate movements have permanent effects, then the first-best response is to eliminate the financial imperfections forcing incumbents to exit and preventing them from reentering, or to use tax and other policies (assuming that these can be enacted in a form compliant with World Trade Organization rules) to address their problems directly.

Readers will detect echoes here of the "lean versus clean debate" (White 2009; Mishkin 2011). The question in that context was whether central banks should lean against asset bubbles. Earlier thinking tended to favor letting other agencies of government—bank supervisors, regulators, those responsible for the conduct of fiscal policy—address problems of asset bubbles using instruments better suited to the task than monetary policy, and for central banks to limit their intervention to cleaning up the aftereffects. It would be presumptuous to assert that recent events have permanently decided the question in favor of one view or the other. But there is no question that those events have shifted the balance by suggesting that central banks should think harder about the need to take preemptive action, both because other agencies of government may not be doing their part and because cleaning up afterwards can be very costly.

The implication is that precisely the same issues will arise, with growing intensity over time, in connection with movements in exchange rates and capital flows. There will be no return to the gold standard of the 1920s or the Bretton Woods System of the 1950s and 1960s. But as the US economy grows even more open and the rest of the world grows larger, international considerations, operating through these channels, will impinge more directly on the central bank's key objectives of price and economic stability. The Fed will have no choice but to incorporate those considerations more prominently into its policy decisions.

■ *The first draft of this paper was prepared for the "Symposium on the First 100 Years of the Federal Reserve," held at the National Bureau of Economic Research in Cambridge, Massachusetts, on July 10, 2013. I thank Christina Romer, David Romer, and the editors of this journal for comments and Chris Krogslund for research assistance. The National Bureau of Economic Research provided an honorarium in support of this paper. Other than this the author has received no significant financial support from interested parties. He serves on the Centennial Advisory Council of the Board of Governors of the Federal Reserve System.*

References

Awalt, Francis Gloyd. 1969. "Recollections of the Banking Crisis in 1933." *Business History Review* 43(3): 347–71.

Baldwin, Richard, and Paul Krugman. 1989. "Persistent Trade Effects of Large Exchange Rate Shocks." *Quarterly Journal of Economics* 104(4): 635–54.

Board of Governors of the Federal Reserve System. 2013. "Central Bank Liquidity Swap Lines." Washington, DC: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/newsevents/reform_swaplines.htm.

Bordo, Michael, and Barry Eichengreen. 2008. "Bretton Woods and the Great Inflation." NBER Working Paper 14532, December.

Brown, William Adams, Jr. 1940. *The International Gold Standard Reinterpreted, 1914–1934.* New York: National Bureau of Economic Research.

Broz, J. Lawrence. 1997. *International Origins of the Federal Reserve System.* Ithaca, NY: Cornell University Press.

Calvo, Guillermo A., and Carmen M. Reinhart. 2002. "Fear of Floating." *Quarterly Journal of Economics* 117(2): 379–408.

Clarke, Stephen V. O. 1967. *Central Bank Cooperation, 1925–31.* Federal Reserve Bank of New York.

Eichengreen, Barry. 2000. "From Benign Neglect to Malignant Preoccupation: U.S. Balance of Payments Policy in the 1960s." In *Economic Events, Ideas and Policies: The 1960s and After*, edited by George Perry and James Tobin, 185–242. Washington, DC: Brookings Institution.

Eichengreen, Barry. 2011. *Exorbitant Privilege: The Rise and Fall of the Dollar and the Future of the International Monetary System.* New York: Oxford University Press.

Eichengreen, Barry, and Marc Flandreau. 2012. "The Federal Reserve, the Bank of England, and the Rise of the Dollar as an International Currency, 1914–1939." *Open Economies Review* 23(1): 57–87.

Federal Open Market Committee. 1936–1994. Minutes of the Federal Open Market Committee. Available at Fraser Federal Reserve Archive. http://fraser.stlouisfed.org/publication/?pid=677.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960.* Princeton University Press for the National Bureau of Economic Research.

Galí, Jordi, and Mark Gertler, eds. 2010. *International Dimensions of Monetary Policy.* University of Chicago Press for the National Bureau of Economic Research.

Irwin, Douglas. 2012. "Gold Sterilization and the Recession of 1937–8." *Financial History Review* 19(3): 249–267.

Irwin, Neil. 2013. *The Alchemists: Three Central Bankers and a World on Fire.* New York: Penguin.

Kennedy, Susan Estabrook. 1973. *The Banking Crisis of 1933.* Lexington: University Press of Kentucky.

Kettl, Donald. 1986. *Leadership at the Fed.* New Haven: Yale University Press.

Kindleberger, Charles P. 1950. *The Dollar Shortage.* Cambridge: MIT Press.

LaRoche, Robert K. 1993. "Bankers Acceptances." *Economic Quarterly,* Federal Reserve Bank of Richmond, 79(1): 75–85.

Meltzer, Allan H. 1991. "U.S. Policy in the Bretton Woods Era." *Federal Reserve Bank of St. Louis Review* 73(3): 54–83.

Meltzer, Allan H. 2003. *A History of the Federal Reserve,* Vol. 1: *1913–1951.* University of Chicago Press.

Meltzer, Allan H. 2013. Comment on "Bretton Woods and the Great Inflation," by Michael D. Bordo and Barry Eichengreen, in *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides, 489–93. University of Chicago Press.

Mishkin, Frederic S., and Miguel A. Savastano. 2001. "Monetary Policy Strategies for Latin America." *Journal of Development Economics* 66(2): 415–44.

Mishkin, Frederic S. 2011. "Monetary Policy Strategy: Lessons from the Crisis." NBER Working Paper 16755, February.

Nelson, Edward. 2005. "The Great Inflation of the Seventies: What Really Happened?" *B. E. Journal of Macroeconomics* 5(1): 1–50.

Rogoff, Kenneth. 1984. "On the Effects of Sterilized Intervention: An Analysis of Weekly Data." *Journal of Monetary Economics* 14(2): 133–50.

Romer, Christina D., and David H. Romer. 2002a. "A Rehabilitation of Monetary Policy in the 1950s." *American Economic Review* 92(2): 121–27.

Romer, Christina D., and David H. Romer. 2002b. "The Evolution of Economic Understanding and Postwar Stabilization Policy." In *Rethinking Stabilization Policy*, 11–78. Federal Reserve Bank of Kansas City.

Sachs, Jeffrey D. 1985. "The Dollar and the Policy Mix, 1985." *Brookings Papers on Economic Activity* 16(1): 117–85.

Shin, Hyun Song. 2011. "Global Banking Glut and Loan Risk Premium." Paper presented at the 12th Jacques Polak Annual Research Conference,

November 10–11. http://www.imf.org/external/np/res/seminars/2011/arc/pdf/hss.pdf.

Solomon, Robert. 1977. *The International Monetary System 1945–1976: An Insider's View*. New York: Harper & Row.

Taylor, John B. 1999. "A Historical Analysis of Monetary Policy Rules." In *Monetary Policy Rules*, edited by John Taylor, 319–348. University of Chicago Press.

Triffin, Robert. 1960. *Gold and the Dollar Crisis*. New Haven: Yale University Press.

Wheeler, Graeme. 2013. "Manufacturing Decline Not Just a Dollar Story." Speech delivered to the New Zealand Manufacturers and Exporters Association, Auckland, on February 20. http://www.rbnz.govt.nz/research_and_publications/speeches/2013/5150125.html.

White, William. 2009. "Should Monetary Policy 'Lean or Clean?'" Globalization and Monetary Policy Institute Working Paper 34, Federal Reserve Bank of Dallas, August.

Wigmore, Barrie A. 1987. "Was the Bank Holiday of 1933 Caused by a Run on the Dollar?" *Journal of Economic History* 47(3): 739–55.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 105–120

An Interview with Paul Volcker

Martin Feldstein

Martin Feldstein interviewed Paul Volcker in Cambridge, Massachusetts, on July 10, 2013, as part of a conference at the National Bureau of Economic Research on "The First 100 Years of the Federal Reserve: The Policy Record, Lessons Learned, and Prospects for the Future." Volcker was Chairman of the Board of Governors of the Federal Reserve System from 1979 through 1987. Before that, he served stints as President of the Federal Reserve Bank of New York from 1975 to 1979, as Deputy Undersecretary for International Affairs in the US Department of the Treasury from 1969 to 1974, as Deputy Undersecretary for Monetary Affairs in the Treasury from 1963–65, and as an economist at the Federal Reserve Bank of New York from 1952 to 1957. During the interludes from public service, he held various positions at Chase Manhattan Bank. He has led and served on a wide array of commissions, including chairing the President's Economic Recovery Advisory Board from its inception in 2009 through 2011.

Ending Gold Convertibility

FELDSTEIN: Let me start with your experience at the Treasury department in the early 1970s. President Nixon suspended gold convertibility in 1971, and that led to the collapse of the Bretton Woods arrangement. I have three questions about that. First, what was your view at the time of the desirability of that policy? Second, what is your view in retrospect? Did the United States have any choice? And finally,

■ *Martin Feldstein is George F. Baker Professor of Economics, Harvard University, and President Emeritus, National Bureau of Economic Research, both in Cambridge, Massachusetts.*

http://dx.doi.org/10.1257/jep.27.4.105 doi=10.1257/jep.27.4.105

how much do you think that action contributed to the sharp rise in inflation in the remainder of the decade?

VOLCKER: I certainly was a major proponent of suspending gold convertibility, in fact the principal planner. I had come to the conclusion that we needed to negotiate a sizable exchange rate adjustment. At the time, we didn't have a choice, as I saw it, to suspending convertibility as a transition to a reformed system. In the end, the Smithsonian Agreement[1] was not a very reasonable outcome from my point of view because I didn't think the changes in exchange rates were big enough to instill confidence. The United States didn't accept any responsibility itself by the way of any kind of convertibility to support the new rates, so the foundation wasn't there for a long-lasting solution in my view.

But my thought always was we suspend convertibility, get the necessary exchange rate change, and then we would redesign the international monetary system. President Nixon had no interest in redesigning the international monetary system, I think it's fair to say. Nor did Mr. Connally[2] have much interest, which led to, I think, a more unsatisfactory situation internationally where it was easy to make the impression that we were being irresponsible.[3]

FELDSTEIN: So given that you didn't get the second part of what you were hoping would happen, does it still look like the right decision in retrospect? Or, as you say, there was really no choice? You couldn't negotiate something different?

VOLCKER: I think it was the right decision. The question was what happened afterwards. The whole international exchange rate situation got out of hand. Federal Reserve policy, in my view, was not particularly credible at that point. I was in the Treasury. Every time I went on a trip abroad, I would try to get to [Federal Reserve Chairman] Arthur [Burns] to say, "Don't ease up while I'm abroad anyway."

The inflation took hold, as you know, and never was really brought under control for some time. It was an unsatisfactory economic situation.

FELDSTEIN: Some people think that that was related to the fact that we had gone off the gold convertibility. How much do you think that actually contributed to the inflation that happened in the remainder of the decade?

VOLCKER: I didn't think we had any choice about the gold convertibility. I suppose, you know, we could have devalued, as we did eventually, and then try to defend a new rate, but we didn't have enough credibility to do that. I think we didn't accept any kind of constraint during that period, the early 1970s. The exchange rate fell abruptly at one point or several points. We never had a reaction until late in the decade. I think there was a lack of discipline that would have been useful at the time, which was not recognized and wasn't acted upon.

[1] At the Smithsonian in December 1971, representatives of ten countries—Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, the United Kingdom, and the United States—established a new set of fixed exchange rates.

[2] John Connally was US Secretary of the Treasury from 1971 to 1972.

[3] Silber (2012), in his biography of Paul Volcker, covers this more fully.

Paul Volcker

October 1979

FELDSTEIN: Let me move to a later time. When you moved from the New York Fed in 1979 to become Chairman of the Board of Governors, consumer prices by then were rising at more than 10 percent a year. You decided to push short-term rates up even higher than the rate of inflation.

VOLCKER: The market pushed those rates up.

FELDSTEIN: Yes. Absolutely. With a little facilitation from folks on Constitution Avenue.[4] Anyway, that succeeded in bringing inflation down to 6 percent in 1982, and 3 percent in 1983. I've got several questions about all of that. First of all, why do you think inflation had gotten so high at the end of the 1970s?

VOLCKER: I guess I have to say the Federal Reserve policy contributed to it, but there was the oil crisis in the early '70s and then repeated again in the later '70s. It made a profound impression on me, if nobody else, that Arthur Burns titled his valedictory speech "The Anguish of Central Banking" (Burns 1979). That was a long lament about how, in the economic and political setting of the times, the Federal Reserve, and by extension presumably any central bank, could not exercise enough restraint to keep inflation under control. It was a pretty sad story. If you were going

[4] The headquarters of the Federal Reserve Board of Governors is on Constitution Avenue in Washington, DC.

to follow that line, you were going to give up, I guess. I didn't think you could give up. If I was in that job, that was the challenge as the Chairman of the Federal Reserve. You inherit a certain challenge.

FELDSTEIN: Some people thought at the time that it was just going to be too costly in terms of lost GDP and higher unemployment to get from where inflation was at that point down to low single digits. I think there was a recommendation, at least among some academic economists, to stop further increases in inflation, but not to try to bring inflation down because that would be so costly.

VOLCKER: The favorite word at the time, which was very popular within the Federal Reserve, but I think popular in the academic community generally, was "gradualism." I don't quite remember them saying, "Don't bring it down at all." But instead, it was "Take it easy. It will be a job of, I don't know, years, decades, whatever, and you can do it without hurting the economy." I never thought that was realistic.

The inflationary process itself brought so many dislocations, and stresses and strains that you were going to have a recession sooner or later. The idea that this was just going to go on indefinitely, and the inflation rate got up to 15 percent, it was going to be 20 percent the next year.

One little story (I think of all these stories): Shortly after we began the disinflation, somebody, I think Arthur Levitt who was the head of the American Stock Exchange, brought in some businessmen—they tend to be small businessmen—to talk to me at the Federal Reserve. I had them for lunch, and I gave them my little patter about, "This is going to be tough, but we're going to stick with it, and the inflation rate is going to come down," and so forth. The first guy that responded said, "That's all very fine, Mr. Volcker, but I just came from a labor negotiation in which I agreed to a 13 percent wage increase for the next three years for my employees. I'm very happy with my settlement." I always wondered whether he was very happy two years later on. But that was symbolic of the depths. He was happy at a 13 percent wage increase.

FELDSTEIN: Did you have a sense of what the "cost" would be—that is how high was unemployment going to be? How long was this going to last?

VOLCKER: I had no sense that interest rates were going to be so high. In October 1979, we took the full panoply of restrictive measures and emphasized the money supply, and so forth. I thought this was all done to convince people we were really serious. We had already raised the discount rate two or three times in the space of two or three months. The last increase in the discount rate was by a 4 to 3 vote. I was a neophyte Federal Reserve Chairman. It was all right with me. I knew I had four votes. If we had to raise it again I'd still have four votes. That's not the way the market interpreted it. They said, "Ah, they're at the end of their string. They can't command the unified majority any more on the Board. So it's the end of the day for any Federal Reserve restraint."

I decided we had to change the playbook a little bit, and we threw everything we could into the October 1979 announcement. I had this naive hope. I knew the short-term rates would go up, but I thought, "Ah, we will instill confidence and

long-term rates will not go up." Long-term rates went up, too, just about as much as the short-term rates, which was a disappointment. But it showed how strong the psychology was.

FELDSTEIN: The inflation came down, I think, much faster than outsiders, in any case, expected. Remember the tax cuts that were put in place in 1981 were based on the idea that inflation would continue to increase tax revenue because of bracket creep. We didn't start indexing income tax brackets for inflation until 1985. Yet inflation, as I quoted a minute ago, was down to 6 percent in 1982 and half that the next year. Were you surprised at how fast the process worked?

VOLCKER: I don't know that I was surprised, but I sure was relieved. I didn't know how long that could have gone on.

FELDSTEIN: Can you say a little more about what it took to persuade the Federal Open Market Committee and the Board of Governors to go along with this?

VOLCKER: First of all, raising interest rates quite visibly and openly is not the easiest thing in the world for central bankers or anybody. It's much easier to lower interest rates than it is to raise interest rates, I think it's fair to say, in almost any circumstances. That 4 to 3 vote that I referred to reflected something of that reluctance.

Three or four years earlier, there was some pressure in the Open Market Committee to adopt a much stricter money supply approach. So it wasn't entirely unknown to the Federal Reserve. Now that approach had always been rejected when it was raised. But there was a little bit of feeling, I knew, among some of the Open Market Committee members—in particular outside of the Board, the regional Reserve Bank presidents had a certain amount of sympathy—you had some kind of instinctive support. I think people were upset and tired about the way things were happening, they were looking for something different, something new, something that had some hope. They realized, I hate to overdramatize, this was a last chance. That's overdramatizing a bit.

People were willing to get together on the new policy. We knew pretty well that it would have a sharp impact on short-term rates. It was meant to be highly restrictive, no doubt about that. But the Board, the Open Market Committee, was pretty united.

FELDSTEIN: If I remember correctly, history has you warning President Carter that something like this was going to happen, getting his at least implicit consent, and then when it actually happened, he was a very unhappy man and acted up.

VOLCKER: I don't think it was quite that way.

FELDSTEIN: You were there, and I wasn't, so tell us.

VOLCKER: When I was appointed in August 1979, I had made clear to him that I thought Federal Reserve policy was too easy at the time. If I was going to be chairman, I was going to be advocating a stronger policy. I remember I thought I wasn't going to be appointed after *that* conversation. But I was appointed anyway, so that shows the difficulty of the time.

Then when it came to making the so-called October 1979 decision, I had warned the Secretary of the Treasury and the Chairman of the Council of Economic Advisers at the time, Bill Miller and Charlie Schultze.

We had to go to the annual IMF meeting in Belgrade at the end of September. I told them on the plane this was what I was planning to do. They were not too happy about it.

FELDSTEIN: You told them on the plane going *to* Belgrade?

VOLCKER: To Belgrade.

FELDSTEIN: To Belgrade? That's an important piece in history.[5]

VOLCKER: It was important for another reason. By some bit of serendipity, Helmut Schmidt, who was then, I guess, Chancellor of Germany, had requested that we stop and talk to him in Hamburg. I don't know how many people here know Helmut Schmidt, but he could be pretty acerbic to say the least. He wasn't exactly happy about the United States.

He sat there and lectured us for about an hour about the irresponsibility of the United States in letting this inflation get out of control, not having tight enough policies, and what was the matter with us, we were the leaders of the world, we'd better shape up and do something. I sat there rather happy about this lecture.

We went to Belgrade. I was aiming for an announcement later that week. This was probably over the weekend, I don't remember exactly. The President was informed about it, not by me. It may have been a deficiency on my part, but I didn't tell him. I would have told him, I guess, when I got back, but he had already been told by the time I got back.

He took the position, I'm told, that "I don't like it much, but I just appointed the guy, and I'm not going to make a public fuss about it." There were one or two people in the administration that I knew who kind of came to me and said, "Go ahead and do it." Now that wasn't the unanimous view, but there was no sharp reprimand, there was no head-on fight.

Credit Controls

FELDSTEIN: But President Carter went on television and he triggered some legal provisions which had some effect.[6]

VOLCKER: No, not then. This was a later catastrophe so far as I was concerned. You may not remember this. This was later when interest rates got up to, I don't know, 20 percent or so maybe by the end of 1979.

Carter announced a budget in early 1980 which was very poorly received. Nothing was happening. The Federal Reserve staff kept saying, "We're having this recession. The recession is beginning." There was no recession. In spite of all this, the economy kept rising.

Carter was obviously under pressure, so he triggered a provision of law that permitted the Federal Reserve to put on credit controls. He said, "I want to put on

[5] Silber (2012, p. 166) reports an interview with Schultze that confirms this account.

[6] For a transcript of President Carter's televised address of March 14, 1980, in which he called upon the Federal Reserve to impose credit controls, see Carter (1980).

credit controls. I want to show I'm on the team, sort of. We're going to go together, and I've got to change the budget. I'm going to make the budget more restrictive. I want to announce there will be credit controls." We were going to tighten policy again. That was all part of the package: "You wait to tighten policy. We'll announce credit controls."

We didn't like the idea because expansion of credit was not a problem at the time, and we didn't want to get into all the mess of managing credit controls. We thought it muddied the picture. I felt I didn't like it, but how can we rebuff the President of the United States who is asking us to put on credit controls? Theoretically, we could have said no, but the Board reluctantly went along at my urging.

FELDSTEIN: He managed to ask you on television in March 1980.

VOLCKER: I guess maybe he did.

FELDSTEIN: It was not a secret to the American public that he had that view.

VOLCKER: No, it wasn't a secret. Anyway, this was a phenomenon engraved in my mind.

We decided, "All right. We'll put on the most modest controls we can think of. We will not put on any control over anything to do with housing," which is the biggest source of credit. I think we exempted automobiles. There's nothing the matter with automobiles.

So the only thing that was left, and in those days it was of little importance, was installment credit that was not related to housing or cars, and credit cards, which was not a very big area of the market. We said, "Okay, you're going to have a reserve requirement on credit cards—if credit cards exceed past peaks, you would have a reserve requirement." We did that knowing, we're now in March, the peak in credit card use comes in November and December. We were way below it so there was no possibility that this was going to become a factor for some time. This was all announced at a big White House ceremony [laughs].

The economy at that point fell like a rock. People were cutting up credit cards, sending in the pieces to the President as their patriotic duty. Mobile home and automobile sales dropped within the space of a week or so. The money supply, we didn't know why the money supply was dropping, but all of the sudden the money supply was down 3 percent in a week or something.

What happened was everybody was paying off their credit card debt by drawing down their demand deposits and the money supply fell. We went into what the National Bureau of Economic Research later determined was a recession, and interest rates and the money supply dropped sharply.[7] Well, it was a recession alright, the economy went down, but it was an artificial recession. As soon as we took off the credit controls in June, the economy began expanding again. Then things really got tough. We reversed the easing during the recession and interest rates resumed rising, including a discount rate increase a month or so before the election. That was the only time President Carter publicly expressed concern.

[7] For the June 3, 1980, press release from the Business Cycle Dating Committee, see National Bureau of Economic Research (1980).

High Real Interest Rates and the Debt Crisis in Latin America

FELDSTEIN: Let me look a little further down the road. Even after inflation had dropped to 3 percent, you kept, or the market kept, interest rates very high. Treasury Bills were 8 and 9 percent in 1983 and 1984, so we had real rates on Treasury Bills of 5 and 6 percent. Why were they so high?

VOLCKER: Oh geez, I don't remember precisely. First of all when the economy began expanding, it was expanding very fast. There was no problem with poor economic activity at that point.[8]

Then, in the spring and early summer of 1984, you had the potential banking crisis of Continental Illinois. A little bank in Oklahoma went bust. But it had sold a lot of oil patch loans to Continental Illinois and other banks, and there was this sort of a banking crisis.[9] When Continental got in trouble, that had an effect on overnight bank lending. That kept, for a while—that was during the election period, actually—kept interest rates higher than we anticipated and really higher than we wanted.

There was quite a debate at the Board of Governors at that time whether we should react to what seemed an artificial increase in interest rates by easing our policies or whether we would tough it out.

FELDSTEIN: You still had very high real rates, short real rates. Was it a concern about inflation coming back?

VOLCKER: If I put myself back in that position, I think we were totally satisfied at what the economy was doing. You still had some inflation. Why would we be easing up when the economy was expanding 6 or 7 percent a year? We're getting back to where we were. Everything was fine. Sort of fine.

FELDSTEIN: Another development at about this time was the debt crisis in Latin America. What role did the high interest rates of the early 1980s play in causing severe debt problems in emerging economies? What options did you consider to deal with this?"

VOLCKER: That's an interesting question. This is something like, I suppose, the subprime mortgage thing. To understand what happened in the early 1980s, we need to start earlier.

The '70s were characterized by a lot of liquidity growing out of the oil crisis and the excess money that the Arabs had, and all the rest. That money was flowing through the big banks to Latin America in a way that arguably looked constructive for a while but was ultimately unsustainable.

[8] Silber (2012, p. 237) cites the 1984 Annual Report of the Council of Economic Advisors on the role of the full employment federal deficit: "[F]ederal borrowing to finance a budget deficit of five percent of GNP . . . means the real rate of interest must rise."

[9] The Oklahoma bank in question was Penn Square Bank. For an overview of the events surrounding the Continental Illinois banking crisis that came to a head in 1984, see Federal Deposit Insurance Corporation (1997).

Arthur Burns, to his credit, was the Paul Revere on this thing. He'd go around and make speeches: "This can't continue. It shouldn't be continued. We've got to do something about it."[10]

I was in the New York Fed then. We tried to do something about it, which was totally ineffectual, I must say. But, it went on and nobody was willing to say, "If you did something that was really effectual—if that's a word—you would have a crisis in Latin America if you shut off the flow." Nobody much wanted to be all that aggressive.

By the early 1980s, interest rates had gotten very high—I don't think it was the interest rates—the banks suddenly stopped lending to Mexico because they thought they were overexposed, so you had a crisis. Once one stopped, they all stopped.

This had been building up. The figures were known. The president of Mexico then, a left-wing guy, was being told by his own finance minister and central bank governor, "We ought to stop this or slow it down." He sent some people around to talk to foreign banks and ask, "Is this a problem?"[11] They all said no. This was in the summer of '81. Now we come to '82. He got rid of the finance minister! He wasn't going to get rid of the borrowing, but his finance minister instead. It's true.

The borrowing continued until the winter when a couple of banks stopped lending. Mexico ran out of money. What do you do? You had a big crisis now. The high interest rates were a burden for Mexico over time, but they didn't make the crisis. They hadn't been in effect all that long. But there's no question that high interest rates aggravated the problem.

What were you going to do? Were you going to conduct an easy-money policy and go back on all the policy you'd undertaken to try to save Mexico, which wouldn't have saved Mexico anyway?

We did save Mexico, but by other means. It wasn't just Mexico. People forget. This is a commentary on age. This was '82. How many years ago was that? Thirty-one or -two years ago. I hear all this talk about crisis. Nobody ever remembers the Latin American debt crisis. Memories only go back to, somehow, the savings and loan crisis in 1990 and don't make the leap back to '80.

The big US banks and some of the big foreign banks had more exposure to Latin America than they had capital. It wasn't something you could just say, "Okay, knock off the loans by 50 percent or something and everybody will be happy." They all would have been bust. You look for other approaches, and it took nearly a decade until Mr. Brady came along and settled them.[12]

[10] For example, see Burns (1977).

[11] The President of Mexico from 1976 to 1982 was José López Portillo. The Director General of the Bank of Mexico during this period was first Gustavo Romero Kolbeck, who was replaced by Miguel Mancera, who resigned in September 1982, protesting the nationalization of the banks. Portillo's finance minister during this time was David Ibarra Muñoz. He was fired in March 1982, nine months before the end of Portillo's term of office, and replaced by Jesus Silva Herzog.

[12] Nicholas Brady was Secretary of the Treasury from 1988 to 1993. In 1989, he announced what came to be known as the Brady Plan for addressing the problem of Latin American debt. It involved negotiating with creditors to accept "Brady bonds" in exchange for their holdings of Latin American debt. The Brady bonds had a lower face value or interest rates than the existing debts, but also were more certain to be repaid.

The Federal Reserve in the Recent Financial Crisis

FELDSTEIN: Let me now turn to more recent events after you left the Fed: first about the crisis and then about current policy. There were of course many causes of the financial crisis, and I don't want to review all of those. I want to ask what role you think the Fed played in causing the crisis? How could Fed policies have prevented it?

VOLCKER: I want to make a point that I think is important and it's underrated. We had a very mild recession, it was hardly a recession, in 2000 and 2001. I remember there was a meeting we both attended. You said it wasn't a recession. I said it wasn't a recession. The National Bureau of Economic Research later said it was a recession, but it was hardly visible. Anyway, we had these low interest rates in the early 2000s. We were running a big balance of payments deficit. It got bigger and bigger. More and more money came from Japan and China in particular. Interest rates were kept very low. It seemed to me, inevitably, this is the kind of doomsday scenario, sooner or later, that you couldn't go on to the point of borrowing 5, 6 percent of the GDP. Interest rates were very low, and parts of the economy were expanding unsustainably rapidly. What to do about it?

I made a speech about it once.[13] I didn't say anything except we've got to make sure we maintain price stability and budget discipline. I didn't directly criticize the Federal Reserve at the time because I wasn't sure—I mean, I would have been happy if he [Alan Greenspan] had been a little tighter, frankly, but I didn't think that was going to cure the situation because it was really an international monetary problem. There was no discipline in either the United States or in China. Nobody even raised the question, and it all ended up very unhappily.

That kind of fed the boom in the United States. I think the Federal Reserve and all the banking regulators did not catch up with this. I didn't know if anyone ever would, but they didn't. It got out of hand and collapsed in a way that I wasn't anticipating particularly, but it did. I had no idea, myself—I'm just sitting around reading newspapers—how big the subprime mortgage problem was. When I found out, it startled me.

You remember these little personal incidents. I was at some meeting, and I was asked to comment on the US economy, I guess in the spring of '07. Question: "What about that mortgage market?" I said, "I don't know much about that mortgage market. But I can't believe the financial system is so weak that this minor business in the mortgage market of subprime mortgages would upset it. I got back and I called some friends in the Federal Reserve. "How big is this subprime mortgage thing?" I must admit, the answer I got from them first was, "I don't know." Then, they called me back later, and they told me, "Well, it looks like it's over a trillion dollars." I had no imagination that this subprime mortgage thing was over a trillion dollars. It

[13] One such speech was given as the keynote address at the "summit" of the Stanford Institute for Economic Policy Research in February 2005, partly available on YouTube at http://www.youtube.com/watch?v=4aTatmAiiuY. Volcker (2005) is a newspaper op-ed piece adapted from the talk.

was a phenomena of, what, three years maybe? From basically a standing start, in three years it was a trillion dollars. Obviously that was kind of the focal point of the crisis when it finally came. Look, this banking regulation stuff is very hard to deal with. But I think there had been a relative lack of interest in it, which was unfortunate, to understate the matter.

The Volcker Rule

FELDSTEIN: Let's talk for a moment about the Volcker Rule, which I remember you were saying at the time . . .

VOLCKER: That's my favorite rule.

FELDSTEIN: It's your favorite rule. Well, if I had a rule named after me, it would probably be my favorite rule as well.

VOLCKER: We were at an international-level meeting last weekend. Somebody had a paper saying that was the most important part of financial reform. I have never said that, but she said it.

FELDSTEIN: That was my question. How important was proprietary trading, which I take it is the essence of the Volcker Rule?[14] How important was it as a cause of the crisis?

VOLCKER: I don't know whether I'd rank it as a prime cause of the crisis, but it was a contributing factor in the sense it led to a lot of, once the crisis started, exposure on proprietary trading and money market funds, and hedge funds.

This crisis kind of started with the hedge funds of Bear Stearns in 2007, and the institution came under strong pressure in early 2008. That failure began shaking psychology and so forth. That was essentially a proprietary trading operation. I have seen figures that say the banks collectively lost as much money in 2008 as they made on proprietary trading and hedge funds in the whole previous decade all in one fell swoop. But obviously the weakest part of the banking system was bad loans.

The difference is banks are there to make loans. That's an essential part of the economy. They're not there, in my opinion, to trade for their own account basically. That's a distinction that I try to make. That's obviously a complicating factor, if it wasn't the prime factor, in the crisis.

The worst part of it in a way, in my view, is a cultural, a psychological question. It's not just the risks that are involved directly for the whole institution.

Take this JPMorgan thing. They lost $6 billion, or whatever it was, with one little play in the derivatives market—one *big* play in the derivatives market.[15] They can

[14] The Volcker Rule, broadly understood, is that financial institutions that are eligible for deposit insurance and have access to the Federal Reserve and FDIC insurance should be limited in the risks they take with their proprietary trading. For an early presentation of this argument, see Group of Thirty (2009).

[15] For an overview of these events, in which a series of derivatives trades in spring 2012 cost JPMorgan approximately $6 billion, see the hearings of the US Senate (2013), titled "JPMorgan Chase Whale Trades: A Case History of Derivatives Risks and Abuses."

survive $6 billion. But what is the psychology that leads people to take that kind of risk? Traders know that the rewards are huge—of a kind that have not been at all normal in commercial banking now or in history. When you've got that kind of cleavage between the culture on the investment banking side of the house and the traditional banking side of the house, obviously the people in the commercial banking side say, "I want to make money, too. Maybe I can make some big risks and I'll get some mortgages together, and I'll package them up. Let's securitize them and stick them out. We'll make a commission on it. It's not a relationship matter. We're going to stick this out, we'll stick somebody else with it." It's a different culture.

The guy that is most eloquent on this, it surprises me because he never used to be friendly toward the Federal Reserve, is John Reed, head of the Citibank at the time. He was a leader in commercial banks going in the investment business. They bought Salomon Brothers investment bank back in 1998, you may recall. Salomon, that subsidiary, went bust later. Not too much later, but it was part of Citibank, so nothing happened. But he is very vocal: "Mea culpa. We made a mistake. It destroyed the culture of the institution."[16] That's my major worry about it.

FELDSTEIN: Another thing you've worried about: the size of the big banks. If I remember correctly, you were in favor of breaking up the big banks during this crisis. Is that true?

VOLCKER: I never took the view to break up the big banks. I wanted to limit their size, which I guess is in the law someplace—not very effectively. I sort of lack imagination. I don't see how you break them up without a lot of disturbance.

But even if you broke them up, you couldn't break them up into small enough pieces so that they wouldn't be systemically significant, or whatever we call it now. You break up JPMorgan in half, or Chase in half, they're the same bank. Bank of America, you slice them in half. They're not one two-trillion-dollar bank, they're two one-trillion-dollar banks. They're still too big.

FELDSTEIN: That view, which I've heard attributed to you, shouldn't have been attributed to you about wanting to break them up?

VOLCKER: No. But I wouldn't mind if somebody else does it.

The Dual Mandate

FELDSTEIN: Let me ask you more broadly about the goals of monetary policy. In the 1977 Humphrey–Hawkins Act, Congress adopted the dual mandate: that is, that monetary policy should be set with an eye on both inflation and unemployment. Recently, the Board has set quantitative goals for inflation and unemployment. If I remember in your time there were no such specific goals. You would say something like, "The job of the Fed is to achieve price stability."

VOLCKER: Right.

[16] For example, see Reed's interview with Bloomberg as reported in Ivry (2009).

FELDSTEIN: I remember that.

VOLCKER: And I'd excise the word "gradually" every time.

FELDSTEIN: My first question is, "What's your opinion of the dual mandate?"

VOLCKER: I'm against it.

FELDSTEIN: You're against the dual mandate? You want to say a little more than that?

VOLCKER: Well, I think it confuses the situation. The danger for the Federal Reserve now is that, implicitly or explicitly, given the circumstances it has acted and has been asked to act in an extraordinary way, it kind of gives the impression that the Federal Reserve has the keys to the kingdom—that they can achieve price stability and low unemployment at the same time, and it doesn't matter what the budget is, and all the structural problems in the economy, and the dislocations in the economy. Monetary policy will solve all problems.

I think that's a bad message to give, because I don't think it's right. I don't think it's possible anyway.

I do think it confuses the situation to say there's a trade-off between price stability, and economic performance, and employment. I think over any reasonable period of time there's not a trade-off. The best contribution of the Federal Reserve can be to maintain price stability.

I frankly don't like this inflation targeting, but that's a minor point—that 2 percent is okay and 1.5 is no good. Everybody kind of knows what price stability is and there's more than one measure of prices.

I think the dual mandate is confusing. I think it makes the Fed's job more difficult. That doesn't mean that policy would be one inch different today than it in fact is. There is no immediate inflation problem or inflation threats, so they can be comfortably very easy, that's what they should be. It doesn't imply any difference, actually, in current policy.

FELDSTEIN: Do you see any political possibility of moving away from the dual mandate any time in the future?

VOLCKER: I think if the Federal Reserve stopped talking about it, nobody else would talk about it. Congress doesn't pay any attention. That law was passed or amended—the Humphrey–Hawkins Act—in '77, just two years before I became Chairman of the Federal Reserve; it was fresh legislation. I do not remember the word "dual mandate" ever passing my lips in all the time that I was Chairman.

Now, I could get by with it because inflation was very high and if somebody asked me the impact on the economy, I would say, "Look, over time the best thing we can do for the economy is to get rid of the inflation." Sitting there saying with 15 percent inflation—well, then when we started we didn't have high unemployment—but even then when the unemployment rate got very high, and the inflation rate was still 10 percent or whatever it was, to say, "Let's get the inflation rate up a little bit so we can get the unemployment rate down." It didn't make sense. The literal reading of the dual mandate presumed that's what you would do. I think you're better off focusing on price stability.

That's the advice I would give to the current Chairman of the Federal Reserve.

FELDSTEIN: You've got a chance to do that now. That probably tells me something about what you think about the idea of unconventional policies. Maybe it doesn't, so you tell us.

VOLCKER: I think this crisis required some unconventional policies, there's no doubt about that. Extremely unconventional is the kindest word you can say about it when you go back a few years ago.

Let's talk about this current version, this so-called QE3.[17] It's a matter of judgment. I don't get alarmed about it, and I think they can manage their way out of it. Chairman Bernanke has made that quite clear and I think he's right.[18]

It does have the dangers of speculative excesses. It's got pluses and minuses. The pluses I don't think are very large. The minuses don't seem to be tremendous right at the moment either.

Since I can say it, if I was conducting these policies, I don't really understand why we're paying interest on excess reserves when we're worried about getting interest rates as low as possible. The Federal Reserve pays more on their excess reserves than the banks can get from lending to each other. So why pay them?

FELDSTEIN: You would stop paying interest on excess reserves?

VOLCKER: Yes, I would. It never dawned on me to pay interest on excess reserves—I guess a limitation of my own imagination. I was always in favor of paying interest on required reserves.[19] But the idea of paying interest on excess reserves never occurred to me until the Federal Reserve began doing it. I can see, in some circumstances, it may have some advantages. But I think you're going to find it has some disadvantages, too. Some day you have to make that rate pretty high if you're going to do it, I guess, when you want to tighten policy. We'll see how that goes.

FELDSTEIN: You mentioned the word "deficits." Let me read another question that came in: "Can the US continue for long as the financial global hegemon with the persistent large fiscal deficits?"

VOLCKER: What's a large fiscal deficit?

FELDSTEIN: A large fiscal deficit is what we have.

[17] QE refers to the "quantitative easing" policies in which the Federal Reserve purchased financial securities like US Treasury bonds and mortgage-backed securities. The first round of these policies, QE1, started in November 2008; the second round, QE2, in August 2010; and the third round, QE3, in September 2012.

[18] For example, see Bernanke's (2010) congressional testimony concerning "The Federal Reserve Exit Strategy."

[19] Traditionally, the Federal Reserve did not pay interest on the reserves that it required banks to hold, nor on any additional or "excess" reserves beyond the legal requirement that banks choose to hold. The Financial Services Regulatory Relief Act of 2006 authorized the Federal Reserve to pay interest on reserves beginning in October 2011. That authority was accelerated to October 2008 by the Emergency Economic Stabilization Act of 2008, and the Federal Reserve began to pay interest on reserves on October 9, 2008. For additional explanation and interest rates that are paid, see the Federal Reserve website at http://www.federalreserve.gov/monetarypolicy/reqresbalances.htm.

VOLCKER: Yes, that's what we have. I think it's hard, particularly the uncertainty over the decades ahead when Medicare, and Social Security, and so forth seem to widen the deficit. The deficit, what's more up to the point to me, is the balance of payments deficit. I think we're back, in a way, in the Triffin dilemma.[20]

In the 1960s, we were in a position in the Bretton Woods system with the other countries wanting to run surpluses and build their reserve positions, so the reserve position of the United States inevitably weakened—weakened to the point where we no longer could support the convertibility of currencies to gold. Now, how long can we expect as a country or world to support how many trillions of dollars that the rest of the world has? So far, so good. The rest of the world isn't in a very good shape, so we look pretty good at the moment.

But suppose that situation changes and we're running big deficits, and however many trillion it is now, it's another few trillions. At some point there is vulnerability there, I think, for the system, not just for the United States. We ought to be conscious of that and do something about it. It's not so easy to do something about it because that comes back to the whole question of international monetary reform, which is a favorite subject.

FELDSTEIN: The Chinese have been major buyers of US government debt. They've been able to do that because they have had a large current account surplus. That surplus has come down from 10 percent a few years ago to less than 2 percent now. If they pursue the policies they say they're going to pursue, it could easily disappear in the next couple of years. How should we think about the implications of that for the US economy?

VOLCKER: I think that's good news from the standpoint of what I just mentioned. China is not the only other country in the world. But China was an important accumulator of US dollars. If they stop accumulating, the kind of worry I just expressed is somewhat alleviated. It's not gone, because our current account deficit continues.

Our current account deficit is smaller, too, and hopefully can remain small. I hate to see our deficit go back to where it was. I'd like to see it disappear, but it's hard to make it disappear. No question about that.

FELDSTEIN: On that semi-optimistic note, let me thank you again for taking the time and giving us your views about all of this.

[20] Robert Triffin (1960) testified before the Joint Economic Committee of Congress that the economy with the world's reserve currency—then and now the US dollar—must be willing to run ongoing trade deficits so that the reserve currency will be available for the global economy, but this in turn means that foreign governments hold large quantities of dollar assets, which at some point as events change are likely to become a source of global financial instability. This situation became known as the "Triffin dilemma."

References

Bernanke, Ben S. 2010. "Federal Reserve's Exit Strategy." Testimony before the Committee on Financial Services, US House of Representatives, Washington, DC, February 10. http://www.federalreserve.gov/newsevents/testimony/bernanke20100210a.htm.

Burns, Arthur F. 1977. "The Need for Order in International Finance." Address at the Columbia University Graduate School of Business, New York, NY, April 12. http://fraser.stlouisfed.org/docs/historical/burns/Burns_19770412.pdf.

Burns, Arthur. 1979. "The Anguish of Central Banking: The 1979 Per Jacobsson Lecture." http://www.perjacobsson.org/lectures/1979.pdf.

Carter, Jimmy. 1980. "Anti-Inflation Program Remarks Announcing the Administration's Program." March 14. The American Presidency Project. http://www.presidency.ucsb.edu/ws/?pid=33142.

Federal Deposit Insurance Corporation. 1997. "Continental Illinois and Too Big to Fail." Chap. 1 of *History of the Eighties: Lessons for the Future,* Vol. 1: *An Examination of the Banking Crises of the 1980s and Early 1990s*. Federal Deposit Insurance Corporation. http://www.fdic.gov/bank/historical/history/.

Group of Thirty. 2009. *Financial Reform: A Framework for Financial Stability*. http://www.group30.org/images/PDF/Financial_Reform-A_Framework_for_Financial_Stability.pdf.

Ivry, Bob. 2009. "Reed Says 'I'm Sorry' for Role in Creating Citigroup." Bloomberg.com, November 6. http://www.bloomberg.com/apps/news?pid=newsarchive&sid=a.z4KpD77s80&pos=6.

National Bureau of Economic Research, Business Cycle Dating Committee. 1980. "Business Cycle Peaked in January." June 3. http://www.nber.org/cycles/june1980.html.

Silber, William L. 2012. *Volcker: The Triumph of Persistence*. New York: Bloomsbury Press.

Triffin, Robert. 1960. "Statement of Robert Triffin to the Joint Economic Committee of Congress." December 8. http://babel.hathitrust.org/cgi/pt?id=mdp.39015081224407;view=1up;seq=234.

US Senate, Committee on Homeland Security and Intergovernmental Affairs. 2013. "JPMorgan Chase Whale Trades: A Case History of Derivatives Risks and Abuses." March 15. http://www.hsgac.senate.gov/subcommittees/investigations/hearings/chase-whale-trades-a-case-history-of-derivatives-risks-and-abuses.

Volcker, Paul A. 2005. "An Economy On Thin Ice." *Washington Post*, April 10, p. B7. http://www.washingtonpost.com/wp-dyn/articles/A38725-2005Apr8.html.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 121–140

Market Reasoning as Moral Reasoning: Why Economists Should Re-engage with Political Philosophy[†]

Michael J. Sandel

There are some things money can't buy—friendship, for example. If I want more friends than I have, it clearly wouldn't work to buy some. A hired friend is not the same as the real thing. Somehow, the money that would buy the friendship dissolves the good I seek to acquire.

But most goods are not of this kind. Buying them does not ruin them. Consider kidneys. Some people favor a market in human organs; others are opposed. But those who oppose the buying and selling of kidneys cannot argue that a market in kidneys would destroy the good being sought. A bought kidney will work, assuming a good match. So if a market in human organs is objectionable, it must be for some other reason. Money can buy kidneys (as the black market attests); the question is whether it should be allowed to do so.

In my book *What Money Can't Buy: The Moral Limits of Markets*, I try to show that market values and market reasoning increasingly reach into spheres of life previously governed by nonmarket norms (Sandel 2012). In procreation and childrearing, health and education, sports and recreation, criminal justice, environmental protection, military service, political campaigns, public spaces, and civic life, money and markets play a growing role. I argue that this tendency is troubling; putting a price on every human activity erodes certain moral and civic goods worth caring about. We therefore need a public debate about where markets serve the public good and where they don't belong.

In this article, I would like to develop a related theme: When it comes to deciding whether this or that good should be allocated by the market or by

■ *Michael J. Sandel is the Anne T. and Robert M. Bass Professor of Government, Harvard University, Cambridge, Massachusetts. His email is msandel@gov.harvard.edu.*

[†] To access the disclosure statement, visit
http://dx.doi.org/10.1257/jep.27.4.121

doi=10.1257/jep.27.4.121

nonmarket principles, economics is a poor guide. On the face of it, this may seem puzzling. Explaining how markets work is a central subject of economics. So why has economics failed to provide a convincing basis for deciding what should, and what should not, be up for sale?

The reason lies in the conception of economics as a value-neutral science of human behavior and social choice. As I will try to show, deciding which social practices should be governed by market mechanisms requires a form of economic reasoning that is bound up with moral reasoning. But mainstream economic thinking asserts its independence from the contested terrain of moral and political philosophy. Economics textbooks emphasize the distinction between "positive" questions and normative ones, between explaining and prescribing. The popular book *Freakonomics* states the distinction plainly: "Morality represents the way we would like the world to work and economics represents how it actually does work." Economics "simply doesn't traffic in morality" (Levitt and Dubner 2006, pp. 11, 46, 190; see also Robbins 1932).

Moral Entanglements

Economists have not always understood their subject in this way. The classical economists, going back to Adam Smith, conceived of economics as a branch of moral and political philosophy. But the version of economics commonly taught today presents itself as an autonomous discipline, one that does not pass judgment on how income should be distributed or how this or that good should be valued. The notion that economics is a value-free science has always been questionable. But the more markets extend their reach into noneconomic aspects of life, the more entangled they become with moral questions.

To be clear, I am not writing here about the standard textbook limitations on markets. A considerable body of economic analysis is devoted to identifying "market failures," or situations in which unaided market forces are unlikely to produce an efficient result, such as imperfectly competitive markets, negative and positive externalities, public goods, imperfect information, and the like. Another body of economic literature addresses questions of inequality. But this literature tends to analyze the causes and consequences of inequality while claiming to be agnostic on normative questions of fairness and distributive justice. Outsourcing judgments about equity and fairness to philosophers seems to uphold the distinction between positive and normative inquiry.

But this intellectual division of labor is misleading, for two reasons. First, as Atkinson (2009) has recently observed, "economics *is* a moral science," despite protestations to the contrary. Efficiency only matters insofar as it makes society better off. But what counts as better off? The answer depends on some conception of the general welfare or the public good. Although "welfare economics has largely disappeared" from mainstream economics in recent decades, Atkinson writes, "economists have not ceased to make welfare statements." Articles in journals of economics "are replete with welfare statements" and reach "clear normative

conclusions," he states, even though the principles underlying those conclusions go largely unexamined. Mostly, the conclusions rest on utilitarian assumptions. But as John Rawls and other philosophers have pointed out, utilitarianism seeks to maximize welfare without regard for its distribution. Atkinson calls for a revival of welfare economics that acknowledges the defects of utilitarianism and considers a broader range of distributive principles.

A second reason to doubt that economics can be a value-free science of social choice points beyond debates about distributive justice to debates about commodification: Should sex be up for sale? What about surrogate motherhood, or pregnancy for pay? Is there anything wrong with mercenary armies, and if so, how should military service be allocated? Should universities sell some seats in the freshman class in order to raise money for worthy purposes, such as a new library, or scholarships for well-qualified students from poor families? Should the United States sell the right to immigrate? What about allowing existing US citizens to sell their citizenship to foreigners and swap places with them? Should we allow a free market in babies up for adoption? Should people be allowed to sell their votes?

Some of these controversial uses of markets would improve efficiency by enabling mutually advantageous exchanges. In some cases, negative externalities might outweigh the benefits to buyers and sellers. Even absent externalities, however, some market transactions are objectionable on moral grounds.

One such ground is that severe inequality can undermine the voluntary character of an exchange. If a desperately poor peasant sells a kidney, or a child, the choice to sell might be coerced, in effect, by the necessities of his or her situation. So one familiar argument in favor of markets—that the parties freely agree to the terms of the deal—is called into question by unequal bargaining conditions. In order to know whether a market choice is a free choice, we have to ask what inequalities in the background conditions of society undermine meaningful consent. This is a normative question that different theories of distributive justice answer in different ways.

A second moral objection is not about fairness and tainted consent, but about the tendency of market practices to corrupt or crowd out nonmarket values worth caring about. For example, we might hesitate to create a market in children on the grounds that putting them up for sale would price less-affluent parents out of the market or leave them with the cheapest, least desirable children (the fairness argument). But we might also oppose such a market on the grounds that putting a price tag on children would objectify them, fail to respect their dignity, and erode the norm of unconditional parental love (the corruption argument).

Even where markets improve efficiency, they may be undesirable if they corrupt or crowd out nonmarket norms of moral importance. So before we can decide whether to create a market in children, for example, we have to figure out what values and norms should govern the social practices of child-rearing and parenting. In this sense, market reasoning presupposes moral reasoning.

For those who assume that all values are merely subjective preferences not open to reasoned argument, it may seem odd to suggest that some ways of valuing goods are more appropriate, or fitting, or morally defensible than others. But such

judgments are unavoidable, and we make them—sometimes implicitly, sometimes explicitly—whenever we decide whether this or that good should be up for sale.

Economists are not unaware of the moral objection to monetizing all relationships. For example, Waldfogel (1993; 2009), like many economists, questions the rationality of gift giving. Analyzing what he calls the "deadweight loss of Christmas," he calculates the utility loss that results from people giving gifts rather than the cash equivalent. He attributes the practice of in-kind gift giving to "the stigma of cash giving." But he does not ask whether this stigma might be justified. He simply assumes it is an irrational obstacle to utility that should ideally be overcome. He does not consider the possibility that the stigma against monetary gifts, at least among lovers, spouses, and other intimates, may reflect norms worth honoring and encouraging, such as attentiveness and thoughtfulness.

Alvin Roth (2007) also recognizes moral objections to the commodification of certain social practices, when he writes of "repugnance as a constraint on markets." To contend with such repugnance, he designs in-kind kidney exchanges and other mechanisms that avoid outright buying and selling. Unlike Waldfogel, he does not treat repugnance as an irrational, utility-destroying taboo; he simply accepts it as a social fact and devises work-arounds. Roth does not morally assess the repugnant transactions he discusses. He does not ask which instances of repugnance reflect unthinking prejudice that should be challenged and which reflect morally weighty considerations that should be honored. This reluctance to pass judgment on repugnance may reflect the economist's hesitation to venture onto normative terrain.

But the project of devising in-kind exchanges presupposes some moral judgment about which instances of repugnance are justified and which ones are not. Consider human organs. Everyone recognizes that lives could be saved by increasing the supply of organs for transplantation. But some object to the buying and selling of kidneys on the grounds that removing an organ from one person and transferring it to another violates the sanctity and integrity of the human body. Others object on the grounds that buying and selling kidneys objectifies the human person by encouraging us to view our bodies as property, as collections of spare parts to be used for profit. Still others favor a market in kidneys on the grounds that we own ourselves and should be free to profit from our bodies in whatever way we choose.

Whether an outright market in kidneys or an in-kind exchange is morally defensible depends, at least in part, on which of these stances toward the body and human personhood is correct. If the first view is right, then all forms of organ transplantation, paid or gifted, are objectionable, notwithstanding the lives that could be saved. If the second view is right, then gifted but not paid kidney transfers are morally defensible. Insofar as kidney exchanges preserve the gift ethic and avoid promoting a mercenary, objectifying attitude toward the human body, they address the moral concern underlying the second view. If the third view is right, we should not limit kidney transfers to in-kind exchanges, but should allow people to buy and sell kidneys for cash.

Some of the most corrosive effects of markets on moral and civic practices are neither failures of efficiency in the economist's sense, nor matters of inequality. Instead, they involve the degradation that can occur when we turn all human

relationships into transactions and treat all good things in life as if they were commodities. The economic literature that acknowledges stigma and repugnance makes implicit judgments about these questions; otherwise, it would be unable to propose either market solutions or quasi-market alternatives. But it does not articulate and defend the basis of these judgments. Doing so would carry economic reasoning beyond the textbook distinction between positive and normative inquiry and call into question the conception of economics as a value-neutral science of social choice. I will try to show how this is so by considering arguments for and against the use of market mechanisms in some contested contexts.[1]

The Line-Standing Business

When Congressional committees hold hearings, they reserve some seats for the press and make others available to the general public on a first-come, first-served basis. Corporate lobbyists are keen to attend these hearings, but are loath to spend hours in line to assure themselves a seat. Their solution: Pay thousands of dollars to professional line-standing companies that hire homeless people and others to queue up for them (Montopoli 2004; Copeland 2005; Lerer 2007; Palmeri 2009).

A company called LineStanding.com describes itself as "a leader in the Congressional line standing business." It charges $50 dollars an hour for line-standing services, of which a portion is paid to the people who stand and wait. The business has recently expanded from Congress to the US Supreme Court. When the Court hears oral arguments in big constitutional cases, the demand for seats far exceeds the supply. But if you are willing to pay, LineStanding.com will get you a ringside seat in the highest court in the land. Business was brisk for the Obama healthcare case in July 2012, when the line began forming three days in advance. For the same-sex marriage cases in June 2013, some people queued up five days in advance, making the price of a seat in the courtroom about $6,000 (for reports of this practice in the popular press, see Cain 2011; Smith 2013; Associated Press 2013; Liptak 2013).

On efficiency grounds, it is hard to find fault with the line-standing business. The homeless people who spend hours queuing up receive a payment that makes the waiting worth their while. Those who employ their services gain access to a Congressional hearing or a Supreme Court argument that they are eager to attend and willing to pay for. And the company that arranges the deal makes money too. All of the parties are better off, and no one is worse off.

[1] A number of the sections of this paper draw upon Sandel (2012), especially from pp. 21–133. For those interested in following up specific discussions, here are the relevant page references to the 2012 book: "Ticket Scalpers and Line Standers," pp. 21–23; "Markets and Corruption," pp. 33–35; "Refugee Quotas," pp. 63–65; "Fines vs. Fees," pp. 65–70; "Tradeable Procreation Permits," pp. 70–72; "Paying to Shoot a Walrus," pp. 82–84; "Incentives and Moral Entanglements," pp. 88–91; "The Case against Gifts," pp. 98–103; "Crowding out Non-market Norms," pp. 113–120; "The Commercialization Effect," pp. 120–22; "Blood for Sale," pp. 122–125; "Two Tenets of Market Faith," pp. 125–127; and "Economizing Love," pp. 127–133.

And yet some people object. Senator Claire McCaskill, a Missouri Democrat, has tried to ban paid Congressional line standing, without success. "The notion that special interest groups can buy seats at congressional hearings like they would buy tickets to a concert or football game is offensive to me," she said (as quoted in O'Connor 2009; see also Hananel 2007).

But what exactly is objectionable about it? One objection is about fairness: It is unfair that wealthy lobbyists can corner the market on Congressional hearings, depriving ordinary citizens of the opportunity to attend. But unequal access is not the only troubling aspect of this practice. Suppose lobbyists were taxed when they hired line-standing companies, and the proceeds were used to make line-standing services affordable for ordinary citizens. The subsidies might take the form, say, of vouchers redeemable for discounted rates at line-standing companies. Such a scheme might ease the unfairness of the present system. But a further objection would remain: turning access to Congress into a product for sale demeans and degrades it.

We can see this more clearly if we ask why Congress "underprices" admission to its deliberations in the first place. Suppose, striving mightily to reduce the national debt, it decided to charge admission to its hearings—say, $1,000 for a front row seat at the House Appropriations Committee. Many people would object, not only on the grounds that the admission fee is unfair to those unable to afford it, but also on the grounds that charging the public to attend a Congressional hearing is a kind of corruption.

We often associate corruption with ill-gotten gains. But corruption refers to more than bribes and illicit payments. To corrupt a good or a social practice is to degrade it, to treat it according to a lower mode of valuation than is appropriate to it (on higher and lower modes of valuation, see Anderson 1993). Charging admission to Congressional hearings is a form of corruption in this sense. It treats Congress as if it were a business rather than an institution of representative government accessible to all citizens.

Cynics might reply that Congress is already a business, in that it routinely sells influence and favors to special interests. So why not acknowledge this openly and charge admission? The answer is that the influence peddling and self-dealing that already afflict Congress are also instances of corruption. They represent the degradation of government in the public interest. Implicit in any charge of corruption is a conception of the purposes and ends an institution (in this case, Congress) properly pursues. The line-standing industry on Capitol Hill is corrupt in this sense. It is not illegal, and the payments are made openly. But it degrades Congress by treating access to public deliberations as a source of private gain rather than an expression of equal citizenship.

This does not necessarily mean that queuing is the best way to allocate access to Congressional hearings or Supreme Court arguments. Another alternative, arguably more consistent with the ideal of equal citizenship than either queuing or paying, would be to distribute tickets by an online lottery, with the provision that they be nontransferable.

How Markets Leave Their Mark

Before we can decide whether a good should be allocated by market, queue, lottery, need, merit, or in some other way, we have to decide what kind of good it is and how it should be valued. This requires a moral judgment that economists, at least in their role as social scientists, hesitate to make.

Part of the appeal of market reasoning is that it seems to offer a nonjudgmental way of allocating goods. Each party to a deal decides what value to place on the goods being exchanged. If someone is willing to pay for sex or a kidney, and a consenting adult is willing to sell, the economist does not ask whether the parties have valued the goods appropriately. Asking such questions would entangle economics in controversies about virtue and the common good and thus violate the strictures of a purportedly value-neutral science. And yet it is difficult to decide where markets are appropriate without addressing these normative questions.

The textbook approach evades this quandary by assuming—usually implicitly—that putting a price on a good does not alter its meaning. It assumes, without argument, that the activity of buying and selling does not diminish the value of the things being bought and sold. This assumption may be plausible in the case of material goods. Whether you sell me a flat screen television, or give me one as a gift, the television will work just as well. But the same may not be true when market practices extend their reach into human relationships and civic practices—sex, child rearing, teaching and learning, voting, and so on. When market reasoning travels abroad, beyond the domain of televisions and toasters, market values may transform social practices, and not always for the better.

Refugee Quotas and Childcare Pickups

Consider, for example, a proposal for a global market in refugee quotas. Each year, more refugees seek asylum than the nations of the world are willing to take in. A law professor, inspired by the idea of tradable pollution permits, suggested a solution: Let an international body assign each country a yearly refugee quota, based on national wealth. Then, let nations buy and sell these obligations among themselves. So, for example, if Japan is allocated 10,000 refugees per year but doesn't want to take them, it could pay Russia, or Uganda, to take them instead. According to standard market logic, everyone benefits. Russia or Uganda gains a new source of national income, Japan meets its refugee obligations by outsourcing them, and more refugees are rescued than would otherwise find asylum (Schuck 1994, 1997).

The argument in favor of the scheme is that countries would likely accept higher refugee quotas if they have the freedom to buy their way out. Yet there is something distasteful about a market in refugees, even if it's for their own good. But what exactly is objectionable about it? It has something to do with the tendency of a market in refugees to change our view of who refugees are and how they should be treated. It encourages the participants—the buyers, the sellers, and also those whose

asylum is being haggled over—to think of refugees as burdens to be unloaded or as revenue sources, rather than as human beings in peril.

One might acknowledge the degrading effect of a market in refugees and still conclude that the scheme does more good than harm. But the example illustrates that markets are not mere mechanisms. They embody certain norms. They presuppose—and promote—certain ways of valuing the goods being exchanged.

Economists often assume that markets are inert, that they do not touch or taint the goods they regulate. But this is untrue. Markets leave their mark on social norms. Market incentives can even erode or crowd out nonmarket motivations.

A well-known study of some childcare centers in Israel shows how this can happen (Gneezy and Rustichini 2000a). The centers faced a familiar problem: parents sometimes came late to pick up their children. A teacher had to stay with the children until the tardy parents arrived. To solve this problem, the centers imposed a fine for late pickups. If you assume that people respond to financial incentives, you would expect the fine to reduce, not increase, the incidence of late pickups. Instead, late pickups increased.

What explains the result? Introducing the monetary payment changed the norms. Before, parents who came late felt guilty; they were imposing an inconvenience on the teachers. Now, parents considered a late pickup as a service for which they were willing to pay. They treated the fine as if it were a fee. Rather than imposing on the teacher, they were simply paying him or her to work longer. If the goal of the payment for late pickups was to cover the additional costs of lateness, they were arguably a success; but if the goal of the payments was to discourage lateness by penalizing it, they were a failure.

Fines versus Fees

It is worth considering the difference between a fine and a fee. Fines register moral disapproval, whereas fees are simply prices that imply no moral judgment. When the government imposes a fine for littering, it makes a statement that littering is wrong. Tossing a beer can into the Grand Canyon not only imposes cleanup costs. It reflects a bad attitude that we want to discourage. Suppose the fine is $100, and a wealthy hiker decides it is worth the convenience. He treats the fine as a fee and tosses his beer can into the Grand Canyon. Even if he pays up, we consider that he's done something wrong. By treating the Grand Canyon as an expensive dumpster, he has failed to appreciate it in an appropriate way.

Or consider the case of parking spaces reserved for use by the physically disabled. Suppose a busy but able-bodied contractor wants to park near his building site. For the convenience of parking his car in a place reserved for the disabled, this contractor is willing to pay the rather large fine. He considers it a cost of doing business. Even if he pays the fine, wouldn't we consider that he is doing something wrong? He treats the fine as if it were simply an expensive parking lot fee. But in treating the fine as a fee, he fails to respect the needs of the physically disabled

and the effort of the community to accommodate them by setting aside certain parking spaces.

In practice, the distinction between a fine and a fee can be unstable. In China, the fine for violating the government's one-child policy is increasingly regarded by the affluent as a price for an extra child. The policy, put in place over three decades ago to reduce China's population growth, limits most couples in urban areas to one child. (Rural families are allowed a second child if the first one is a girl.) The fine varies from region to region, but reaches 200,000 yuan (about $31,000) in major cities—a staggering figure for the average worker, but easily affordable for wealthy entrepreneurs, sports stars, and celebrities (Moore 2009; Bristow 2007; Coonan 2011; Ming'ai 2007).

China's family planning officials have sought to reassert the punitive aspect of the sanction by increasing fines for affluent offenders, denouncing celebrities who violate the policy and banning them from appearing on television, and preventing business executives with extra kids from receiving government contracts. "The fine is a piece of cake for the rich," explained Zhai Zhenwu, a Renmin University sociology professor (Moore 2009). "The government had to hit them harder where it really hurt, at their fame, reputation, and standing in society" (for discussion, see also Xinhua News Agency 2008; Liu 2008).

The Chinese authorities regard the fine as a penalty and want to preserve the stigma associated with it. They don't want it to devolve into a fee. This is not mainly because they're worried about affluent parents having too many children; the number of wealthy offenders is relatively small. What is at stake is the norm underlying the policy. If the fine were merely a price, the state would find itself in the awkward business of selling a right to have extra children to those able and willing to pay for them.

Tradable Procreation Permits

Some Western economists have called for a market-based approach to population control strikingly similar to the one the Chinese seem determined to avoid: that is, they have urged countries that seek to limit their population to issue tradable procreation permits. For example, Kenneth Boulding (1964) proposed a system of marketable procreation licenses as a solution to overpopulation. Each woman would be issued a certificate (or two, depending on the policy) entitling her to have a child. She would be free to use the certificate or sell it at the going rate. Boulding (pp. 135–36) imagined a market in which people eager to have children would purchase certificates from (as he indelicately put it) "the poor, the nuns, the maiden aunts, and so on."

The plan would be less coercive than a system of fixed quotas, as in a one-child policy. It would also be economically more efficient, since it would get the goods (in this case, children) to the consumers most willing to pay for them. Recently, two Belgian economists revived Boulding's proposal. They pointed out that, since

the rich would likely buy procreation licenses from the poor, the scheme would have the further advantage of reducing inequality by giving the poor a new source of income (de la Croix and Gosseries 2006).

Some people oppose restrictions on procreation, whether mandatory or market-based. Others believe that reproductive rights can legitimately be restricted to avoid overpopulation. Set aside for the moment that disagreement of principle and imagine a society that was determined to implement mandatory population control. Which policy would be less objectionable: a fixed quota that limits each couple to one child and fines those who exceed the limit, or a market-based system that issues each couple a tradable procreation voucher entitling the bearer to have one child?

From the standpoint of economic reasoning, the second policy is clearly preferable. The freedom to choose whether to use the voucher or sell it makes some people better off and no one worse off. Those who buy or sell vouchers gain (by making mutually advantageous trades), and those who don't enter the market are no worse off than they would be under the fixed quota system; they can still have one child.

And yet, there is something troubling about a system in which people buy and sell the right to have kids. Part of what is troubling is the unfairness of such a system under conditions of inequality. We hesitate to make children a luxury good, affordable by the rich but not the poor. Beyond the fairness objection is the potentially corrosive effect on parental attitudes and norms. At the heart of the market transaction is a morally disquieting activity: parents who want an extra child must induce or entice other prospective parents to sell off their right to have a child.

Some might argue that a market in procreation permits has the virtue of efficiency; it allocates children to those who value them most highly, as measured by the ability to pay. But trafficking in the right to procreate may promote a mercenary attitude toward children and corrupt the norm of unconditional love of parents for their children. For consider: Wouldn't the experience of loving your children be tainted if you acquired some of them by bribing other couples to remain childless? Might you be tempted, at least, to hide this fact from your children? If so, there is reason to conclude that, whatever its advantages, a market in procreation permits would corrupt parenthood in ways that a fixed quota, however odious, would not.

In deciding whether to commodify a good, we must consider more than efficiency and fairness. We must also ask whether market norms will crowd out nonmarket norms, and if so, whether this represents a loss worth caring about.

Paying to Shoot a Walrus

Consider another kind of tradable quota—the right to shoot a walrus. Although the Atlantic walrus was once abundant in the Arctic region of Canada, the massive, defenseless marine mammal was easy prey for hunters, and by the late nineteenth century the population had been decimated. In 1928, Canada banned walrus

hunting, with a small exception for aboriginal subsistence hunters whose way of life had revolved around the walrus hunt for 4,500 years.

In the 1990s, Inuit leaders approached the Canadian government with a proposal. Why not allow the Inuit to sell the right to kill some of their walrus quota to big-game hunters? The number of walruses killed would remain the same. The Inuit would collect the hunting fees, serve as guides to the trophy hunters, supervise the kill, and keep the meat and skins as they had always done. The scheme would improve the economic wellbeing of a poor community, without exceeding the existing quota. The Canadian government agreed.

Today, rich trophy hunters from around the world make their way to the Arctic for the chance to shoot a walrus. They pay $6,000 to $6,500 for the privilege. They do not come for the thrill of the chase or the challenge of stalking an elusive prey. Walruses are unthreatening creatures that move slowly and are no match for hunters with guns. In a compelling account in the *New York Times Magazine*, Chivers (2002) compares walrus hunting under Inuit supervision to "a long boat ride to shoot a very large beanbag chair." The guides maneuver the boat to within 15 yards of the walrus and tell the hunter when to shoot. Chivers describes the scene as a game hunter from Texas shot his prey: "[The] bullet smacked the bull on the neck, jerking its head and knocking the animal to its side. Blood spouted from the entry point. The bull lay motionless. [The hunter] put down his rifle and picked up his video camera." The Inuit crew then pull the dead walrus onto an ice floe and carve up the carcass.

The appeal of the hunt is difficult to fathom. It involves no challenge, making it less a sport than a kind of lethal tourism. The hunter cannot even display the remains of his prey on his trophy wall back home. Walruses are protected in the United States, and it is illegal to bring their body parts into the country.

So why shoot a walrus? Apparently, the main reason is to fulfill the goal of killing one specimen of every creature on lists provided by hunting clubs—for example, the African "Big Five" (leopard, lion, elephant, rhino, and cape buffalo), or the Arctic "Grand Slam (caribou, musk ox, polar bear, and walrus).

It hardly seems an admirable goal; many find it repugnant. But from the standpoint of market reasoning, there is much to be said for allowing the Inuit to sell their right to shoot a certain number of walruses. The Inuit gain a new source of income, and the "list hunters" gain the chance to complete their roster of creatures killed—all without exceeding the existing quota. In this respect, selling the right to kill a walrus is like selling the right to procreate, or to pollute. Once you have a quota, market logic dictates that allowing tradable permits improves the general welfare. It makes some people better off without making anyone worse off.

And yet there is something morally disagreeable about the market in walrus killing. Let's assume, for the sake of argument, that it is reasonable to permit the Inuit to carry on with subsistence walrus hunting as they've done for centuries. Allowing them to sell the right to kill "their" walruses is nonetheless open to two moral objections. First, it can be argued that this bizarre market caters to a perverse desire

that should carry no weight in any calculus of social utility. Whatever one thinks of other forms of big-game hunting, the desire to kill a helpless mammal at close range, without any challenge or chase, simply to complete a list, is not worthy of being fulfilled. To the contrary, it should be discouraged. Second, for the Inuit to sell outsiders the right to kill their allotted walruses arguably corrupts the meaning and purpose of the exemption accorded their community in the first place. It is one thing is to honor the Inuit way of life and to respect its long-standing reliance on subsistence walrus hunting. It is quite another to convert that privilege into a cash concession in killing on the side.

Of course, the moral judgments underlying these objections are contestable. Some might defend the system of tradable walrus-hunting quotas on the grounds that the desire to shoot a walrus is not perverse but morally legitimate, worthy of consideration in determining the general welfare. It might also be argued that the Inuit themselves, not outside observers, should determine what counts as respecting their cultural traditions. My point is simply this: deciding whether or not to permit the Inuit to sell their right to shoot walruses requires debating and resolving these competing moral judgments.

Crowding out Nonmarket Norms

Markets in refugee quotas, procreation permits, and the right to shoot a walrus, however efficient in economic terms, are questionable policy to the extent that they erode the attitudes and norms that should govern the treatment of refugees, children, and endangered species. The problem I am emphasizing here is not that such markets are unfair to those who can't afford the goods being sold (although this may well be true), but that selling such things can be corrupting.

Standard economic reasoning assumes that commodifying a good—putting it up for sale—does not alter its character; market exchanges increase economic efficiency without changing the goods themselves. But this assumption is open to doubt. As markets reach into spheres of life traditionally governed by nonmarket norms, the notion that markets never touch or taint the goods they exchange becomes increasingly implausible. A growing body of research confirms what common sense suggests: financial incentives and other market mechanisms can backfire by crowding out nonmarket norms.

The day care study offers one example. Introducing a monetary payment for late arrivals increased rather than reduced the number of parents arriving late. It is no doubt true that, if the fine were high enough (say, $1,000 an hour), the standard price effect would win out. But all that matters for my argument is that introducing a monetary incentive or disincentive can sometimes corrupt or crowd out nonmarket attitudes and norms. When and to what extent the "crowding out" effect may trump the price effect is an empirical question. But even the existence of a "crowding out" effect shows that markets are not neutral; introducing a market mechanism may change the character and meaning of a social practice. If this is true, deciding to

use a cash incentive or a tradable quota requires that we evaluate, in each case, the nonmarket values and norms such mechanisms may displace or transform.

Several other studies also demonstrate the crowding out effect:

Nuclear Waste Siting

When residents of a Swiss town were asked whether they would be willing to approve a nuclear waste site in their community if the Parliament decided to build it there, 51 percent said yes. Then the respondents were offered a sweetener: Suppose the Parliament proposed building the nuclear waste facility in your community *and* offered to compensate each resident with an annual monetary payment. (Frey, Oberholzer-Gee, Eichenberger 1996; Frey and Oberholzer-Gee 1997; see also Frey 1997, pp. 67–78). Adding the financial inducement did not increase the rate of acceptance. In fact, it cut it in half—from 51 percent to 25 percent. Similar reactions to monetary offers have been found in other places where local communities have resisted radioactive waste repositories (Frey, Oberholzer-Gee, and Eichenberger 1996, pp. 1300, 1307; Frey and Oberholzer-Gee 1997, p. 750; Kunreuther and Easterling 1996, pp. 606–608).

Why would more people accept nuclear waste for free than for pay? For many, the willingness to accept the waste site apparently reflected public spirit—a recognition that the country as a whole depended on nuclear energy, and that the waste had to be stored somewhere. If their community was found to be the safest site, they were willing to sacrifice for the sake of the common good. But they were not willing to sell out their safety and put their families at risk for money. In fact, 83 percent of those who rejected the monetary proposal explained their opposition by saying they could not be bribed (Frey, Oberholzer-Gee, and Eichenberger 1996, p. 1306). The offer of a private payoff had transformed a civic question into a pecuniary one. The introduction of market norms crowded out their sense of civic duty (Kunreuther and Easterling 1996, pp. 615–19; Frey, Oberholzer-Gee, and Eichenberger 1996, p. 1301; for an argument in favor of cash compensation, see O'Hare 1977).

Donation Day

Each year, on a designated day, Israeli high school students go door-to-door to solicit donations for worthy causes—cancer research, aid to disabled children, and so on. Gneezy and Rustichini (2000b) did an experiment to determine the effect of financial incentives on the students' motivations. They divided the students into three groups. One group of students was given a brief motivational speech about the importance of the cause, and sent on its way. The second and third groups were given the same speech, but also offered a monetary reward based on the amount they collected—1 percent and 10 percent respectively. The rewards would not be deducted from the charitable donations, but would come from a separate source.

Not surprisingly, the students who were offered 10 percent collected more in donations than those who were offered 1 percent. But the unpaid students collected more than either of the paid groups, including those who received the high commission. Gneezy and Rustichini (2000b, 802–807) conclude that, if you're going to use

financial incentives to motivate people, you should either "pay enough or don't pay at all." While it may be true that paying enough will get what you want, there is also a lesson here about how money crowds out norms.

Why did both paid groups lag behind those doing it for free? Most likely, it was because paying students to do a good deed changed the character of the activity. Going door-to-door collecting funds for charity was now less about performing a civic duty and more about earning a commission. The financial incentive transformed a public-spirited activity into a job for pay. As with the Swiss villagers, so with the Israeli students: the introduction of market norms displaced, or at least dampened, their moral and civic commitment.

Why worry about the tendency of markets to crowd out moral and civic ideals? For two reasons—one fiscal, the other ethical. From an economic point of view, social norms such as civic virtue and public spiritedness are great bargains. They motivate socially useful behavior that would otherwise cost a lot to buy. If you had to rely on financial incentives to get communities to accept nuclear waste, you would have to pay a lot more than if you could rely instead on the residents' sense of civic obligation. If you had to hire school children to collect charitable donations, you would have to pay more than a 10 percent commission to get the same result that public spirit produces for free.

But to view moral and civic norms simply as cost-effective ways of motivating people ignores the intrinsic value of the norms. Relying solely on cash payments to induce residents to accept a nuclear waste facility is not only expensive; it is corrupting. The reason it is corrupting is that it bypasses persuasion and the kind of consent that arises from deliberating about the risks the facility poses and the larger community's need for it. In a similar way, paying students to collect charitable contributions on donation day not only adds to the cost of fundraising; it dishonors their public spirit and disfigures their moral and civic education.

The Commercialization Effect

Many economists now recognize that markets change the character of the goods and social practices they govern. In recent years, one of the first to emphasize the corrosive effect of markets on nonmarket norms was Fred Hirsch, a British economist who served as a senior advisor to the International Monetary Fund. In a book published the same year that Gary Becker's (1976) influential work *An Economic Approach to Human Behavior* appeared, Hirsch (1976) challenged the assumption that the value of a good is the same whether provided through the market or in some other way. Hirsch (pp. 87, 93, 92) argued that mainstream economics had overlooked what he called the "commercialization effect." By this he meant "the effect on the characteristics of a product or activity of supplying it exclusively or predominantly on commercial terms rather than on some other basis—such as informal exchange, mutual obligation, altruism or love, or feelings of service or obligation." The "common assumption, almost always hidden, is that the commercialization

process does not affect the product." Hirsch observed that this mistaken assumption loomed large in the rising "economic imperialism" of the time, including attempts, by Becker and others, to extend economic analysis into neighboring realms of social and political life. The empirical cases we've just considered support Hirsch's (1976) insight—that the introduction of market incentives and mechanisms can change people's attitudes and crowd out nonmarket values.

A growing body of work in social psychology offers a possible explanation for this commercialization effect. These studies highlight the difference between intrinsic motivations (such as moral conviction or interest in the task at hand) and external ones (such as money or other tangible rewards). When people are engaged in an activity they consider intrinsically worthwhile, offering them money may weaken their motivation by depreciating or "crowding out" their intrinsic interest or commitment. (For an overview and analysis of 128 studies on the effects of extrinsic rewards on intrinsic motivations, see Deci, Koestner, and Ryan 1999).

Standard economic theory assumes that all motivations, whatever their character or source, are additive. But this misses the corrosive effect of money. The "crowding out" phenomenon has far-reaching implications for economics. It calls into question the use of market mechanisms and market reasoning in many aspects of social life, including the use of financial incentives to motivate performance in education, health care, the workplace, voluntary associations, civic life, and other settings in which intrinsic motivations or moral commitments matter (Janssen and Mendys-Kamphorst 2004).

Blood for Sale

Perhaps the best-known illustration of markets crowding out nonmarket norms is a classic study of blood donation by the British sociologist Richard Titmuss. In his book *The Gift Relationship*, Titmuss (1971) compared the system of blood collection used in the United Kingdom, where all blood for transfusion was given by unpaid, voluntary donors, and the system in the United States, where some blood was donated and some bought by commercial blood banks from people, typically the poor, who were willing to sell their blood as a way of making money. Titmuss presented a wealth of data showing that, in economic and practical terms alone, the UK blood collection system worked better than the American one. Despite the supposed efficiency of markets, he argued, the American system led to chronic shortages, wasted blood, higher costs, and a greater risk of blood contaminated by hepatitis (pp. 231–32).

But Titmuss (1971) also leveled an ethical argument against the buying and selling of blood. He argued that turning blood into a market commodity eroded people's sense of obligation to donate blood, diminished the spirit of altruism, and undermined the "gift relationship" as an active feature of social life. "Commercialization and profit in blood has been driving out the voluntary donor," he wrote. Once people begin to view blood as a commodity that is routinely bought and sold, Titmuss (pp. 223–24, 177) suggested, they are less likely to feel a moral responsibility to donate it.

Titmuss's book prompted much debate. Among his critics was Kenneth Arrow (1972). In taking issue with Titmuss, Arrow invoked two assumptions about human nature and moral life that economists often assert but rarely defend (for an insightful contemporary reply to Arrow, see Singer 1973). The first is the assumption I have examined above, that commercializing an activity doesn't change it. According to this assumption, if a previously untraded good is made tradable, those who wish to buy and sell it can do so, thereby increasing their utility, while those who regard the good as priceless are free to desist from trafficking in it. This line of reasoning leans heavily on the notion that creating a market in blood does not erode the value or meaning of donating blood out of altruism. Titmuss attaches independent moral value to the generosity that motivates the gift. But Arrow (1972, p. 351) doubts that such generosity could be diminished or impaired by the introduction of a market: "Why should it be that the creation of a market for blood would decrease the altruism embodied in giving blood?"

The answer is that commercializing blood changes the meaning of donating it. In a world where blood is routinely bought and sold, giving it away for free may come to seem a kind of folly. Moreover, those who would donate a pint of blood at their local Red Cross might wonder if doing so is an act of generosity or an unfair labor practice that deprives a needy person of gainful employment selling his blood. If you want to support a blood drive, would it be better to donate blood yourself, or to donate $50 that can be used to buy an extra pint of blood from a homeless person who needs the income?

The second assumption that figures in Arrow's (1972) critique is that ethical behavior is a commodity that needs to be economized. The idea is this: We should not rely too heavily on altruism, generosity, solidarity, or civic duty, because these moral sentiments are scarce resources that are depleted with use. Markets, which rely on self-interest, spare us from using up the limited supply of virtue. So, for example, if we rely on the generosity of the public for the supply of blood, there will be less generosity left over for other social or charitable purposes. "Like many economists," Arrow (1972, pp. 354–55) writes, "I do not want to rely too heavily on substituting ethics for self-interest. I think it best on the whole that the requirement of ethical behavior be confined to those circumstances where the price system breaks down . . . We do not wish to use up recklessly the scarce resources of altruistic motivation."

It is easy to see how this economistic conception of virtue, if true, provides yet further grounds for extending markets into every sphere of life. If the supply of altruism, generosity, and civic virtue is fixed, as if by nature, like the supply of fossil fuels, then we should try to conserve it. The more we use, the less we have. On this assumption, relying more on markets and less on morals is a way of preserving a scarce resource.

Economizing Love

The classic statement of this idea was offered by Sir Dennis H. Robertson (1954), a Cambridge University economist and former student of John Maynard Keynes, in an address at the bicentennial of Columbia University. The title of Robertson's

lecture was a question: "What does the economist economize?" He sought to show that, despite catering to what he called (p. 148) "the aggressive and acquisitive instincts" of human beings, economists nonetheless serve a moral mission.

Robertson (1954) claimed that by promoting policies that rely, whenever possible, on self-interest rather than altruism or moral considerations, the economist saves society from squandering its scarce supply of virtue. "If we economists do [our] business well," Robertson (p. 154) concluded, "we can, I believe, contribute mightily to the economizing . . . of that scarce resource Love," the "most precious thing in the world."

To those not steeped in economics, this way of thinking about the generous virtues is strange, even far-fetched. It ignores the possibility that our capacity for love and benevolence is not depleted with use but enlarged with practice. Think of a loving couple. If, over a lifetime, they asked little of one another, in hopes of hoarding their love, how well would they fare? Wouldn't their love deepen rather than diminish the more they called upon it? Would they do better to treat one another in more calculating fashion, to conserve their love for the times they really needed it?

Similar questions can be asked about social solidarity and civic virtue. Should we try to conserve civic virtue by telling citizens to go shopping until their country really needs them? Or do civic virtue and public spirit atrophy with disuse? Many moralists have taken the second view. Aristotle (*Nicomachean Ethics*, Book II, chap. 1, pp. 1103a–1103b) taught that virtue is something we cultivate with practice: "We become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts."

Rousseau (1762 [1973] Book III, chap. 15, pp. 239–40) held a similar view. The more a country asks of its citizens, the greater their devotion to it. "In a well-ordered city every man flies to the assemblies." Under a bad government, no one participates in public life "because no one is interested in what happens there" and "domestic cares are all-absorbing." Civic virtue is built up, not spent down, by strenuous citizenship. Use it or lose it, Rousseau says, in effect. "As soon as public service ceases to be the chief business of the citizens, and they would rather serve with their money than with their person, the state is not far from its fall."

The notion that love and generosity are scarce resources that are depleted with use continues to exert a powerful hold on the moral imagination of economists, even if they don't argue for it explicitly. It is not an official textbook principle, like the law of supply and demand. No one has proven it empirically. It is more like an adage, a piece of folk wisdom, to which many economists nonetheless subscribe.

Almost half a century after Robertson's lecture, Lawrence Summers, then the president of Harvard University, was invited to offer the Morning Prayers address in Harvard's Memorial Church. He chose as his theme what "economics can contribute to thinking about moral questions." Economics, Summers (2003) stated, "is too rarely appreciated for its moral as well as practical significance." Summers observed that economists place "great emphasis on respect for individuals—and the needs, tastes, choices, and judgments they make for themselves." He illustrated the moral implications of economic thinking by challenging students who had advocated a boycott of goods

produced by sweatshop labor: "We all deplore the conditions in which so many on this planet work and the paltry compensation they receive. And yet there is surely some moral force to the concern that as long as the workers are voluntarily employed, they have chosen to work because they are working to their best alternative. Is narrowing an individual's set of choices an act of respect, of charity, even of concern?"

Summers (2003) concluded with a reply to those who criticize markets for relying on selfishness and greed: "We all have only so much altruism in us. Economists like me think of altruism as a valuable and rare good that needs conserving. Far better to conserve it by designing a system in which people's wants will be satisfied by individuals being selfish, and saving that altruism for our families, our friends, and the many social problems in this world that markets cannot solve."

Here was Robertson's (1954) adage reasserted. This economistic view of virtue fuels the faith in markets and propels their reach into places they don't belong. But the metaphor is questionable. Are altruism, generosity, solidarity, and civic spirit like commodities that are depleted with use? Or are they more like muscles that develop and grow stronger with exercise?

Market Reasoning as Moral Reasoning

To answer this question is to take sides in a long-standing debate in moral and political philosophy. We have now seen two ways in which economic reasoning rests on contestable normative assumptions. One is the assumption that subjecting a good to market exchange does not alter its meaning; the other is the claim that virtue is a commodity that is depleted with use.

The extension of market thinking into almost every aspect of social life complicates the distinction between market reasoning and moral reasoning, between explaining the world and improving it. Where markets erode nonmarket norms, we need to ask whether this represents a loss worth caring about. Do the efficiency gains of tradable refugee quotas outweigh the degrading effect they may inflict on refugees? Are the economic benefits of commercialized walrus hunts worth the coarsened attitudes toward endangered species they may engender and promote? Should we worry if cash compensation for civic sacrifice turns patriotic sentiments to pecuniary ones?

Questions such as these carry us beyond predicting whether a market mechanism will "work" in a narrow sense. They require that we make a moral assessment: What is the moral importance of the attitudes and norms that money may crowd out? Would their loss change the character of the activity in ways we would regret? If so, should we avoid introducing financial incentives into the activity, even though they might offer certain benefits?

To decide when to use cash incentives, or tradable permits, or other market mechanisms, economists must go beyond identifying the norms that inform social practices; they must also evaluate those norms. The more economic thinking extends its reach into social and civic life, the more market reasoning becomes inseparable from moral reasoning. If economics is to help us decide where markets serve the

public good and where they don't belong, it should relinquish the claim to be a value-neutral science and reconnect with its origins in moral and political philosophy.

■ *I am grateful to the editors of this journal, David Autor, Timothy Taylor, and Ulrike Malmendier, for their challenging comments and criticisms. Timothy Besley's recently published essay (Besley 2013) on my book* What Money Can't Buy *helped me sharpen the arguments of this paper, as did a valuable conversation with Peter Ganong. I would also like to thank Robert Frank and the participants in New York University's Paduano seminar, and my colleagues in Harvard Law School's summer faculty workshop, for instructive and penetrating discussions of an earlier version of this paper.*

References

Anderson, Elizabeth. 1993. *Value in Ethics and Economics.* Cambridge: Harvard University Press.

Aristotle. 1925. *Nicomachean Ethics.* Translated by David Ross. New York: Oxford University Press.

Arrow, Kenneth J. 1972. "Gifts and Exchanges." *Philosophy and Public Affairs* 1(4): 343–62.

Associated Press. 2013. "Tickets to Watch U.S. Supreme Court Decisions are Costing People Time, Money." *San Francisco Examiner*, March 25. http://www.sfexaminer.com/sanfrancisco/tickets-to-watch-us-supreme-court-decisions-on-gay-marriage-are-costing-people-time-money/Content?oid=2338811.

Atkinson, A. B. 2009. "Economics as a Moral Science." *Economica* 76(Issue Supplement S1): 791–804.

Besley, Timothy. 2013. "What's the Good of the Market? An Essay on Michael Sandel's *What Money Can't Buy*." *Journal of Economic Literature* 51(2): 478–95.

Boulding, Kenneth E. 1964. *The Meaning of the Twentieth Century.* New York: Harper & Row.

Bristow, Michael. 2007. "Grey Areas in China's One-Child Policy." BBC News, September 21. http://news.bbc.co.uk/2/hi/asia-pacific/7002201.stm.

Cain, Robyn Hagan. 2011. "Need a Seat at Supreme Court Oral Arguments? Hire a Line Stander." FindLaw, September 2. http://blogs.findlaw.com/supreme_court/2011/09/need-a-seat-at-supreme-court-oral-arguments-hire-a-line-stander.html.

Chivers, C. J. 2002. "A Big Game." *New York Times Magazine*, August 25. http://www.nytimes.com/2002/08/25/magazine/a-big-game.html?pagewanted=all&src=pm.

Coonan, Clifford. 2011. "China Eases Rules on One Child Policy." *Independent* (London), April 1. http://www.independent.co.uk/news/world/asia/china-eases-rules-on-one-child-policy-2258875.html.

Copeland, Libby. 2005. "The Line Starts Here." *Washington Post*, March 2, p. C1. http://www.washingtonpost.com/wp-dyn/articles/A64231-2005Mar1.html.

de la Croix, David, and Axel Gosseries. 2006. "Procreation, Migration and Tradable Quotas." CORE Discussion Paper No. 2006/98, November. Available at SSRN: http://ssrn.com/abstract=970294.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* 125(6): 627–68.

Frey, Bruno S. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation.* Cheltenham, UK: Edward Elgar Publishing.

Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15(5): 589–611.

Frey, Bruno S., and Felix Oberholzer-Gee. 1997. "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out." *American Economic Review* 87(4): 746–55.

Frey, Bruno S., Felix Oberholzer-Gee, Reiner Eichenberger. 1996. "The Old Lady Visits Your Backyard: A Tale of Morals and Markets." *Journal of Political Economy* 104(6): 1297–1313.

Gneezy, Uri, and Aldo Rustichini. 2000a. "A Fine is a Price." *Journal of Legal Studies* 29(1): 1–17.

Gneezy, Uri, and Aldo Rustichini. 2000b. "Pay Enough or Don't Pay at All." *Quarterly Journal of Economics* 15(3): 791–810.

Hananel, Sam. 2007. "Lawmaker Wants to Ban Hill Line Standers." *Washington Post,* October 17. http://www.washingtonpost.com/wp-dyn/content/article/2007/10/17/AR2007101701034.html.

Hirsch, Fred. 1976. *The Social Limits to Growth.* Cambridge, MA: Harvard University Press.

Janssen, Maarten C. W., and Ewa Mendys-Kamphorst. 2004. "The Price of a Price: On the Crowding Out and In of Social Norms." *Journal of Economic Behavior and Organization* 55(3): 377–95.

Kunreuther, Howard, and Doug Easterling. 1996. "The Role of Compensation in Siting Hazardous Facilities." *Journal of Policy Analysis and Management* 15(4): 601–622.

Lerer, Lisa. 2007. "Waiting for Good Dough." *Politico,* July 26. http://www.politico.com/news/stories/0707/5109.html.

Levitt, Steven D., and Stephen J. Dubner. 2006. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything,* revised and expanded edition. New York: William Morrow.

Liptak, Adam. 2013. "Supreme Court Spectator Line Acts as a Toll Booth." *New York Times,* April 15. http://www.nytimes.com/2013/04/16/us/supreme-court-spectator-line-acts-as-a-toll-booth.html?_r=0.

Liu, Melinda. 2008. "China's One Child Left Behind." *Newsweek*, January 19. http://www.thedailybeast.com/newsweek/2008/01/19/china-s-one-child-left-behind.html.

Ming'ai, Zhang. 2007. "Heavy Fines for Violators of One-Child Policy." China.org.cn, September 18. http://www.china.org.cn/english/government/224913.htm.

Montopoli, Brian. 2004. "The Queue Crew." *Legal Affairs*, January/February. http://www.legalaffairs.org/issues/January-February-2004/scene_montopoli_janfeb04.msp.

Moore, Malcolm. 2009. "China's One-Child Policy Undermined by the Rich." *Telegraph* (London), June 15. http://www.telegraph.co.uk/news/worldnews/asia/china/5540739/Chinas-one-child-policy-undermined-by-the-rich.html.

O'Connor, Sarah. 2009. "Packed Agenda Proves Boon for Army Standing in Line." *Financial Times*, October 13. http://www.ft.com/intl/cms/s/0/5fa69a24-b78f-11de-9812-00144feab49a.html.

O'Hare, Michael. 1977. "'Not on *My* Block You Don't': Facility Siting and the Strategic Importance of Compensation." *Public Policy* 25(4): 407–458.

Palmeri, Tara. 2009. "Homeless Stand in for Lobbyists on Capitol Hill." CNN Politics.com, July 13. http://edition.cnn.com/2009/POLITICS/07/13/line.standers/.

Robbins, Lionel. 1932. *An Essay on the Nature and Significance of Economic Science.* London: Allen and Unwin.

Robertson, Dennis H. 1954. "What Does the Economist Economize?" Speech delivered at Columbia University, May 1954. (Reprinted in 1956 in Dennis H. Robertson, *Economic Commentaries,* Staples Press, London, 147–54.)

Roth, Alvin E. 2007. "Repugnance as a Constraint on Markets." *Journal of Economic Perspectives* 21(3): 37–58.

Rousseau, Jean-Jacques. 1762[1973]. *The Social Contract.* Translated by G. D. H. Cole. London: J. M. Dent & Sons.

Sandel, Michael J. 2012. *What Money Can't Buy: The Moral Limits of Markets.* New York: Farrar, Straus and Giroux.

Singer, Peter. 1973. "Altruism and Commerce: A Defense of Titmuss against Arrow." *Philosophy and Public Affairs* 2(3): 312–20.

Schuck, Peter H. 1994. "Share the Refugees." *New York Times*, August 13. http://www.nytimes.com/1994/08/13/opinion/share-the-refugees.html.

Schuck, Peter H. 1997. "Refugee Burden-Sharing: A Modest Proposal." *Yale Journal of International Law*, vol. 22, pp. 243–97.

Smith, Jada F. 2013. "A Line at the Supreme Court, and Some are Paid to Be There." *New York Times blog*, March 24. http://thecaucus.blogs.nytimes.com/2013/03/24/days-early-a-line-forms-at-the-supreme-court/.

Summers, Lawrence H. 2003. "Economics and Moral Questions." Morning Prayers address, Memorial Church, September 15. Reprinted in *Harvard Magazine*, November–December 2003. Also available at http://www.harvard.edu/president/speeches/summers_2003/prayer.php.

Titmuss, Richard Morris. 1971. *The Gift Relationship: From Human Blood to Social Policy.* New York: Pantheon.

Waldfogel, Joel. 1993. "The Deadweight Loss of Christmas." *American Economic Review* 83(5): 1328–36.

Waldfogel, Joel. 2009. *Scroogenomics: Why You Shouldn't Buy Presents for the Holidays.* Princeton University Press.

Xinhua News Agency. 2008. "Beijing to Fine Celebrities Who Break 'One Child' Rule." January 20. http://english.sina.com/china/1/2008/0120/142656.html.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 141–164

Reclaiming Virtue Ethics for Economics

Luigino Bruni and Robert Sugden

Economists have made use of, and have contributed to the development of, many branches of moral theory, including utilitarianism, social contract theory, libertarianism, and maximin and capability theories of justice. In contrast, virtue ethics—the study of moral character—has been an important strand in moral philosophy for literally thousands of years, but has received little attention from contemporary economists. That neglect has not been reciprocated. A significant body of philosophical work in virtue ethics is associated with a radical critique of the market economy and of economics. Expressed crudely, the charge sheet is this: The market depends on instrumental rationality and extrinsic motivation; market interactions therefore fail to respect the internal value of human practices and the intrinsic motivations of human actors; by using market exchange as its central model, economics normalizes extrinsic motivation, not only in markets but also (in its ventures into the territories of other social sciences) in social life more generally; therefore economics is complicit in an assault on virtue and on human flourishing. We will argue that this critique is flawed, both as a description of how markets actually work and as a representation of how classical and neoclassical economists have understood the market. We will show how the market and economics can be defended against the critique from virtue ethics.

Crucially, our response to that critique will be constructed *using the language and logic of virtue ethics*. In this respect, it is fundamentally different from a response that many economists would find more natural—to point to the enormous benefits,

■ *Luigino Bruni is Professor of Economics, LUMSA University, Rome, Italy. Robert Sugden is Professor of Economics, University of East Anglia, Norwich, United Kingdom. Their emails are l.bruni@lumsa.it and r.sugden@uea.ac.uk.*

http://dx.doi.org/10.1257/jep.27.4.141 doi=10.1257/jep.27.4.141

including income and leisure that can be devoted to intrinsically motivated activities, that we all enjoy as a result of the workings of markets, and to the essential role of economics in explaining how markets work. Set against those benefits, it can be argued, questions about whether market motivations are virtuous are second-order concerns that economists can safely leave to moral philosophers. Thus, for example, responding to the philosopher Michael Sandel's objection to markets in carbon dioxide emissions on the grounds that they express nonvirtuous attitudes to the environment (Sandel 2012, pp. 72–76), Coyle (2012) writes, "I would rather see an effective scheme to reduce greenhouse gas emissions, but then I'm an economist." We are economists too, and have some sympathy with such sentiments. Nevertheless, the virtue-ethical critique of economics is gaining credence in public debate. Many people see it as providing intellectual support for popular attitudes of opposition to capitalism and globalization, and of hostility to economics as a discipline. Philosophically, the critique is grounded in an ancient and respected tradition of ethical thought: it is not something that economics can or should simply brush aside. Our premise is that economics needs a response to this critique that takes virtue ethics seriously.

Another possible reply, made for example by van Staveren (2009) and Besley (2013), is that, *in their critique of economics*, the virtue ethicists fail to recognize the diversity of the discipline. Economics has never been unanimous or unconditional in advocating markets; indeed, it is possible to read the development of normative economics in terms of a continually expanding catalog of market failures and their remedies. In particular, a recent development in economics has been the growth of a literature in which concepts of intrinsic motivation are used to explain individual behavior. Although this work is not explicitly virtue-ethical in the normative sense, it allows economics to model a "crowding-out" mechanism that is similar to the virtue ethicists' account of the corrupting effects of markets. However, pointing to the diversity of economics merely deflects the virtue-ethical critique from economics in general to a particular but surely major tradition of economic thought—that liberal tradition that understands the market as a domain in which socially desirable consequences emerge from the pursuit of private interests. In contrast, our response meets the critique head-on. We aim to show that economists can teach about and defend the market without standing for nonvirtue against virtue.[1]

The logic of our response requires that we use the modes of argument of virtue ethics. We write as philosophically and historically inclined economists, hoping to be read both by philosophers and by our fellow economists. For the benefit of the economists and with apologies to the philosophers, we assume no prior knowledge of virtue ethics on the part of the reader. Thus, we begin with a brief introduction

[1] In this respect, our approach has more in common with McCloskey's (2006) account of the "bourgeois virtues." However, our analysis is more systematic and economics-specific than McCloskey's imaginative but discursive exploration of the seven virtues of traditional Christian thought and their role in economic life.

to virtue ethics. We then describe some prominent critiques of the market that are grounded in virtue ethics and in the related economic and psychological literature on intrinsic motivation.

Following this introduction, we use the methods of virtue ethics to develop a conception of market virtue that is consistent with many classical and neoclassical economists' accounts of how markets work and of what purposes they serve. Our central idea is that the public benefits of markets should be understood as the aggregate of the mutual benefits gained by individuals as parties to voluntary transactions, and that the market virtues are dispositions that are directed at this kind of mutual benefit. For a virtuous market participant, mutual benefit is not just a fortunate by-product of the individual pursuit of self-interest: he or she *intends* that transactions with others are mutually beneficial.

Using this idea, we identify some specific character traits that have the status of virtues within the domain of the market. Our list of market virtues (which we do not claim is complete) includes universality, enterprise and alertness, respect for the tastes of one's trading partners, trust and trustworthiness, acceptance of competition, self-help, non-rivalry, and stoicism about reward. We will argue that these market virtues, grounded on ideas of reciprocity and mutual benefit, are closely associated with virtues of civil society more generally. It is therefore a mistake to think that the market is a virtue-free zone, or that the character traits that best equip individuals to flourish in markets are necessarily corrosive of virtue in other domains of life.

The idea that economic agents should understand their interactions as mutual assistance is characteristic of a tradition of natural-law philosophy from which mainstream economic thought turned away in the later eighteenth century. Nevertheless, as we will show, the idea that mutual benefit is in some sense the purpose of the market is implicit in the writings of many major economists from the eighteenth century to the present day. The specific market virtues that we present feature in some canonical accounts of the desirable properties of markets. In this sense, our paper can also be read as an attempt to reconstruct a submerged current of virtue-ethical thought in economics.

What is Virtue Ethics?

The central concern of virtue ethics, broadly interpreted, is with *moral character*—with what sort of person one is and should be. Virtues are acquired character traits or dispositions that are judged to be good. Crucially, virtues are not judged to be good *because* they tend to induce actions that, for other moral reasons, are good or right. In virtue ethics, actions are judged to be good because they are in character for a virtuous person—they are constitutive of living well, of "flourishing." A morally well-constituted individual cultivates virtues not as rules of thumb for moral action, but because such virtues are characteristic of the kind of person she is or wants to be.

Aristotle's *Nicomachean Ethics* (c. 350 BC [1980]) is traditionally seen as the founding text of virtue ethics. Aristotle's account of virtue begins from the idea that

within any "practice" or domain of life, goodness is understood in relation to the *telos* (literally, "end" or "purpose") of that domain—"that for whose sake everything is done." For example, Aristotle (Book 1, section 1) treats medicine as a domain whose *telos* is "health" and military strategy as a domain whose *telos* is "victory." In relation to a given domain, an acquired character trait is a virtue to the extent that the person who possesses it is thereby better able to contribute to the *telos* of that domain. The underlying idea is that human happiness or flourishing (*eudaimonia*) requires that people are oriented towards their various activities in ways that respect the intrinsic ends of the domains to which those activities belong.

How is the *telos* of a domain determined? Aristotle seems to think of the *telos* as a natural fact that can be ascertained by intuition, but many modern virtue ethicists favor a communitarian approach. This approach, exemplified by the work of MacIntyre (1984), understands the concept of flourishing as internal to specific communities and cultural traditions. Thus, to identify the *telos* of a practice, one must discover the meaning of that practice within the community of practitioners. In this view, a claim about the *telos* of an practice is not just the expression of a personal value judgement; it involves some (perhaps creative) interpretation of what is already there (Sandel 2009, pp. 184–192, 203–207; Anderson 1993, p. 143). As Sandel (p. 98) puts it, "we identify the norms appropriate to social practices by trying to grasp the characteristic end, or purpose, of those practices."

There is much common ground between Aristotelian virtue ethics, with its emphasis on the intrinsic value of practices, and those strands of modern "positive psychology" that emphasise the importance of intrinsic motivation for human happiness, in particular the *self-determination theory* of Deci and Ryan (1985). In this theory, the analog of flourishing is a concept of psychological health or well-being. The core hypothesis is that individual autonomy is a source of psychological well-being, and thus that human flourishing is linked with authenticity and self-realization. In Ryan and Deci's (2000) taxonomy of motivation, there is a continuum from "amotivation," through increasingly autonomous forms of "extrinsic motivation," to the full autonomy of "intrinsic motivation." A person who is extrinsically motivated performs an activity "in order to obtain some separable outcome." Extrinsic motivations can become more "internal" (and thereby more autonomous) to the extent that the individual has a sense of having chosen the objective on which he acts and endorsed its value. But an intrinsically motivated person performs an activity "for its inherent satisfactions rather than for some separable consequence"; such a person "is moved to act for the fun or challenge entailed rather than because of external prods, pressures, or rewards" (pp. 56–60). Thus, the analog of *telos* is the meaning that an individual attaches to an activity when he sees the activity as an end in itself.

The Instrumentality of the Market: The Critique from Virtue Ethics

In critiques of economics by virtue ethicists, a recurring theme is that markets rely on extrinsic and thereby nonvirtuous motivations. This idea can also be traced

back to Aristotle, who wrote (Book 1, §5): "The life of money-making is one undertaken under compulsion, and wealth is evidently not the good we are seeking; for it is merely useful and for the sake of something else." This sentence makes two claims that are echoed in critiques of economics made by modern virtue ethicists. The first claim is that when individuals participate in markets, they show a lack of autonomy—they act *under compulsion*. The suggestion seems to be that a truly autonomous person would not need to seek wealth (perhaps because he would already have as much as he needed without having to seek for it).[2] The second claim is that the motivation for economic activity is extrinsic and thereby of an inferior kind—the things that economic activity can achieve are *merely useful and for the sake of something else*.

Here, we will focus on how three prominent contemporary virtue ethicists apply these themes in their writings about economics and the market. Of these criticisms of the market, MacIntyre's (1984) book *After Virtue* is the most radical. Taken literally, MacIntyre's elegant despair has no real point of contact with modern economics. But precisely because it takes the critique of the instrumentality of markets to its logical conclusion, it offers a useful point of reference. MacIntyre (p. 187) presents an account of morality that is built on the concept of a *practice*. A practice is a "coherent and complex form of socially established cooperative human activity" which realizes "goods internal to that form of activity." A practice has intrinsic ends, and internal standards of excellence that make sense in relation to those ends. Associated with the practice are certain acquired character traits that assist in the achievement of excellence, or in recognizing and internalizing communal understandings of the meaning of the practice. The traits can be viewed as the virtues of the practice.

For MacIntyre (1984), a person who fails to treat an activity as a practice with an internal end is failing to display virtue—either because the activity falls within a practice whose internal ends the person is failing to respect, or because the activity is of such a morally impoverished and instrumental kind that it is not a practice at all—MacIntyre's (p. 187) questionable example of an activity that does not count as a practice is bricklaying. This way of thinking immediately makes markets morally suspect. The market motivation of creating goods *for exchange* conflicts with the idea that activities, or the goods that they realize, are ends *in themselves*. Thus, according to MacIntyre, the exposure of a practice to market forces is liable to corrupt its excellences and virtues. MacIntyre does not quite claim that practices can never coexist with market exchange. For example, he maintains that portrait painting from the time of Giotto to that of Rembrandt was a practice with internal ends and standards of excellence. He recognizes that many excellent painters were also able to achieve (and presumably cared about) goods external to the practice of art, including the income they were able to earn from the sale of their services (pp. 189–190). The suggestion is that the corrupting tendencies of the market can be contained only

[2] In a witty account of the history of Western intellectuals' criticisms of capitalism, Alan Kahan (2010, p. 31) presents the "Three Don'ts" of anti-capitalism. The first is "Don't make money (just have it)".

to the extent that individuals are at least partially motivated by the internal ends of practices (as, in MacIntyre's account, the great painters were).

However, as MacIntyre (1984) recognizes, practices as he understands them are not, and cannot be, characteristic of ordinary economic life in the world in which we live. Treating the household as a paradigm case of communal life, he argues that "[o]ne of the key moments in the creation of modernity" occurs when production moves from the household to an impersonal domain of "means-ends relationships" (p. 227). This thought reflects the presupposition that production for exchange belongs to the domain of external goods. The implication is that an economy of practices cannot make effective use of comparative advantage and the division of labor. MacIntyre's ultimate response to economic reality is a yearning for an imagined and ill-defined economy of communal production somehow devoid of the hierarchical power relationships found in real historical economies.

Similar themes, developed in somewhat less unworldly forms, are prominent in the work of Anderson (1993) and Sandel (2009, 2012). These writers recognize, at times reluctantly, that markets are a necessary part of social organization. But they argue that the instrumental logic of markets is liable to corrupt virtues that are proper to other domains of social life, and that it is therefore appropriate for the state to impose limits on the scope of markets.

Thus, the first sentence of Anderson's *Value in Ethics and Economics* (1993) is: "Why not put everything up for sale?" This rhetorical question signals several elements of her position: to allow all areas of social life to be governed by market relationships would be morally objectionable; this truth ought to be obvious to a morally aware reader; but some opinion-formers *do* want to put everything up for sale, and their arguments need to be countered. More specifically, the people against whom she is arguing fail to understand that there are "ways we ought to value people and things that can't be expressed through market norms" (pp. xi–xiii).

Anderson (1993) proposes a "pluralist theory of value" in which different kinds of goods ought to be valued in different ways (p. 12). She tries to delimit the proper scope of the market by identifying the norms that are characteristic of market relations, and the corresponding class of goods that are properly valued in terms of those norms. For Anderson, the ideal economic good is a "pure commodity." The mode of valuation appropriate to pure commodities is "use." She writes (p. 144): "Use is a lower, impersonal, and exclusive mode of valuation. It is contrasted with higher modes of valuation, such as respect. To merely use something is to subordinate it to one's own ends, without regard for its intrinsic value." This definition immediately introduces the Aristotelian ranking of intrinsic value over instrumental value. Anderson is presenting market norms as a kind of second-rate morality: the market's mode of valuation is *lower* than that of other domains of social life; it is *merely* use; it has *no regard for* intrinsic value. In this account, market norms are *impersonal* and *egoistic*. Impersonality is the idea that market transactions are viewed instrumentally: each party to a transaction considers it only as a means to the satisfaction of his own ends. Egoism is the idea that those ends are defined in terms of self-interest.

Anderson (1993) acknowledges that market norms embody a moral ideal of "economic freedom." However, this ideal is presented in negative terms—as freedom from the kinds of moral constraints that one would face if one recognized the intrinsic value of goods, the obligations of personal relationships, and the potential validity of other people's judgements about value (pp. 144–146). Indeed, Anderson seems comfortable with the ideal of economic freedom only in the context of inessential but harmless consumer products. Accepting (if condescendingly) that "the market . . . also has its proper place in human life," her examples of goods that properly belong to the domain of economic freedom are "the conveniences, luxuries, delights, gadgets, and services found in most stores" (166–67). There is no mention of the role of the market in supplying private goods like food, clothing, fuel, and shelter, on which we all depend for our survival.

Anderson (1993) develops her critique of the instrumentality of the market by considering the intrinsic value of the goods and services provided by professional workers such as doctors, academics, athletes, and artists. Like MacIntyre (1984) in his discussion of portrait painters, Anderson recognizes that professionals can be intrinsically motivated even though they produce for sale. But she argues (pp. 147–150) that the norms of the market can conflict with "the norms of excellence internal to their professional roles." The result is that, when professionals sell their services, intrinsically valuable goods are "partially commodified." She does not claim that commodification is wholly undesirable, but the thrust of her argument is that the internal goals of professional practices must be partially insulated from the extrinsic motivations that are fostered by markets. If necessary, taxpayers should bear some of the costs of this insulation, for example through subsidies to the arts and to pure research.

Sandel (2009) develops a different but complementary critique of the market, focusing on the virtue ethics of justice.[3] Like MacIntyre, he works with a concept of social practices; each practice has its Aristotelian *telos* and its associated excellences and virtues. However, Sandel's concern is less with the cultivation of proper attitudes towards goods and practices, and more with how individuals are honored and rewarded for showing appropriate virtues. Justice, for Sandel, is about "giving people what they deserve." That requires judgements about "what virtues are worthy of honor and reward, and what way of life a good society should promote" (p. 9).

Sandel (2009) begins his book by describing some recent issues of public debate in America, intended to support his claim that virtue ethics is alive and well in ordinary political discourse. Two of these issues concern what Sandel sees as the ethical limitations of the market. The first issue is the conduct of those firms that charged scarcity prices for such goods as motel rooms, emergency repairs, and

[3] In a more recent book, Sandel (2012) presents an argument about the "moral limits of markets." His paper in this issue takes up some of these arguments. As he acknowledges (p. 208, note 18), this argument is similar to that of Anderson (1993). Sandel sees economics as complicit in the inappropriate propagation of "market values." Sandel is less precise than Anderson in explaining what those values are, but it is clear that he sees them in opposition to the civic virtues of "social solidarity," including "shar[ing] in a common life," and "car[ing] for the common good" (pp. 128, 203).

bottled water in the aftermath of Hurricane Charley in Florida in 2004. At the time, some economists argued that market-clearing prices promote efficiency in the use of resources, and that this truth is not invalidated by hurricanes. Sandel sides with the opinion that this kind of "price gouging" should be illegal. His reason is an application of virtue ethics: the firms that charged scarcity prices were motivated by greed; since greed is "a vice, a bad way of being," the state should discourage it (pp. 7–8). The second issue is the remuneration of senior corporate executives. Sandel asks whether the chief executive officers of large American corporations deserved the payments they received in the years leading up to 2008, when their firms were generating large profits. We are invited to conclude that effort and talent are qualities that are worthy of reward in business, but that when the market rewards executives for profits that are *not* attributable to effort or talent, a principle of justice is being violated (pp. 12–18). The message from both examples, developed over the course of the book, is that the market generates incomes that are not properly aligned with the virtues of the people who receive them.

To an economically trained reader, these critiques of economics and the market often seem divorced from the reality of everyday economic life. MacIntyre (1984) and Anderson (1993) seem to find it hard to find moral significance in the ordinary useful jobs by which most people earn their livings. Sandel (2009) seems to find it hard to come to terms with the fact that market rewards depend on luck as well as talent and effort. We will argue that virtue ethicists are failing to find virtue in markets because they are not seeing the market as a practice in its own right.

Intrinsic Motivation and Economics

Although there is little explicit analysis of virtue in modern economics, a large literature in behavioral economics echoes Anderson's (1993) argument about the importance of insulating intrinsic motivation from contamination by the market (for example, Gneezy, Meier, and Rey-Biel 2011). The concept of intrinsic motivation has come to economics from social psychology, and particularly from Ryan and Deci's self-determination theory. That theory has strong undertones of Aristotelian hostility to markets. Recall that according to Ryan and Deci's (2000) definition, an intrinsically motivated person does an activity for its inherent satisfactions *rather than for some separable consequence*; such a person is not motivated by external prods, pressures, *or rewards*. Notice how this definition excludes all ordinary market activities. It should be no surprise that the economic literature on intrinsic motivation has been seen as supporting the virtue-ethical critique of markets (for example, Sandel 2012, pp. 64–65, 113–120).

An important hypothesis in this psychological literature is that external rewards can *crowd out* intrinsic motivation (Deci 1971; Lepper and Greene 1978); a parallel hypothesis in relation to social policy is due to Titmuss (1970). Titmuss's famous example is the effect of introducing financial incentives for blood donors. In a regime in which donors are entirely unpaid, blood donation is motivated by altruism,

reciprocity, or public spirit. If financial incentives are introduced into such a setting, this prompts the thought that people who supply blood may be self-interested sellers rather than altruistic donors. This can undermine the sense of would-be donors that giving blood is a morally significant and socially valued act, and so lead to a reduction in the supply of blood. A similar interpretation is now often given for the much-discussed finding that fines for lateness in collecting children from a day-care center led to an increase in the incidence of lateness (Gneezy and Rustichini 2000a).

The economic implications of the hypothesis of motivational crowding-out were first explored by Frey (1994, 1997).[4] Defining intrinsic motivation in essentially the same way as Deci and Ryan do, Frey (1997, p. 2) maintains that it is "neither possible nor desirable to build a society solely or even mainly on monetary incentives"; intrinsic motivation has an essential role to play.

Within economics, there is growing interest in theorizing about how intrinsic motivation can be shielded from market forces. One approach is summarized in the slogan "getting more by paying less." Suppose there is some occupation, say nursing, in which workers are better able to provide the services that their employers value if they are intrinsically motivated to pursue the internal ends of that occupation—if, in Ryan and Deci's (2000) terminology, they are attracted by its "inherent satisfactions" and "challenges." Viewed in the standard conceptual framework of economics, a person with such a motivation for nursing has a lower reservation wage for working as a nurse than for working in other occupations. So employers may be able to separate the better workers from the worse by offering *low* wages—they can get more by paying less (Brennan 1996; Katz and Handy 1998; Heyes 2005). When a person accepts the low wages of an employer who is looking for intrinsic motivation, she signals to herself and to others that she is intrinsically motivated. So there need be no crowding-out effect.

We suspect that many readers will share our unease about this argument. Nelson (2005) formulates this unease by raising two objections. First, because low wages may screen out intrinsically motivated individuals who need to support themselves and their families, access to intrinsically rewarding occupations may be restricted to people with private incomes or well-off partners or parents. Second, when social norms treat self-sacrifice as a characteristic virtue of "caring" occupations such as nursing, they act as a cover for, and an incitement to, exploitation. These objections draw attention to a questionable assumption of the "getting more by paying less" argument—that a person is virtuous or authentic to the extent to which that person is willing to sacrifice material rewards in the pursuit of intrinsic ends. In a model in which all motivations are represented as properties of individuals' preferences, that assumption is almost unavoidable, since an individual's preference for "consuming"

[4] It is only very recently that economists have taken this hypothesis seriously. Titmuss's (1970) work was well-known to economists in the 1970s, but his crowding-out argument was viewed skeptically (for example, Arrow 1972). Gneezy and Rustichini (2000a, 2000b) discussed motivational crowding-out as a possible explanation of their findings, but favored a more conventional economic interpretation in terms of incomplete contracts.

an intrinsic good is defined in terms of how much of other goods she is willing to give up in exchange. However, it is not an essential part of a virtue-ethical approach in which the exercise of virtue is associated with flourishing rather than sacrifice, nor of a decision-theoretic approach in which intentions for mutual benefit are represented as "team reasoning" (Bruni and Sugden 2008).

Folbre and Nelson (2000) suggest that the crowding-out problem can be countered by separating the payment of intrinsically motivated workers from the specific services they provide, so that payment can be construed as an *acknowledgement* of intrinsic motivation rather than as one side of a market exchange. The implication seems to be that authentic caring is compromised if carers and cared see their relationship as that of seller and buyer. There is another echo here of the Aristotelian idea that market relationships are instrumental and thereby nonvirtuous.

But how is the payment of service suppliers to be separated from exchange relationships? One possibility is to use gift relationships. Consider the case of restaurant waiters who are paid less than the market wage, but with the expectation that their earnings will be supplemented by tips from customers. Perhaps this practice supports dispositions towards friendliness and efficiency that restaurant owners value in their waiters and find costly to monitor, but one might think that it impairs rather than supports the waiter's sense of autonomy.

A different model (and probably the one that Folbre and Nelson 2000 have in mind) is that of a salaried professional. Think of the role of the tenured academic in a well-financed university, as that role used to be (and sometimes still is) understood. The academic is awarded tenure in the expectation of a continuing intrinsic motivation to pursue excellence in teaching and research, but is subject to only the lightest of monitoring. He is paid a good salary that has no direct relationship to the services he provides, but is seen as expressing a social valuation of the excellence that is expected. Actual excellence in teaching will be rewarded by the gratitude of students; excellence in research, by the respect of peers. This kind of separation of payment from services rendered can give professionals an enviable degree of autonomy; and it can protect whatever intrinsic motivation they have from crowding-out effects. But it also insulates them from pressures to respond to the interests of the people to whom their services are being provided. Just as the waiter loses autonomy in having to depend on the good will of the customer, so does the client in having to depend on the professional's intrinsic motivation.

These examples illustrate the difficulty of shielding intrinsic motivation from the supposedly corrosive effects of exchange relationships. These difficulties have a common source: it is inherent in the concept of intrinsic motivation that an individual's autonomy and authenticity are compromised whenever she enters into exchange relationships, but such relationships are fundamental to the workings of any economy that relies on comparative advantage and the division of labor. The literature of intrinsic motivation invites us to aspire to the ideal of an economy in which everyone's actions and efforts are coordinated to realize gains from trade, but in which no one is actually motivated to seek those gains. This ideal seems as profoundly unrealistic as MacIntyre's (1984) imaginary world of an economy built

on practices. If we are to reconcile the ideas of virtue and authenticity with real economic life, we need a way of understanding market relationships that acknowledges that gains from trade are not realized by accident: they are realized because individuals seek them out.

The *Telos* of the Market

In the literature of virtue ethics, the market is seen as opposed to virtue and authenticity because behavior in markets fails to respect intrinsic value. Intrinsic value is attributed to practices in which goods are produced—for example, the practices of art, scientific enquiry, or nursing—as well as to *nonmarket* practices which transfer goods between individuals, like gift-giving and the honoring of excellence. But there is a reluctance to treat the market as a practice in its own right, with its own forms of intrinsic value and authenticity. We suggest that the first step in a virtue ethics of the market is to think of the market in this way.

It must be said that economists have been partly responsible for the difficulty that virtue ethicists have had in seeing the market as a practice. After all, generations of economists have pictured the market as a domain in which socially desirable consequences emerge as unintended consequences of individuals' pursuit of their private interests. Two famous expressions of this idea are due to Adam Smith (1776 [1976], pp. 26–27, 456)—the assertion that "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest," and the description of the merchant who "intends only his own gain, [but is] led by an invisible hand to promote an end which was no part of his intention." In Smith's theory of markets, the primary motivation for action is self-love, even though *in fact* everyone's self-interested actions combine to create benefits for all. To say this is not to assert that Smith shared his successors' lack of interest in virtue ethics. The virtues of sympathy and benevolence are important in Smith's (1759 [1976]) earlier work *The Theory of Moral Sentiments*, even though they play only minor roles in his economic analysis. And for Smith, self-interest expressed within the rules of a commercial society is not *opposed to* virtue. To the contrary, character traits associated with the pursuit of long-term self-interest, particularly prudence, temperance, and self-command, are virtues (on this, see Hirschman 1997, especially pp. 18–19). We take it as given that such traits are indeed virtues of economic life, but our focus will be on how, within a market economy, individuals relate to one another.

Can the market be viewed as a practice with its own intrinsic values? In terms of MacIntyre's (1984) definition of practices, the market is certainly a coherent and complex form of socially established cooperative human activity. But does it have moral goods that are internal to itself? Does it have internal standards of excellence? From the standpoint of virtue ethics, the answer to these questions begins by asking: "What is the *telos* of the market?" For many readers (and perhaps particularly for those who are economists), it will be tempting to reply that the presupposition of the question is either false or meaningless. We ask such readers

to set aside their skepticism for a moment, and to translate this question into common-sense terms. What is the characteristic end or purpose or *raison d'être* of the market? How would you describe, in the simplest and most general terms, what markets do that is valuable? If you had to write a mission statement for the market, what would it say?

Thoughtful economists have offered answers to such questions. For example, Friedman (1962, p. 13) wrote that, in relation to the problem of coordinating economic activity, "the technique of the market place" is "voluntary cooperation of individuals." Buchanan and Tullock (1962, p. 103) wrote: "The *raison d'être* of market exchange is the expectation of mutual gains." We are not claiming here that Friedman, Buchanan, and Tullock are virtue ethicists. All we are attributing to them is the idea that markets have a point or purpose, and that that purpose is mutual benefit. Most economists, faced with our questions, would probably invoke in one way or another the idea of mutual benefit or gains from trade through voluntary transactions.

If economists were asked to nominate one simple diagrammatic representation of a market, the "Edgeworth box" would surely be one of the commonest choices, and the point of that diagram is to understand markets as networks of mutually beneficial voluntary transactions. Edgeworth (1881, pp. 16–17) himself, in a famous passage in which he declares that the first principle of economics is that every agent is activated only by self-interest, distinguishes between "war" and "contract," differentiated by whether "the agent acts without, or with, the consent of others affected by his actions"; his analysis of competitive markets is presented as an analysis of contract. If economists were asked to nominate a theorem to represent the market in its best light, many would opt for the first fundamental theorem of welfare economics, which is essentially equivalent to showing that in competitive equilibrium, no opportunities for mutually beneficial transactions, however complex, remain unexploited. Another strong contender would be Ricardo's (1817, Ch. 7) comparative advantage theorem, which shows that there are typically opportunities for gains from trade between any pair of countries (and by extension, any pair of individuals), whatever their respective endowments and productivity.

How else might one answer our question about the *telos* of the market? One obvious alternative answer is that the *telos* of the market is wealth creation: after all, the founding text of economics is called *The Wealth of Nations*. But even for the author of that text, the fundamental mechanism by which wealth is created is the division of labor and the extension of the market, and the division of labor is the consequence of the human propensity "to truck, barter and exchange one thing for another" (Smith 1776 [1976], p. 25). Other economists have emphasised how the market creates wealth by exploiting comparative advantage (Ricardo 1817), the division of knowledge (Hayek 1948), and increasing returns to scale (Marshall 1920, pp. 222–242; Arrow 1984, p. 188); but all of these mechanisms operate through mutual gains from trade. Another possible answer is that the *telos* of the market is economic freedom. The association between the market and freedom is a recurring theme in economics; famous expositors of this idea include Mill (1848 [1910]), Marshall (1920, p. 8), Hayek (1948), and Friedman (1962). But economic freedom

is not the freedom of each person to get what he wants *tout court*; it is his freedom to use his own possessions and talents as he sees fit and to trade with whoever is willing to trade with him.

We suggest that the common core of these understandings of markets is that markets facilitate mutually beneficial voluntary transactions. Such transactions can be seen as valuable because individuals want to make them, because they satisfy individuals' preferences, because they create wealth, and because the opportunity to make them is a form of freedom. We therefore propose to treat mutual benefit as the *telos* of the market.

Market Virtues

On the supposition that the *telos* of the market is mutual benefit, a market virtue in the sense of virtue ethics is an acquired character trait with two properties: possession of the trait makes an individual better able to play a part in the creation of mutual benefit through market transactions; and the trait expresses an intentional orientation towards and a respect for mutual benefit. In this section, we present a catalog of traits with these properties, without claiming that our catalog is exhaustive.

According to the logic of virtue ethics, such traits are properly or consistently viewed as praiseworthy within the practice of the market, when that practice is understood as directed at mutual benefit. Thus, we should expect the traits in our catalog to have been evaluated favorably in the tradition of liberal economic thought from which we have distilled the *telos* of mutual benefit. We maintain that this is the case, and will point to illustrative examples. Recall that virtue ethicists claim to uncover the virtues of practices by philosophical reflection, and not simply by sociological observation. It is in the spirit of such enquiry to look to thoughtful economists as well as to market participants for insights into the nature of market virtues.

We will not claim that all market participants display the market virtues. (The logic of virtue ethics does not require that kind of implausibility: virtue ethicists can, for example, describe bravery as a military virtue without asserting that all soldiers are brave.) But we do maintain that the market virtues are broadly descriptive of traits that many people, including people who are successful in business, display when they participate in markets. Readers who are accustomed to equating virtue with self-sacrifice may suspect that this claim is overoptimistic, but we repeat that such an equation is alien to virtue ethics. It is fundamental to the classical and neoclassical understanding of markets that, under normal circumstances, each party to a market transaction benefits from involvement in it. Thus, a disposition to seek mutual benefit in markets will normally incline individuals towards the kinds of individually beneficial behavior that economic theory has traditionally described. Our account of market virtue is not a new theory of nonselfish behavior. It is a description of a distinctive moral attitude to market relationships—an attitude characterized not by altruism but by reciprocity.

Universality

Our first market virtue is universality—the disposition to make mutually beneficial transactions with others on terms of equality, whoever those others may be. If the market is to be viewed as an institution that promotes the widest possible network of mutually beneficial transactions, universality has to be seen as a virtue. Its opposites—favoritism, familialism, patronage, protectionism—are all barriers to the extension of the market.

It is intrinsic to the virtue of universality that market relations are not based on personal ties of kinship, community, friendship, or gratitude—the kind of ties that Anderson (1993) sees as characteristic of "higher" modes of valuation. As Smith (1776 [1976], p. 27) makes clear in his account of how we get our dinners, it is because the market is based on free horizontal relations between equals that it allows us to satisfy our economic needs with independence and self-respect. This independence can be compromised if economic transactions depend on relations other than mutual benefit. However, this is not to say that market relations must be impersonal in the sense that each party treats the other merely as a means to an end. When trading partners intend their transactions to be mutually beneficial, it is possible for their relations to have the characteristics of friendliness and goodwill that we (Bruni and Sugden 2008) describe as "fraternity."

Friedman (1962, pp. 108–118) identifies another valuable aspect of universality when he argues that market forces tend to counter racial and religious prejudice. His leading example is the case of the Jews of medieval Europe, who (between outbreaks of outright persecution) were able to survive in a hostile social environment by working on their own account and trading with non-Jews. For Friedman, it must be said, universality is a desirable but unintended consequence of the pursuit of self-interest, rather than a virtue in our sense; but nonetheless, the customer who chooses where to shop on the basis of price and quality rather than the shopkeeper's religion can be thought of as exhibiting a market virtue.

Enterprise and Alertness

If the *telos* of the market is mutual benefit, enterprise in seeking out mutual benefit must be a virtue. Discovering and anticipating what other people want and are willing to pay for is a crucial component of entrepreneurship. (Think of Freddie Laker's pioneering of no-frills aviation, Steve Jobs's development of graphical user interfaces, or Art Fry's discovery of the commercial potential of the Post-it.) Successful entrepreneurship requires empathy and imagination, as Jevons (1871 [1970], pp. 102–103) recognized in one of the founding texts of neoclassical economics: "Every manufacturer knows and feels how closely he must anticipate the tastes and needs of his customers: his whole success depends on it."

The virtue of alertness to *mutual* benefit applies to both sides of the market: for mutual benefit to be created, the alertness of a seller has to engage with the alertness of a buyer. Thus, the inclination to shop around, to compare prices, and to experiment with new products and new suppliers must be a virtue for consumers. Arguing that the law of one price has more application to wholesale than to retail

markets, Mill (1848 [1909], p. 441) wrote: "Either from indolence, or carelessness, or because people think it fine to pay and ask no questions, three-fourths of those who can afford it give much higher prices than necessary for the things they consume." Notice how Mill's empirical claim that well-off consumers are not inclined to search for the lowest prices is linked with moral criticism.

Respect for the Tastes of One's Trading Partners

One is more likely to succeed in making mutually beneficial transactions if one is disposed to respect the preferences of potential trading partners. The spirit of this virtue is encapsulated in the business maxim that the customer is always right. This virtue is closely related to the idea that market transactions are made on terms of equality, and opposed to the paternalistic idea that the relationship of supplier to customer is that of guardian to ward. It is also opposed to the idea of virtues based on intrinsic motivation, or on professional and craft standards. It is perhaps true (as MacIntyre 1984 and Anderson 1993 claim) that when professionals and craft workers sell their services, they are liable to compromise the standards of excellence that are internal to their respective practices, but that does not invalidate the proposition that producing what customers *do* want to buy is an aspect of a practice—the practice of the market—with its own standards of excellence and its own forms of authenticity. From this perspective, it is unsurprising that Smith (1776 [1976], pp. 758–764) favored the payment of university teachers by their students on a fee-for-service basis—a practice that gives the relationship between professional and client essentially the same status as that between shopkeeper and customer.

In speaking of *respect* for the preferences of trading partners, we mean something more than the recognition that satisfying those preferences is a source of profit. Consider a famous case in which this virtue is lacking. Gerald Ratner, the chief executive of a (then) successful low-price British jewelery business, made a speech in 1991 to the Institute of Directors in which he referred to his firm's products with the joke: "People say, 'How can you sell this for such a low price?' I say, 'because it's total crap.'" When this was reported in the press, the business lost £500 million in market value and eventually had to be relaunched with a new name—and Ratner lost his job (Ratner 2007). Notice that Ratner was not saying, as suppliers of lower-priced products often and quite properly do, that what he was selling was cheap and cheerful and aimed at those consumers for whom value for money was a priority. But nor, as we understand this story, was he confessing to taking advantage of some lack of information on the part of his customers, and so failing to return their trust: the *objective* properties of his products were transparent enough. He was expressing contempt for the tastes to which his business catered, and thereby for the idea that the relationship between supplier and customer is one of mutual benefit.

Trust and Trustworthiness

Because the monitoring and enforcement of contracts is often difficult or costly, dispositions of trust and trustworthiness (qualified by due caution against

being exploited by the untrustworthy) facilitate the achievement of mutual benefit in markets. If that is right, these dispositions must be market virtues.

The idea that markets rely on trust and trustworthiness has a long history in economics. Smith (1763 [1978], pp. 538–539) recognizes the importance of "probity" for the workings of markets and describes this trait as a "virtue." Significantly, Smith sees this virtue as consistent with long-term self-interest. He claims that it is most prevalent in the most commercial societies, and explains this observation by arguing that a reputation for probity is more valuable, the more one engages in trade. The idea that commercial transactions typically depend on an element of trust has continued to be recognized by leading economists, including Marshall (1920, p. 6) and Arrow (1972). Following the work of Akerlof (1982), trust relationships have featured in many economic models.

A recent public discussion about the role of trustworthiness in business was initiated by an open resignation letter written by a senior executive in Goldman Sachs and published in the *New York Times*. The executive, Greg Smith (2012, p. A27), wrote that the "culture" of Goldman Sachs had changed in a way that he could no longer identify with. At one time, "always doing right by our clients" had been at the heart of this culture, but now "I attend derivatives sales meetings where not one single minute is spent asking questions about how we can help clients. It's purely about how we can make the most possible money off of them." Like Adam Smith, and in the spirit of virtue ethics, Greg Smith argued that the virtue (or "culture") of trust was not opposed to long-term self-interest: "It astounds me how little senior management gets a basic truth: If clients don't trust you they will eventually stop doing business with you."

Acceptance of Competition

If the *telos* of the market is mutual benefit, a virtuous trader will not obstruct other parties from pursuing mutual benefit in transactions *with one another*, even if that trader would prefer to transact with one or another of them instead. The spirit of this virtue is expressed in the "Thank you and goodbye" messages of some airlines, in which, before expressing the hope that its own services will be used again, the airline acknowledges that customers have a choice of carriers. The suggestion is that the airline is confident that its offer is better than those of its competitors and welcomes being put to the test of comparison.

A virtuous trader will not be motivated to seek to be protected by barriers to entry, or to ask potential trading partners to trade for reasons other than price and quality. Nor will a virtuous trader be inclined to make agreements with other traders on the same side of the market to restrict supply or demand, or to partition the market and then not compete. It might be objected that such cartel agreements are mutually beneficial transactions for the firms that are parties to them. But they are not the transactions *in goods and services* that constitute the market, and with respect to which mutual benefit is understood by those economists who see mutual benefit as the *telos* of the market. If obstructing other parties' transactions is nonvirtuous, so too is participation in cartels.

This market virtue seems inescapable, given our approach, but there is no denying that traders often find it hard to live by. For example, Adam Smith famously claimed: "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or some contrivance to raise prices" (1776 [1976], p. 45). Nevertheless, it is obvious from the tone of these and similar remarks—for example about "the wretched spirit of monopoly" (p. 461)—that Smith does not *approve* of this trait. The idea that cartel agreements are unethical—unworthy of a virtuous trader—is a recurring theme in the writings of pro-market economists. Even Friedman (1962, pp. 131–132), who argues that market power is not a serious problem unless it is positively supported by governments, approves the common law doctrine that combinations in restraint of trade are unenforceable in the courts.

This is a convenient place to ask whether being concerned about externalities resulting from one's activities should be included among the market virtues. One way of posing this question is to ask whether the *telos* of the market is mutual benefit *among the parties to market transactions* (considered severally), or mutual benefit *among everyone in a society*. We suggest the former. On this view, the existence of externalities can be a reason for governments to regulate markets, but self-regulation is not part of the internal practice of the market.[5]

Self-Help

Within the practice of a market that is structured by mutual benefit, each individual's wants and aspirations are relevant to others only in so far as they can be satisfied in mutually beneficial transactions. Thus, it is a market virtue to accept without complaint that others will be motivated to satisfy your wants, or to provide you with opportunities for self-realization, only if you offer something that they are willing to accept in return. Smith (1776 [1976], p. 45) appeals to the virtue of self-help or independence when, in relation to how we get our dinners, he writes: "Nobody but a beggar chuses to depend chiefly upon the benevolence of his fellow-citizens." (The phrase "chuses to" is important here. Smith is not denigrating dependence on others by people who have no other means of subsistence.)

A person who upholds the virtue of self-help will avoid asking others to reward her for producing goods that those others do not value. Thus, for example, an artist will not treat the intrinsic value of her work, as judged within the practice of art, as a reason to be paid by people (whether as consumers or as taxpayers) who do not recognize that work as beneficial to them. Nor will she treat the self-realization that she achieves through that work as a reason to be paid. In this respect, the market virtue of self-help conflicts with the positions taken by Anderson (1993) and Sandel

[5] To this extent, we agree with Friedman (1962, pp. 133–36) that "social responsibility" is not a proper role of business. However, Friedman argues that the only responsibility of business is "to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game, which is to say, engages in open and free competition, without deception or fraud." Our idea that market virtue involves intentions for mutual benefit is broader than this claim.

(2009). From the perspective of market virtue, the commodification of a practice is nothing more than its orientation towards mutual benefit. Expecting others to pay for one's preferred form of self-realization is a kind of civil (as distinct from clinical) narcissism. One might add a person who thinks of her interactions with others in terms of self-realization is treating those others as means to her own ends rather than as partners in a cooperative relationship.

Self-help is also opposed to self-sacrifice, and so to the conception of virtue and intrinsic motivation that underlies the idea of "getting more by paying less." A relationship in which one party incurs a loss so that another person can gain is not a mutually beneficial transaction between equals, and so does not express market virtue on either side. The motivational asymmetry of such a relationship—which might be revealed in the giver's expectation of gratitude or status recognition, or in either party's assumption that the recipient's desires or interests take precedence over the giver's—contrasts with the symmetry of a normal market transaction. The "trade not aid" slogan of the fair trade movement is an expression of the market virtue of self-help.

Seeing self-help as a virtue makes it easier to understand how people can find satisfaction in work that they would not choose to do if they were not paid for it. Large parts of most people's working lives are not "fun" or "challenging" in the sense of self-determination theory. Nor are they most naturally understood as the pursuit of artistic, professional, or craft excellence, or as self-sacrificing caring. They are simply activities by which one earns a living by being useful to other people in ways that they are willing to pay for. But that surely does not mean that these activities lack authenticity or virtue.

Non-Rivalry

If opportunities for mutual benefit are to be realized, individuals must perceive the market as a domain in which such opportunities exist. Thus, it must be a market virtue to see others as potential partners in mutually beneficial transactions rather than as rivals in a competition for shares of a fixed stock of wealth or status. A disposition to be grudging or envious of other people's gains is a handicap to the discovery and carrying through of mutually beneficial transactions. The corresponding virtue is that of being able to take pleasure in other people's gains—particularly those that have been created in transactions from which you have gained too.

As viewed in the liberal tradition of economics, the market is not the archetypal locus of positional competition, with success measured by relative wealth. Indeed, positional competition may be more typical of professions that have maintained some insulation from the market and have developed nonmarket institutions for ranking excellence, such as literary, artistic, and scientific honors and prizes. Perhaps one of the reasons why academic writers (including some economists) often find it difficult to understand how markets can be structured by mutual benefit is that competition in the intellectual community is so positional.

From the earliest days of economics, prominent economists have argued against positional understandings of market competition, and have presented nonpositional

attitudes as virtuous. For example, Hume (1760 [1985], pp. 327–28) argues against the "narrow and malignant opinion" that the relationship between commercial economies is that of zero-sum rivalry: "[T]he encrease of riches and commerce in any one nation, instead of hurting, commonly promotes the riches and commerce of all its neighbours."[6] Writing almost a century later, Mill (1848 [1909], pp. 581–82) expresses the same sentiment: "[C]ommerce first taught nations to see with good will the wealth and prosperity of one another. Before, the patriot . . . wished all countries weak, poor, and ill-governed, but his own: now he sees in their wealth and progress a direct source of wealth and progress to his own country."

What about rivalry between firms, and in particular the case in which the successful entry of one firm into an industry squeezes out another? Even in these cases, the *motivation* of the entrant need not be positional. Indeed, even a self-interested entrant would have no reason to want to displace an incumbent firm, except as a means of making profit; and that profit can be earned only through mutually beneficial transactions with customers. A virtuous entrant, one might say, intends that the transactions he offers to make are mutually beneficial *for the parties that will be involved in them*; the entrant does not intend or take satisfaction in the failure of competitors, even if that external effect is a predictable consequence of successful entry.

Stoicism about Reward

In a market structured by mutual benefit, each individual benefits according to the value that other people place on their transactions with that individual. In terms of any defensible concept of what people deserve, this form of economic organization cannot consistently reward people according to their deserts. Desert is a backward-looking concept: what people deserve can depend on how they behaved in the past. But mutual benefit, in the sense that markets can be said to facilitate its achievement, is defined in terms of people's circumstances and beliefs *at the time at which they trade*. Because economic circumstances can change unpredictably, efforts that were made with reasonable expectations of return may turn out not to be rewarded by the market. Conversely, being in a position to gain from mutually beneficial transactions with others at a particular time and place can involve luck as well as foresight. Sandel's (2009) example of being able to benefit from possessing the human and physical capital of a hotelier or builder in the aftermath of a hurricane is just an extreme case of this general feature of market reward. If Sandel's interpretation of the pay of senior corporate executives in the pre-2008 period is that that those executives were benefiting from the good luck of being able to exercise their trade in a bull market, that example illustrates the same point.

[6] That international trade promotes peace by making nations dependent on one another was argued even earlier, by Montesquieu (1748 [1914], Book 20, Section 2). However, Hume is more explicit in arguing that trade gives each country an interest in the prosperity of its trading partners.

To recognize this feature of markets is not to oppose all redistributive policies. Indeed, one might argue that a market economy is politically sustainable only if everyone can expect to benefit in the long run from the wealth that markets create, and that might require some collective commitment to redistribution. But if the market is to function, rewards cannot be perfectly aligned with desert (Sugden 2004, 2012). To some critics, this disconnect between reward and desert comprises a moral failure of the market. Sandel (2009) refers to a passage in which Milton and Rose Friedman (1980, pp. 136–137) argue that this aspect of the unfairness of life is a price we have to pay for the freedom and opportunity that the market gives us. Sandel (pp. 164–165) thinks this a "surprising concession" from advocates of the market. His thought seems to be that material wealth is the currency of market reward, and that individuals' earnings from the market ought therefore to be in due proportion to effort and talent.

Of course it is true that most people value material wealth, and that, in this morally neutral sense, wealth is a currency of reward in the market, as it is in other domains of life. But an adequate account of market virtue cannot maintain that what a person earns from market transactions is a reward *for the exercise of virtue*, in the sense that a literary prize can be seen as a reward for artistic excellence. A person can expect to benefit from market transactions only to the extent that she provides benefits that trading partners value at the time they choose to pay for them. To expect more is to create barriers to the achievement of mutual benefit. Thus, market virtue is associated with *not* expecting to be rewarded according to one's deserts, *not* resenting other people's undeserved rewards, and (if one has been fortunate) recognizing that one's own rewards may *not* have been deserved.

This attitude of fortitude or stoicism towards the distribution of rewards in a market economy is fundamental to Hayek's (1976) account of the moral status of the market and "the mirage of social justice." Hayek accepts that the market often fails to reward desert, but writes: "It is precisely because in the cosmos of the market we all constantly receive benefits which we have not deserved in any moral sense that we are under an obligation also to accept equally undeserved diminutions of our incomes. Our only moral title to what the market gives us we have earned by submitting to those rules which make the formation of the market order possible" (p. 94).

Conclusion

We have presented a view of the market as a domain of human life with a distinctive constellation of virtues. We have argued that this view of the market is compatible with, and to some extent implicit in, a long tradition of liberal economic thought. The virtues we have discovered do not, as some moral critics of the market might have expected, merely normalize egoism and instrumentality: they are genuine virtues that can be upheld with authenticity.

We stress again that virtues are defined relative to practices. The traits that make a person good *as a participant in markets* need not be evaluated positively in all domains of human life. To acknowledge that there are market virtues is not to claim that the market is the only morally relevant domain, nor that the market virtues are the only virtues. We have argued (in agreement with some but not all virtue ethicists) that the virtues of different domains can conflict with one another. Thus, the market virtue of universality can conflict with loyalty to community and tradition. Respect for one's trading partners' tastes can conflict with upholding standards of professional and craft excellence. The virtue of self-help, as viewed by a potential philanthropist, can conflict with benevolence. Stoicism about market reward can conflict with the pursuit of social justice. However, it should not be thought that the market virtues apply *only* within the practice of the market. On our account, the *telos* of the market is mutual benefit. Thus, market virtues will apply in other domains of human life that are understood as cooperation among equals for mutual benefit and that, as Mill (1861 [1976], pp. 29–30) argues, thereby provide the environment in which the "social feelings of mankind" can develop. As Mill and many later theorists of social capital recognize, market relations form one part of the network of cooperative relations of which civil society is made up (for example, Putnam 1993). Thus, the market virtues are also virtues of civil society in general.

We close with an expression of this idea by Antonio Genovesi (1765–67 [2005]), an Italian contemporary of Adam Smith who, like Smith, tried to understand the motivations driving the growth of commercial societies in his time and who made an attempt to build a theory of commercial society based on the idea of mutual assistance (Bruni and Sugden 2000). Significantly, the name that Genovesi tried to give our discipline was not *political* economy but *civil* economy. We quote the final words of his *Lectures on Commerce, or on Civil Economy* (Genovesi, 1765–67 [2005], our translation), delivered at the University of Naples, where he was the world's first professor of economics. Having taught his students how a commercial society works, he concludes: "Here is the idea of the present work. If we fix our eyes at such beautiful and useful truths, we will study [civil economy] . . . to go along with the law of the moderator of the world, which commands us to do our best to be useful to one another."

■ *We are grateful for comments from participants at various conferences and workshops at which earlier versions of this paper were presented, and from the editorial team at the* Journal of Economic Perspectives. *Sugden's work was supported by the Economic and Social Research Council through the Network for Integrated Behavioural Science (grant reference ES/K002201/1).*

References

Akerlof, George. 1982. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics* 97(4): 543–69.

Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Aristotle. c350 BCE [1980]. *The Nicomachean Ethics of Aristotle*, translated by William David Ross. Oxford University Press.

Arrow, Kenneth J. 1972. "Gifts and Exchanges." *Philosophy and Public Affairs* 1(4): 343–62.

Arrow, Kenneth J. 1984. *Collected Papers of Kenneth J. Arrow,* Vol. 1: *Social Choice and Justice*. Oxford: Basil Blackwell.

Besley, Timothy. 2013. "What's the Good of the Market? An Essay on Michael Sandel's *What Money Can't Buy*." *Journal of Economic Literature* 51(2): 478–95.

Brennan, Geoffrey. 1996. "Selection and the Currency of Reward." In *The Theory of Institutional Design*, edited by Robert E. Goodin, 256–75. Cambridge University Press.

Bruni, Luigino, and Robert Sugden. 2000. "Moral Canals: Trust and Social Capital in the Work of Hume, Smith and Genovesi." *Economics and Philosophy* 16(1): 21–45.

Bruni, Luigino, and Robert Sugden. 2008. "Fraternity: Why the Market Need Not Be a Morally Free Zone." *Economics and Philosophy* 24(1): 35–64.

Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Coyle, Diane. 2012. "Should You Pay to Jump the Queue—Or for a New Kidney? It's Hard to Define Where Cash Has No Place." A Review of *What Money Can't Buy: The Moral Limits of Markets* by Michael Sandel. *The Independent*, May 5. http://www.independent.co.uk/arts-entertainment/books/reviews/what-money-cant-buy-the-moral-limits-of-markets-by-michael-sandel-7711785.html.

Deci, Edward L. 1971. "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology* 18(1): 105–115.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.

Edgeworth, Francis Ysidro. 1881. *Mathematical Psychics*. London: Kegan Paul.

Folbre, Nancy, and Julie A. Nelson. 2000. "For Love or Money—Or Both?" *Journal of Economic Perspectives* 14(4): 123–40.

Frey, Bruno S. 1994. "How Intrinsic Motivation Is Crowded In and Out." *Rationality and Society* 6(3): 334–52.

Frey, Bruno S. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar.

Friedman, Milton. 1962. *Capitalism and Freedom*. University of Chicago Press.

Friedman, Milton, and Rose Friedman. 1980. *Free to Choose*. New York: Houghton Mifflin Harcourt.

Genovesi, Antonio. 1765–67 [2005]. *Lezioni di commercio o sia di economia civile*. Instituto Italiano per gli Studi Filofící, Napoli.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25(4): 191–210.

Gneezy, Uri, and Aldo Rustichini. 2000a. "A Fine Is a Price." *Journal of Legal Studies* 29(1): 1–17.

Gneezy, Uri, and Aldo Rustichini. 2000b. "Pay Enough or Don't Pay at All." *Quarterly Journal of Economics* 115(3): 791–810.

Hayek, Friedrich A. 1948. *Individualism and Economic Order*. University of Chicago Press.

Hayek, Friedrich A. 1976. *Law, Legislation and Liberty*, Vol 2: *The Mirage of Social Justice*. University of Chicago Press.

Heyes, Anthony. 2005. "The Economics of Vocation, or 'Why Is a Badly-Paid Nurse a Good Nurse?'" *Journal of Health Economics* 24(3): 561–69.

Hirschman, Albert O. 1997. *The Passions and the Interests: Political Arguments for Capitalism before Its Triumph*. Twentieth Anniversary Edition. Princeton, NJ: Princeton University Press. (First edition 1977.)

Hume, David. 1760 [1985]. "On the Jealousy of Trade." In *Essays Moral, Political and Literary*, by David Hume, pp. 327–31. Indianapolis: Liberty Fund.

Jevons, William Stanley. 1871 [1970]. *The Theory of Political Economy*. London: Penguin.

Kahan, Alan S. 2010. *Mind vs. Money: The War Between Intellectuals and Capitalism*. Piscataway, NJ: Transaction Publishers.

Katz, Eliakim, and Femida Handy. 1998. "The Wage Differential between Non-Profit Institutions and Corporations: Getting More by Paying Less?" *Journal of Comparative Economics* 26(2): 246–61.

Lepper, Mark R., and David Greene, eds. 1978. *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*. Hillsdale, NY: Erlbaum.

MacIntyre, Alisdair. 1984. *After Virtue: A Study in Moral Theory*, Second edition. Notre Dame, IN: University of Notre Dame Press. (First edition 1981).

Marshall, Alfred. 1920. *Principles of Economics*, Eighth edition. London: Macmillan.

McCloskey, Deidre N. 2006. *The Bourgeois Virtues: Ethics for an Age of Commerce*. University of Chicago Press.

Mill, John Stuart. 1848 [1909]. *Principles of Political Economy*. London: Longmans.

Mill, John Stuart. 1861 [1976]. *Utilitarianism*. London: Dent.

Montesquieu, Charles-Louis de Secondat, Baron de. 1748 [1914]. *The Spirit of the Laws*. Translated by Thomas Nugent. London: Bell and Sons.

Nelson, Julie. 2005. "Interpersonal Relations and Economics: Comments from a Feminist Perspective." In *Economics and Social Interaction: Accounting for Interpersonal Relations*, edited by Benedetto Gui and Robert Sugden, 250–61. Cambridge University Press.

Putnam, Robert D. 1993. *Making Democracy Work*. Princeton University Press.

Ratner, Gerald. 2007. *Gerald Ratner: The Rise and Fall . . . and Rise Again*. North Mankato, MN: Capstone.

Ricardo, David. 1817. *On The Principles of Political Economy and Taxation*. London: John Murray.

Ryan, Richard M., and Edward L. Deci. 2000. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology* 25(1): 54–67.

Sandel, Michael J. 2009. *Justice: What's the Right Thing to Do?* London: Penguin.

Sandel, Michael J. 2012. *What Money Can't Buy: The Moral Limits of Markets*. New York: Farrar, Straus and Giroux.

Smith, Adam. 1759 [1976]. *The Theory of Moral Sentiments*. Oxford: Clarendon Press.

Smith, Adam. 1763 [1978]. *Lectures on Jurisprudence*. Oxford University Press.

Smith, Adam. 1776 [1976]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press.

Smith, Greg. 2012. "Why I Am Leaving Goldman Sachs." *New York Times,* March 14.

Sugden, Robert. 2004. "Living with Unfairness: The Limits of Equality of Opportunity in a Market Economy." *Social Choice and Welfare* 22(1): 211–236.

Sugden, Robert. 2012. "The Market as a Cooperative Endeavour." *Public Choice* 152(3): 365–370.

Titmuss, Richard. 1970. *The Gift Relationship: From Human Blood to Social Policy*. London: Allen and Unwin.

van Staveren, Irene. 2009. "Comunitarianism and the Market: A Paradox." *Review of Social Economy* 67(1): 25–47.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 165–186

Gifts of Mars: Warfare and Europe's Early Rise to Riches

Nico Voigtländer and Hans-Joachim Voth

Today, per capita income differences around the globe are large, varying by as much as a factor of 35 across countries (Hall and Jones 1999). These differentials mostly reflect the "Great Divergence" (a term coined by Huntington 1996)—the fact that Western Europe and former European colonies grew rapidly after 1800, while other countries grew much later or stagnated. What is less well-known is that a "First Divergence" preceded the Great Divergence: Western Europe surged ahead of the rest of the world long before technological growth became rapid. Europe in 1500 already had incomes twice as high on a per capita basis as Africa, and one-third greater than most of Asia (Maddison 2007). In this essay, we explain how Europe's tumultuous politics and deadly penchant for warfare translated into a sustained advantage in per capita incomes.

Much of the European advantage in per capita incomes emerged after the Black Death of 1350, which killed between one-third and one-half of the European population. In the three centuries after 1400, European per capita incomes grew rapidly, while Africa and Asia stagnated (Maddison 2001). By 1700, Western Europeans produced 2.5 times more than Africans on a per capita basis, and 70–85 percent more than Indians, Chinese, and Japanese.

■ *Nico Voigtländer is Assistant Professor of Economics, Anderson School of Management, University of California at Los Angeles. He is also a Research Fellow, National Bureau of Economic Research, Cambridge, Massachusetts. Hans-Joachim Voth is ICREA Research Professor of Economics, DEE, at Universitat Pompeu Fabra, Barcelona, Spain. He is also a Research Fellow, Centre for Economic Policy Research, London, United Kingdom. Their email addresses are nico.v@anderson.ucla.edu and jvoth@crei.cat.*

http://dx.doi.org/10.1257/jep.27.4.165 doi=10.1257/jep.27.4.165

Table 1
The "First Divergence"—Europe versus China

Urbanization rate (percentage of population living in cities with more than 10,000 inhabitants)			*GDP per capita (in 1990 international dollars)*		
Year	*China*	*Europe*	***Year***	*China*	*Europe*
762	3%		**1**	$450	$550
1000		0%	**960**	$450	$422
1120	3.1%		**1300**	$600	$576
1500	3.8%	5.6%			
1650	4%	8.3%	**1700**	$600	$924
1820	3.8%	10%	**1820**	$600	$1,090

Source: Maddison (2007).

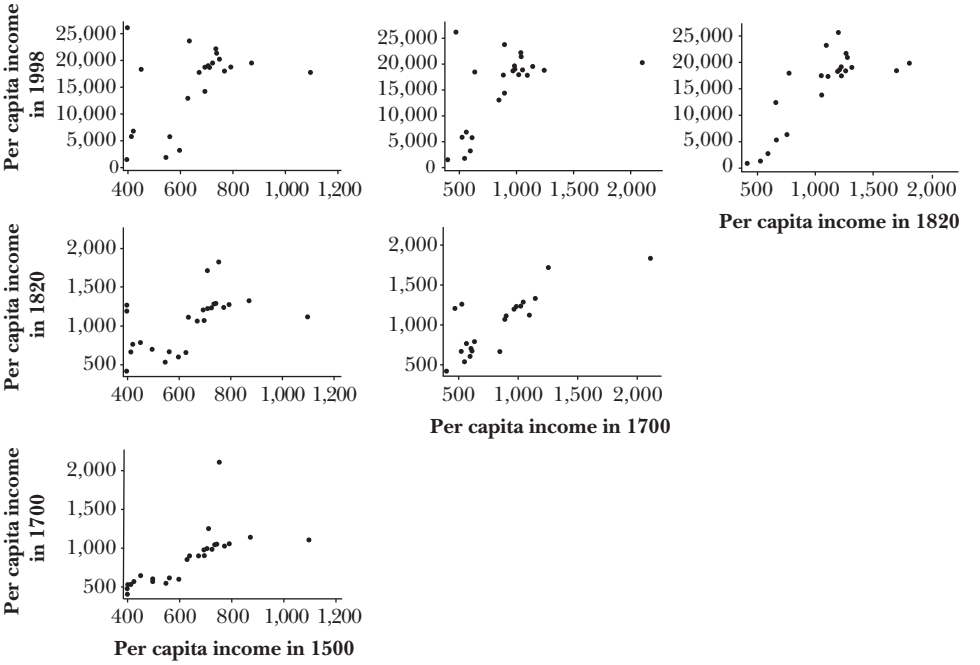
Within Europe, there was also divergence: UK incomes were 75 percent higher in 1700 than they had been in 1500; Dutch incomes increased by 180 percent. At the opposite end of the spectrum were the laggards: Italy probably showed essentially no increases in productive capacity over these two centuries; Spain grew by only 28 percent. Urbanization rates tell a similar story. Where data on per capita income is poor, urbanization rates make a good substitute (Wrigley 1985; Nunn and Qian 2011). This is because urbanization will reflect both the productivity of the urban sector (creating goods that can be traded for food) and of agriculture (which needs to generate a surplus above subsistence to feed cities).

Table 1 shows comparative figures for Europe and China, for both urbanization rates and GDP per capita. (The urbanization rate is defined as the percentage of the population living in cities with more than 10,000 inhabitants.) Europe may have been slightly ahead of China in terms of per capita incomes in Roman times; by the High Middle Ages, it had declined both absolutely and in relative terms, before showing rapid increases. The urbanization rate in China was already around 3 percent in the eighth century; Europe, in contrast, probably lagged behind substantially.[1] By 1500, European urbanization rates were already higher than in China; by 1650, they were twice those in the Far East.[2]

[1] Maddison (2007) estimates that the share of Europeans living in cities of more than 10,000 inhabitants was zero. Bosker, Buringh, and van Zanden (2008) present alternative figures, showing higher urban shares in the Iberian peninsula, which was under Arab rule at the time.

[2] The evidence in favor of Chinese underperformance has been questioned. Pomeranz (2001) points out that comparing the most advanced countries of Europe such as England and the Netherlands with all of China is unfair. The Yangtze area, China's leading agricultural producer, did much better than the rest of the country. However, a decade of detailed research has now firmly established that early modern European incomes were indeed much higher than Chinese ones. Broadberry and Gupta (2012) estimate that Chinese and Indian wages already lagged European wages as early as 1550; by 1800, the gap was huge. Allen, Bassino, Ma, Moll-Murata, and van Zanden (2011) similarly show that real urban wages in China were much lower than in Europe. Allen (2009a) shows that even in the Yangtze area, per capita incomes were on a downward path during the early modern period.

Figure 1
Scatterplot of per Capita Incomes in 1500, 1700, 1820, and 1998



Note: Data are from Maddison (2001), and the countries with data availability for 1500–1820 include the European countries as well as Brazil, Mexico, China, India, Indonesia, Japan, Philippines, Iran, Iraq, Turkey, Egypt, and Morocco.

The First Divergence matters not only for incomes at that time: The countries that surged ahead also conquered vast parts of the globe in the 19th century, and remain amongst the first rank of economic nations today. Countries that failed to grow in the early modern period remained poor for centuries; only some caught up more recently.

For example, the same countries that surged ahead after 1500 were also the first to undergo an Industrial Revolution (Comin, Easterly, and Gong 2010). Figure 1 illustrates the persistence of per capita income over the long term, plotting levels in 1500, 1700, 1820, and 1998 against each other. The correlation coefficient ranges from 0.46 to 0.8, and is highly significant in every pairwise comparison. A naïve regression of income levels in 1998 on per capita income in 1500 can explain more than 20 percent of total variance; incomes in 1820 predict 64 percent of cross-sectional differences. If the relationship is already strong when looking at countries, it is even stronger when adjusted for ancestral population movements (Putterman and Weil 2010). One of the best predictors of an individual's income today is the level of riches attained by that person's ancestors hundreds of years ago.

In this paper, we argue that Europe's rise to riches during the First Divergence was driven by the nature of its politics after 1350—it was a highly fragmented continent characterized by constant warfare and major religious strife. Our explanation emphasizes two crucial and inescapable consequences of political rivalry: war and death. No other continent in recorded history fought so frequently, for such long periods, killing such a high proportion of its population. When it comes to destroying human life, the atomic bomb and machine guns may be highly efficient, but nothing rivaled the impact of early modern Europe's armies spreading hunger and disease.

In a Malthusian world, the amount of land per person was the prime determinant of per capita output. Wars were so common, and their impact was so severe, that they raised average death rates in early modern Europe significantly.[3] In turn, this spelled higher land-labor ratios in agricultural production and thus higher per capita income (Voigtländer and Voth 2009, 2013). War therefore helped Europe's precocious rise to riches because the survivors had more land per head available for cultivation. We argue that the feedback loop from higher incomes to more war and higher land-labor ratios was set in motion by the Black Death in the middle of the 14th century. As surplus incomes over and above subsistence increased, tax revenues surged. These in turn financed near-constant wars on an unprecedented scale. Wars raised mortality not primarily because of fighting itself; instead, armies crossing the continent spread deadly diseases such as the plague, typhus, or small pox. The massive, continued destruction of human life that followed led to reduced population pressure. In our view, it was a prime determinant of Europe's unusually high per capita incomes before the Industrial Revolution.

A rapidly growing literature on persistence in economic performance has sought explanations for the long arm of history—the puzzling extent to which past economic performance continues to predict present economic outcomes. Focusing on the British case, in Voigtländer and Voth (2006), we model productivity advance as an externality from capital use, and show how higher premodern incomes can improve the chances of industrializing. Comin, Easterly, and Gong (2010) argue that technological leadership is bequeathed from generation to generation, while Spolaore and Wacziarg (2009) conclude that genetic distance to the technological leader—a proxy for how long ago two populations shared a common ancestor—is a key predictor of per capita incomes.

We begin by describing the economic logic of the Malthusian world, explain why existing interpretations struggle to make sense of Europe's early and sustained lead in per capita income, and introduce the evidence for our own interpretation in more detail. We also compare our results for Europe with the Chinese case before discussing why alternative interpretations of the First Divergence are ultimately unconvincing.

[3] In England, for example, average life expectancy fell from 40 years in 1580 to around 32 years in 1700.

The Puzzle: Sustained Riches in a Malthusian World

Malthus today is a byword for economic stagnation. His "iron law of wages" implies that technical advances cannot lead to greater riches: Whenever additional income became available, it would translate into population growth (Malthus 1798). In the Malthusian worldview, fertility reacted faster to positive income shocks than technology could grow, and wages quickly returned to their previous levels. In the polemical novel *A Modern Utopia*, H. G. Wells (1905) summarized this view by writing that earlier generations "spent the great gifts of science as rapidly as it got them in a mere insensate multiplication of the common life."

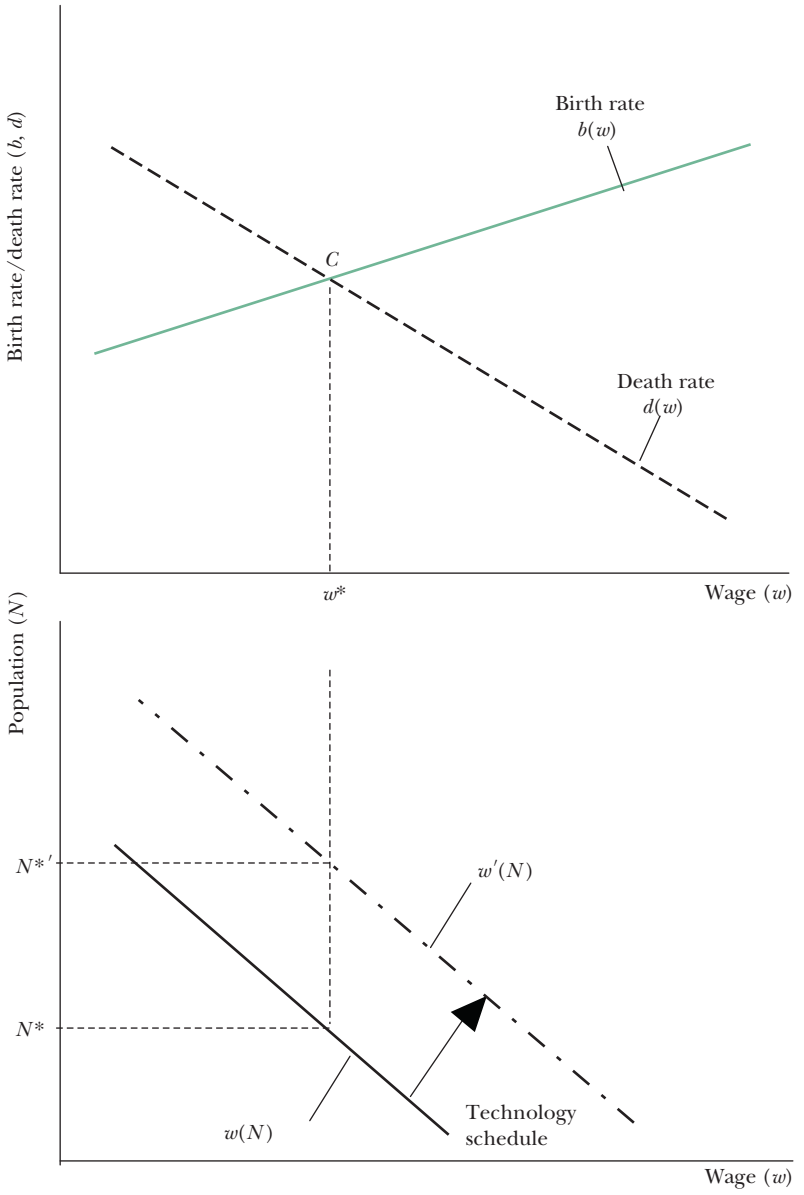
The underlying reason is that—with land as a key factor of production in fixed supply—marginal returns to labor declined quickly. Population growth spelled lower land-labor ratios, and fewer units of land per head meant lower productivity. After the Industrial Revolution, the world escaped the Malthusian trap by finding ways to produce output that relied less and less on nonreproducible factors of production (Hansen and Prescott 2002). Land is a negligible part of the capital stock in most modern economies; as a result, population growth only has second-order effects for output per head.

A well-known implication of the Malthusian model is that death and birth rates alone pin down the long-run equilibrium wage. Fewer births, or a higher death rate, both translate into higher incomes, because of reduced population pressure. Thus, high mortality could be good news for living standards of the survivors. Indeed, European incomes peaked after the Black Death in 1350, reaching levels not attained again until the 19th century. In other words, it took until the reign of Queen Victoria (1819–1901) for per capita incomes to return to the levels last seen under Richard III (1452–1485).

Figure 2 illustrates the standard Malthusian model. Death rates fall as wages increase; birth rates rise. Where the two intersect at point C, population growth is zero. This defines the equilibrium wage w^*. As the lower panel shows, there is a trade-off between population size and average wages, reflecting declining marginal returns. The system is self-equilibrating: Temporary reductions in population size (due to a one-off mortality shock, such as the Black Death) will lower population and raise wages. Population growth will set in, and it will continue until the economy returns to the same equilibrium levels of population N^* and w^*. An innovation in technology will raise wages above w_C temporarily (to see this, move horizontally from N^* to the new technology schedule $w'(N)$ in Figure 2). As a result, population grows. This, in turn will reduce land-labor ratios, leading to lower wages. In the long run, the economy will return to w_C, but at a higher population level $N^{*\prime}$. Therefore, one-off technological improvements will not raise wages in a sustainable fashion.

The Black Death boosted effective resources per unit of labor to an extent that could not be reversed within a few generations. In the long run, incomes should have returned to the pre-plague level. After 1350, wages indeed spiked in Europe, and then trended downwards as population recovered. But in this particular case,

Figure 2
Unique Equilibrium in the Malthusian Model



Notes: In the Malthusian model, birth rates increase with wages, while death rates decline (upper panel). Wages, in turn, depend negatively on population (lower panel)—due to decreasing returns to labor in an economy with fixed land supply. The intersection of birth and death rates yields zero population growth, and thus a stable population N^*. If mortality shocks move wages beyond w^*, population grows. Rising population exerts downward pressure on wages, and the economy returns to point C. If technology improves, the $w(N)$ schedule shifts out, so that a higher population can be sustained at any given wage. However, technological change does not affect the steady state wage w^*.

incomes never returned to pre-plague levels. To understand why, we need to think through the logic of Malthusian forces more generally.

The world prior to the Industrial Revolution was largely governed by Malthusian forces (Ashraf and Galor 2011; Clark 2007): that is, higher wages caused population growth to accelerate; higher population pressure reduced incomes. Some authors have doubted that Malthusian forces were strong in Europe before 1800. Since population influences wages, and wages influence population size (via marriage rates and mortality), it is not easy to identify directions of causality. Three approaches have been pursued. Nicolini (2007) and Crafts and Mills (2009) use vector autoregressions—which begin by not making any assumptions about directions of causality—to argue that Malthusian forces were relatively weak. Anderson and Lee (2002) offer similar findings. Kelly (2005) instead uses weather as an instrument for wages and finds significant evidence for rising fertility and declining mortality in response to positive shocks to agricultural productivity in medieval and early modern England. This suggests that Malthusian forces were strong, with population growth responding quickly to increasing income. The results by Kelly seem more in line with the aggregate evidence: in England before 1750, there was a sharp trade-off between population size and per capita incomes (Wrigley and Schofield 1989). This pattern also holds true outside Europe: Chen and Kung (2012) use weather variability to show that Malthusian forces were important in Qing China.

In a Malthusian world, neither technological advances nor improvements in institutions can lead to sustained increases in per capita output. A high rate of technological change in the premodern era was 0.25–0.5 percent annually—which implies a 28–64 percent increase over a century—while the average was about 0.1 percent annually (Galor 2005). On the other hand, human populations can easily expand at an annual rate of 3 percent or more, which implies an increase of more than 1,800 percent over a century. In other words, in a Malthusian economy the race between technology and population size is the turtle against the hare—technological change can almost never rise fast enough to overcome the deleterious effects of population growth. The same logic applies for institutional improvements. They, too, can improve the mapping from resources to output, just like technological advances—but it is highly unlikely that institutional improvements outpaced the ability of human populations to grow.

High and Stable Incomes after the Black Death: The Effects of War

At its worst, early modern war from about 1400 to 1700 was more deadly than World War II in the most affected areas. During 1941–45, for example, the Soviet Union lost an estimated 24 million citizens, both combatants and civilians, out of a population of 168 million. This amounts to a loss of nearly 15 percent. German losses were somewhat smaller in proportion to the size of the population, while Polish ones were greater—17 percent of Poland's population died during wartime after 1939. By comparison, the United States and the United Kingdom

during World War II registered losses of less than 1 percent. In contrast, the two greatest periods of conflict in the early modern period—the Religious Wars in late 16th-century France and the Thirty Years War in Germany—claimed approximately 20 and 33 percent of the population, respectively.[4] While these estimates have large margins of error, war in the age of the musket could clearly be more devastating than in the age of tanks and aerial bombardment. How do we explain this puzzling fact?

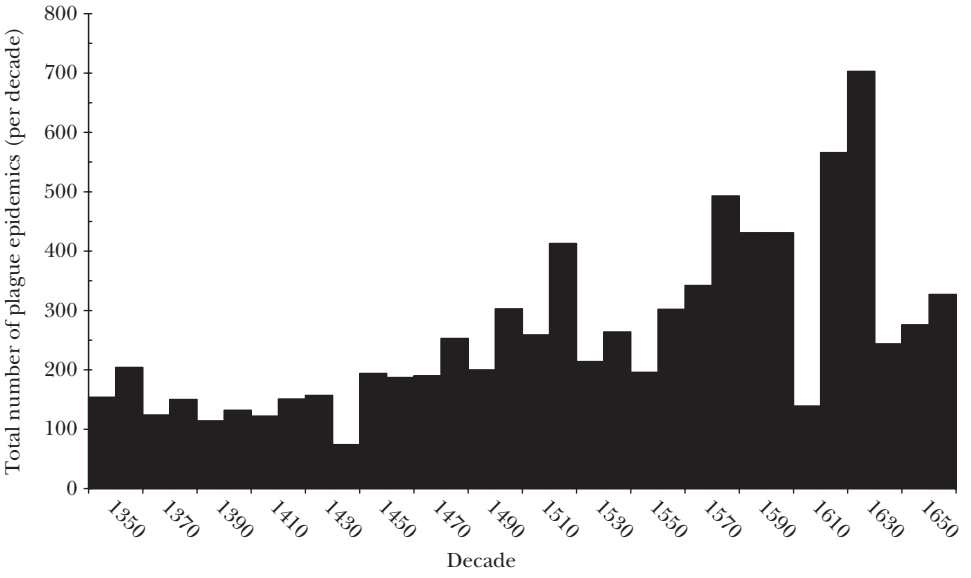
The deadliness of war principally depends on two factors—the lethal power of weaponry, and the frequency with which noncombatants and soldiers succumb to hunger and disease. The killing power of modern arms is many times greater than it was in the past (Ferguson 2002, 2006), but death from hunger and disease has become less frequent over time. Before the 19th century, the disease channel was the most important driver of war-related mortality (Landers 2005). There are many examples: When Europeans arrived in the Americas, even minor ailments like the flu killed natives in large numbers. Major diseases like smallpox wiped out entire populations (Diamond 1997). It has been estimated that European diseases caused a collapse in Meso-American population size by 75 percent or more.

Something similar, if milder, occurred when an army marched through isolated villages in the European countryside. They brought the local population into contact with new diseases, causing major spikes in mortality. Trade could also spread diseases: the last plague outbreak in Western Europe, in Marseilles in 1720, was caused by a ship from the Levant. But armies were more potent vectors of disease. Typhus probably reached Europe via Spanish soldiers who contracted it on Cyprus; syphilis might have been brought back from the Americas; and plague famously spread throughout the Old World after a Mongol army infected the Genoese defenders of a trading outpost on the Crimea in 1347 (McNeill 1976). The more isolated populations were, the greater the mortality impact of a new disease. Exposure to new diseases was deadly for soldiers as well: during colonial wars in Africa, annual death rates could reach one-fifth or more.[5] As late as the early 19th century, Russian troops occupying Swedish islands caused a major increase in death rates—without any fighting. Prussian troops contracted smallpox when campaigning in France during the Franco-Prussian War, leading to an epidemic at home once they returned (Landers 2003).

[4] Total military and civilian deaths during World War II come from Clodfelter (2002) for the United States, United Kingdom, Japan, and Poland; from Hubert (1998) for Germany; and from Ellman and Maksudov (1994) for the USSR. Population estimates come from US Census Bureau (2000), Mitchell (1988), Statistics Bureau of Japan (2011), Hubert (1998), Ellman and Maksudov (1994), and Piotrowski (1997). For French Religious Wars, we use the death toll in Knecht (1996) and the population estimate in Dupâquier (1988), and we use Clodfelter (2002) for the German Thirty Years War.

[5] During the two-month "Logo expedition" to what is now the country of Mali in 1874, the implied annual death rate was 2,940/1,000, which means that the average soldier had a life expectancy of around four months (Curtin 1998).

Figure 3
Plague Outbreaks in Europe, 1350–1650



Source: Biraben (1975).

War did not just create temporary spikes in death rates; it raised average death rates by up to one-third because it was so common.[6] One way to understand the effectiveness of war as a vector for disease is to look at the pattern of plague outbreaks in Europe. Following the catastrophic outbreak of the plague in 1348, a wave of epidemics followed. The detailed historical records demonstrate that many of them were spread by marauding armies. Figure 3 shows the number of epidemics per year for the period 1350–1650. They increased gradually, from 150 per decade in the second half of the 14th century to more than 400 on average during the first half of the 17th century.

The disease channel was a particularly potent killer in Europe because of geographical fragmentation. The continent is divided by large mountain ranges such as the Alps and the Pyrenees, so the movement of armies brought populations into contact with new germs. Political fragmentation also mattered: it went hand-in-hand with frequent warfare. Since the fall of Rome, Europe has never been dominated

[6] We explain the details of this calculation in Voigtländer and Voth (2013), where we marry micro-evidence on changes in death rates in regions affected by war (traced from data on which areas saw fighting or were traversed by armies) with estimates of the share of population exposed and the likely mortality impact.

Table 2
Frequency of War

Century	*Number of wars*	*Average duration (years)*	*Percentage of years under warfare*
16th	34	1.6	95%
17th	29	1.7	94%
18th	17	1.0	78%
19th	20	0.4	40%
20th	15	0.4	53%

Source: Tilly (1990).
Notes: A year is considered "under warfare" if there is at least one war involving the great powers taking place during any part of that year. The great powers are England, Spain, France, Austria, Russia, and the Ottoman Empire.

by a single power—the bids for supremacy by the Habsburgs, by France, and by Germany all conspicuously failed. Religious strife and dynastic conflict provided a large number of potential flashpoints; it took very little for war to erupt. The European states in the early modern era fought each other like no other continent in recorded history has done before or since. Mortality was so high partly because wars lasted so long: the Religious Wars in 16th-century France lasted for 36 years; the Thirty Years War, 30 years; the War of Spanish Succession, 13 years; the war of the Austrian Succession, eight years; and the Napoleonic Wars, 23 years. Military technology was partly to blame—the rise of early modern fortifications made long drawn-out sieges a necessity. In comparison, twentieth-century wars were relatively brief affairs, with World War I lasting four years, World War II for six years, and the Korean War for three years.

During the period 1500–1700, on average, almost every year saw a war between great powers under way. Warfare was not only frequent after 1500; it was a near-permanent feature of the political landscape. Tilly (1990) calculates that for every 100 years in the 16th and 17th centuries, there was a great power war under way in 95 of them; the rate for the 18th century is only marginally lower. In comparative terms, the 19th century saw much lower frequency of conflict, with only 40 out of 100 years affected, as shown in Table 2. Even the 20th century—termed the "Age of Extremes" by Hobsbawm (1994)—only saw a major armed conflict in a little more than half of all years.

Of course, wars are not equal in their destructiveness; the Thirty Years War lasted for three decades and killed millions; the War of Jenkin's Ear (1739–41) lasted for two years and only caused minor casualties—although it did eventually merge into the War of the Austrian Succession. One indicator of intensity is battle frequency. Europe saw fewer than 100 battles per century from the 9th through the 13th century. The number of battles then jumped to 138 in the 14th century and rose steadily to 521 major battlefield engagements in the 18th century. If one looks at the percentage of the population affected, the numbers tell a similar story of extremely high levels of

conflict: at the height of early modern warfare, during the Thirty Years War, close to half of the European population was affected by military conflict in a given year (for a derivation of this estimate, see Voigtländer and Voth 2013).

Given these rising levels of war and disease, why were Europeans so much more productive by 1700 than they had been in the Middle Ages—and so much more productive than the rest of the world? We argue that the particular type of warfare that characterized Europe after the Middle Ages is an important part of the answer. Before 1800, battlefield deaths and direct victims of armed force were few; civilian and military deaths from disease were plentiful. The imbalance between violent killing, and death from disease, also has important implications for the economic impact of war. War in early modern Europe acted more like a neutron bomb: because of disease, it destroyed human life quickly while not wreaking havoc on infrastructure on a scale comparable to modern wars.[7]

Recovery could be quick in places where it was only wooden houses that needed to be rebuilt. Since European agriculture did not rely on elaborate irrigation systems (as did the Middle East, for example), the direct effects of war were limited to destroyed farm buildings, stores, and livestock. All of these typically could recover or be rebuilt in short order. As Malthus (1798) observed: "The fertile province of Flanders, which has been so often the seat of the most destructive wars, after a respite of a few years, has appeared always as fruitful and as populous as ever. Even the Palatinate lifted up its head again after the execrable ravages of Louis the Fourteenth." Similarly, after the Turkish siege of Vienna in 1683, the speed of recovery was astonishing. As one observer put it: "the suburbs . . . as well as the neighboring countryside . . . have been completely rebuilt in a short space of time" (Tallett 1992). Land left fallow increased in fertility. Livestock had high rates of reproduction, so herds could be rebuilt quickly.

War as practiced in this time therefore combined two characteristics that mattered for economic outcomes: it was highly destructive of human life, and it was largely ineffective in destroying infrastructure and capital stock. While the amount of useful land and the size of the capital stock fell only a little as a direct result of war, military conflict before 1800 was massively destructive of human life. In other words, war was highly effective in increasing the ratio of land and capital relative to the size of the population. In a Malthusian world, frequent war could act as a powerful force raising per capita incomes for the survivors.

[7] A key exception is the destruction of cities. When early modern cities were taken after a siege, they sustained damage that is comparable or worse than that of aerial bombardment during World War II. For example, when Magdeburg was taken by Imperial forces in 1631, the entire city was burned to the ground, more than 90 percent of homes were destroyed, and an estimated 20,000 (out of 35,000) inhabitants lost their lives. The bombing of Dresden caused 20,000–25,000 casualties, out of a population of 650,000 (Cunningham 2000; Neutzner 2010). While the destruction of cities was not minor, there are few countries where the urban share of total population exceeded 10 percent before 1700.

Table 3
Tax Revenues in Europe

Year	*Total tax revenue (tons of silver)*	*Average tax per capita (daily urban wage equivalents)*
1509	214	3.7
1559	456	3.6
1609	1,116	4.9
1659	2,215	5.7
1709	2,667	8.1
1759	3,808	9.9
1789	6,846	12.2

Source and Notes: Data are from Karaman and Pamuk (2010), who use country-level historical compilations of revenue statistics. The main database is the European State Finance Database (ESFD), available at http://www.esfdb.org. They use silver as the measure of fiscal revenue, because all national currencies were convertible into it. The original source for the urban wage series is Allen (2001).

Empirical Evidence for Two Key Hypotheses

Our argument for a "Malthusian circle" involves two key steps: 1) rising incomes after the Black Death led to higher government revenues, with most of these spent on war; and 2) warfare had a silver lining—by shifting death schedules upwards, it spelled higher per capita incomes for survivors. Thus, countries that fought more wars should have seen greater increases in their per capita incomes over the early modern period. Here, we consider evidence bearing on these two hypotheses.

Incomes and Taxes

Government revenues in Europe exploded after 1500. In Table 3, we show total tax revenue and the average tax burden per capita as a multiple of urban daily wages. In 1509, the main European powers had annual silver revenues equivalent to 214 tons; by 1789, this had risen 32-fold to 6,845 tons. Only a small part of the overall increase reflected a decline in the value of silver (inflation after 1550 ran at less than 2 percent per year); most of it came from much higher taxes on (rising) income. The data are ultimately derived from detailed, country-level data on fiscal revenues over time, extracted from the administrative archival records preserved for each state. All tax revenues are expressed in terms of silver and have been compiled in a comparable format by Karaman and Pamuk (2010). Countries included are England, France, Venice, Prussia, Poland, the Dutch Republic, Spain, Austria, and the Ottoman Empire. Data on urban wages are from Allen, Bassino, Ma, Moll-Murata, and van Zanden (2011). The rise in tax revenues is also dramatic when expressed relative to income: in 1509, the average European taxpayer had to

render unto his prince the equivalent of less than four days' pay; by 1789, on the eve of the French Revolution, this had tripled to more than 12 days' wages.

Taxes after 1500 rose much faster than population size. One factor that supported growing tax revenue was growth in per capita incomes—the very fact at the root of the First Divergence. O'Rourke and Williamson (2002) estimate that European *surplus* incomes—that is, the amount above subsistence—on average grew by 0.4–0.5 percent per year from a very low base, much faster than incomes themselves. As incomes grew above subsistence levels, they could increasingly be taxed by the belligerent princes of early modern Europe. Out of every unit increase in "surplus" (above-subsistence) incomes, European states successfully appropriated about one-third (Voigtländer and Voth 2013).

The vast majority of early modern tax revenues were spent on war. European states at this time routinely spent 70–80 percent of their income on armies and navies. In wartime, spending exceeded revenue by a large margin; repaying the past debts thus accounted for a good share of the remaining expenditure.

At this time, financial strength mattered a great deal for military success. Warfare during the early middle ages had been a relatively cheap affair; armies were small, and usually consisted mainly of vassals who were obliged to follow their prince into battle (Landers 2003). In the early modern period, a "military revolution" transformed the face of battle. Armies used firearms, which required extensive training; standing armies became the norm. Huge new fortifications were necessary to protect cities because medieval city walls had been rendered ineffective by the invention of the cannon. These costs quickly bankrupted fiscally weaker states, and financial prowess became a prime determinant of military success. As a Spanish 16th-century soldier and diplomat, Don Bernardino de Mendoza, eloquently put it (as quoted in Parker 1977): "[V]ictory will go to whoever possesses the last escudo."

War and Rising Riches

Did the continuous and near-universal warfare on the European continent lead to higher incomes? We combine data on the incidence of early modern warfare at the country level with two indicators for economic development: urbanization and per capita GDP. The size of urban centers is from Bairoch, Batou, and Chèvre (1988), which we combine with population estimates from McEvedy and Jones (1978) to obtain the percentage living in urban areas. As a result, we have country-level urbanization rates covering the early modern period between 1300 and 1700. As a consistency check, we also use urbanization data from DeVries (1984), which are available from 1500 onwards. In addition, we use per capita income data from Maddison (2001), which is available in 100-year intervals from 1500 onwards. To measure the extent of warfare, we employ data on the years of warfare from Kohn (1999).

We analyze a cross-section of states in early modern Europe. To measure economic progress we use the change in urbanization between 1300 and 1700 and the change in per capita income between 1500 and 1700. Our explanatory variable

is war frequency over the time interval corresponding to each of the two outcome variables. Figure 4, panel A, shows that a higher war frequency between 1300 and 1700 is associated with a larger increase in urbanization. Countries with above-average frequencies of armed conflict, such as the Netherlands, France, and England, gained urban population at a quick rate; those fighting fewer wars, such as Ireland, Switzerland, and Norway, saw only limited progress. The correlation coefficient is 0.40. The same pattern is visible for per capita income (panel B), with a correlation coefficient of 0.28. Here, we use data starting in 1500, as calculated by Maddison (2001).

We look at the relationship between warfare and income growth in a variety of ways. First, we split the countries into two groups, those with below-average and above-average war frequency. We find that both measures of urbanization as well as per capita GDP grew significantly faster in countries with an above-average number of wars. For example, urbanization rates grew by 7.4 percentage points between 1300 and 1700 in countries with above-mean war frequency, versus 2.8 in the remainder. Per capita income grew almost twice as fast over 1500–1700 in countries with above-average warfare. These differences are statistically significant at the 5 percent level.

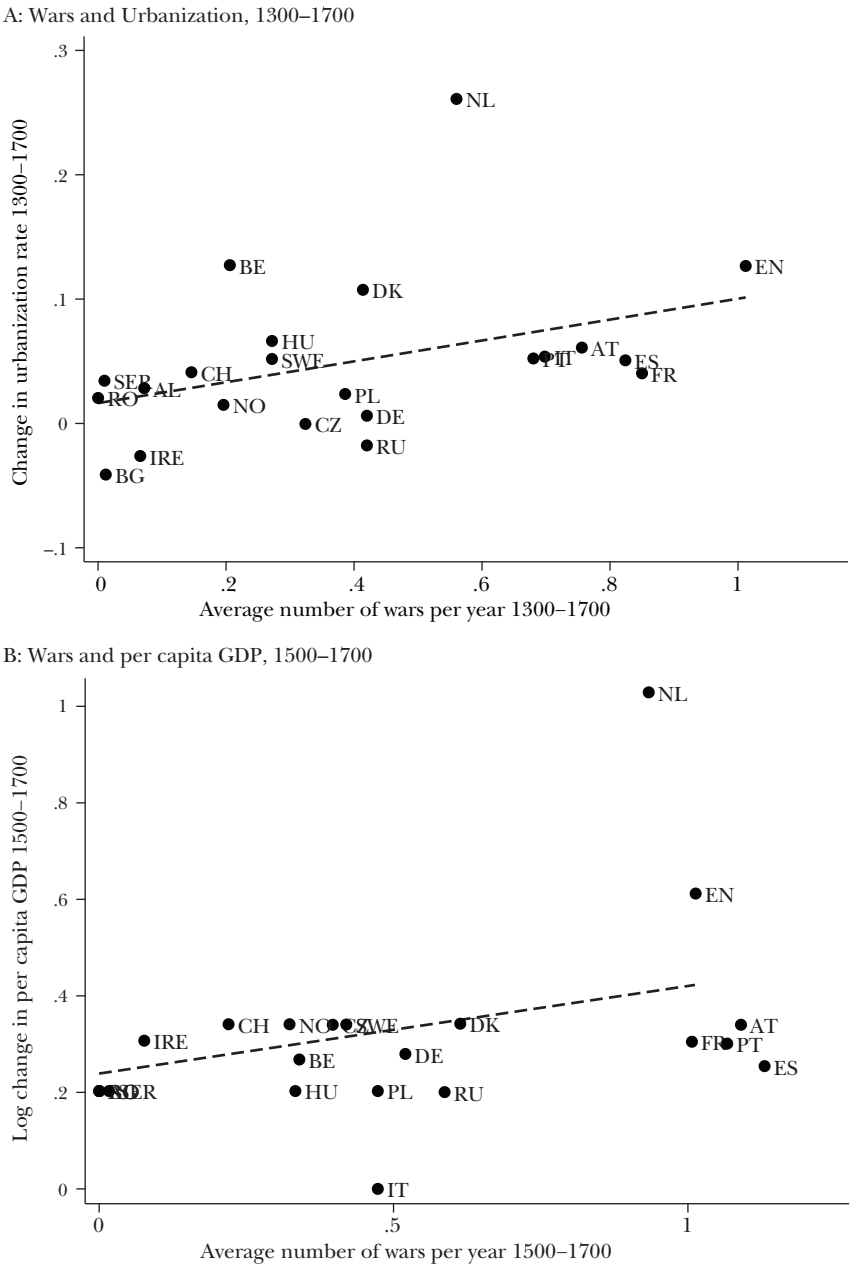
In simple bivariate regressions, we also find a large and statistically significant relationship between the number of wars and increases in urbanization. The baseline specification implies that in a country with one war per year on average between 1300 and 1700, urbanization rates rose faster, by 7.6 percentage points, over the same period, as compared to a country without warfare. Two close examples for these numbers are England, with 1.012 wars per year between 1300 and 1700 and a relatively high degree of urbanization, and Romania, with zero wars.[8] The former saw an increase in the urbanization rate by almost 13 percentage points, versus 2 in the latter.

The Chinese Mirror

So far, we have focused on patterns within Europe, but this begs the question why other parts of the globe did not "benefit" on a per capita basis in the same way. After all, war (and plagues) were hardly a European prerogative. Here, we focus on the case of China, which is particularly instructive because its starting position in the Middle Ages was seemingly so strong: It was politically unified, had a career bureaucracy chosen by competitive exam, and an impressive track record in terms of useful inventions and innovations.

[8] We also experimented with using country fixed effects, as well as adding a dummy variable for Western Europe, for Roman heritage, and so on. None of these alternative specifications undermines our result. Below, we discuss one prominent alternative interpretation (Acemoglu, Johnson, and Robinson 2005) in more detail. When we add their preferred variable, "Atlantic Coastline," the variable is statistically significant, but war remains highly significant. There is good evidence to suggest that trade opportunities across the Atlantic also helped to raise incomes. But all of our comparisons suggest that war mattered over and above the effects of trade.

Figure 4
Warfare and European Development



Sources: Change in urbanization from Bairoch, Batou, and Chèvre (1988), in combination with population estimates from McEvedy and Jones (1978), as explained in Voigtländer and Voth (2013); changes in per capita income from Maddison (2001). Average wars per year are derived from the dataset used in Acemoglu, Johnson, and Robinson (2005), who construct years of warfare by time period based on Kohn (1999).

That China had fallen behind significantly in per capita income by the early 19th century is not in doubt. Urbanization rates were low; incomes were a small fraction of their European equivalents. The country was also politically and militarily weak, and was about to be humiliated at the hands of European powers.

Our interpretation attributes a good part of China's relative decline to its low levels of military conflict. After the Yuan Dynasty (1271–1368) was deposed in a series of revolts, comparative peace reigned. Under the Ming and Qing Dynasties (1368–1644; 1644–1912), the country remained politically unified for a half a millennium. Frequency of military conflict was dramatically lower in China: Europe saw 443 wars during the period 1500–1800 (a frequency of 1.48 wars per year), involving 1,071 major battles. The corresponding figures for China are 91 wars between 1350 and 1800 (a frequency of 0.2 per year) and only 23 major battles—most conflicts were peasant revolts (Tilly 1992; Jaques 2007). In other words, the frequency of war per year was 85 percent lower, and the number of major battles was 98 percent lower, than in Europe.

Not only was war less frequent in China, it also caused fewer deaths. The majority of the population tilled the soil within a few hundred kilometers of the eastern seaboard. There were few natural obstacles to population movements and trade. The epidemiological evidence, where it exists, suggests that disease pools were largely integrated by the year 1000 (McNeill 1976). As a result, diseases spread by troops did not have the same devastating impact in China as they did in Europe.

In this setting, China experienced considerable population pressure. During the early modern period, Chinese population increased by an estimated 280 percent; the corresponding figure for Europe is 140 percent (Maddison 2007). Europeans visiting China noticed the abundance and the cheapness of labor. As Malthus (1798) observed: "The country [China] is rather over peopled . . . and labour is, therefore, so abundant, that no pains are taken to abridge it."

The view that China fell behind economically, and that its demography is partly to blame, is controversial. In *One Quarter of Humanity*, Lee and Wang (2001) challenged the earlier consensus that Chinese fertility rates were much higher than European ones. While marriage was universal, they argued, within-marriage fertility rates were relatively low. The current consensus view is that there is some merit in the argument but that total fertility rates in Europe were probably still markedly lower (especially taking into account how much higher incomes were).[9]

If population pressure in China was much higher than in Europe but higher fertility rates were only part of the answer, then lower mortality must be part of the story. Notice that China being poorer should actually have produced relatively *high* mortality rates: after all, many health risk factors and diseases before 1900 were nutrition sensitive, and lower incomes probably resulted in higher death rates from tuberculosis and the like. Thus, the absence of major war-induced mortality is a

[9] The relevant variable in a Malthusian setting is the income-adjusted fertility rate, accounting for the upward-sloping birth schedule in the Malthusian model. Total fertility rates measure the expected number of children a woman would have over the course of her life.

plausible explanation for why Ming and Qing China experienced such a substantial population boom.

Alternative Interpretations

We are not the first to examine Europe's relative rise to riches after 1500. Alternative interpretations have emphasized the role of technological innovations, of institutional improvements, and of fertility restriction. Theoretically, it is possible that a positive income shock driven by all or one of these factors gave rulers the means to fight more—explaining the positive correlation between war and income growth without any causal connection. Such an alternative interpretation is unlikely, for several reasons.

Acemoglu, Johnson, and Robinson (2005) show that European outperformance was largely driven by states bordering the Atlantic in combination with institutions that fostered commerce. In England and the Dutch Republic, trading opportunities strengthened the bourgeoisie, which in turn succeeded in constraining the powers of rulers. On the Iberian Peninsula, in contrast, the discovery of the Americas gave extra resources to powerful monarchs; as a result, institutional quality declined. The implication of this argument is that North-Western Europe owed its precocious lead over the rest of the world to institutional improvements, most of which occurred along the Western seaboard of the continent. Their interpretation is part of a broader approach to European—and in particular, British—outperformance.

Another prominent interpretation emphasizes Europe's growing ability to innovate (Mokyr 1992) and contrasts it with technological decline elsewhere. While medieval Europe had even forgotten some of the useful inventions of Rome—such as concrete—technological creativity flourished after 1500. From the invention of the printing press with movable letters and the barometer to vastly improved sailing ships, steel ploughs, and hot air balloons, Europe excelled at producing new and useful goods in the early modern period. In contrast, the famous "four great inventions" of China—compass, gunpowder, printing, and papermaking—marked an even earlier period of technological advance that found no echo in the early modern period. The underlying reasons for Europe's technological advance at this time are still a subject of research, but it seems plausible that the shortage of labor helped to encourage a search for labor-saving devices and that ongoing military conflict created pressure for innovations and a conduit for spreading ideas (Allen 2009b).

The principal problem with both the technological and the institutional interpretation is that they are not well-suited to explaining income divergence in a world dominated by demographic forces. Ashraf and Galor (2011) demonstrate that there are no significant gains in per capita incomes from productivity improvements during the Malthusian era. The reason is that human populations typically grow rapidly when faced with abundance. For productivity improvements to push up per capita living standards, they would have to be faster than the rate of population

growth.[10] In terms of orders of magnitude, this was highly improbable in the period before 1950. As noted earlier, human populations can easily grow at more than 3 percent per year, while technological change was probably less than 0.1 percent per year on average.

This leaves demographic interpretations. In any Malthusian system, incomes are ultimately determined by mortality and fertility rates. Europe's level of mortality was uniquely high, and war was an important component of it. Other factors also contributed to the fact that, at least for some part of this time period, mortality rates could rise at the same time as incomes grew. As incomes rose, Europeans crowded into more and larger urban centers. Cities in Europe before 1850 were veritable death traps, with mortality rates much higher than fertility rates. Poor sanitation and urban overcrowding were to blame. Therefore, not only did the development of cities reflect rising per capita incomes; these cities also helped to sustain incomes, much in the same way as war did, by keeping land-labor ratios high.

While the mechanism of disease-ridden cities adding to mortality is the same as for war, it is quantitatively much smaller. Even as late as 1800, only 10 percent of Europeans lived in cities with more than 10,000 inhabitants (DeVries 1984). Even if these city-dwellers suffered markedly higher mortality, they could not influence aggregate death rates by much.[11]

A similar argument applies to the effect of trade. Trade typically increases with incomes, and it can act as a potent vector for diseases. While little is known quantitatively about the volume of trade before 1800, it is reasonable to assume that increasing contact between distant population centers led to the exchange of germs and higher mortality. In related work, we estimate the size of the plausible effects, and find that they are quantitatively small, adding no more than 0.25 percentage points to annual European death rates of approximately 3 percent (Voigtländer and Voth 2013).

Another factor that could have helped hold down Europe's population growth and thus improve its performance in a Malthusian world is fertility restriction. Europe evolved a unique form of fertility limitation. Europeans of this time typically married late—in their mid-20s, not much earlier than they do today. A significant share of women also remained unmarried (Hajnal 1965). The reasons for this phenomenon (which only existed west of a line from St. Petersburg in Russia to Trieste in Italy) are complex. Most interpretations emphasize economic factors, such as the culturally determined need for newlyweds to set up a new household ("neo-locality"), the access of women to urban labor markets, inheritance rules, and the increasing use of females in pastoral agriculture (De Moor and van Zanden 2010; Voigtländer and Voth forthcoming).

[10] In Voigtländer and Voth (2013), we show that in a calibrated Malthusian model, even a sudden jump from technological stagnation to ongoing technological growth at the early modern rate of 0.1 percent per year would not have had a substantial impact on per capita income or urbanization.

[11] There are some exceptions: In the Netherlands, for example, the urbanization rate was so high that urban mortality on its own may have increased overall death rates by 0.5 percentage points, relative to a baseline of 3 percent (Voigtländer and Voth 2013).

Conclusion

The "First Divergence"—Europe pulling ahead long before the Industrial Revolution—has long posed a puzzle for growth theorists and economic historians. In a world with strong Malthusian forces, incomes should not have had much scope to rise and then stay elevated over long periods. And yet, this is what happened in early modern Europe.

In this paper, we argue that a good part of Europe's precocious rise to riches reflected "gifts from Mars"—permanently high per capita incomes for the survivors were an indirect consequence of near-constant, and deadly, warfare. We first show that despite small army size and relatively primitive weapons technology, war in the centuries before 1800 was a potent destroyer of human lives. The main cause of death was not armed force itself, but the spreading of disease: A single army of 6,000–8,000 men, dispatched from La Rochelle by Cardinal Richelieu to fight in the Mantuan war in 1628 may have killed up to a million people by spreading the plague on its march from France to northern Italy (Landers 2003; Biraben 1975).

In a Malthusian world, population losses generally created higher incomes for the survivors—there was more land available per capita. These effects should have been transitory: as population recovered, gains in per capita output ought to have been reversed. After the Black Death hit in 1348–50, population losses were massive and so were gains in per capita income, but one would expect these gains to fade over time. However, with income gains much greater than what could be eroded by population growth in a generation or two, rulers found ample surplus income (over and above subsistence) to tax. As they appropriated this surplus to a growing extent after 1350, war frequency surged. Rulers effectively treated war as a luxury good, and as money became available, fought ever more of them. The high frequency of war in turn made it easier to sustain the gains in living standards for those who survived.

The war channel for greater riches was particularly potent in Europe because of political fragmentation. Plagues also hit Justinian Rome, China, and the Middle East (McNeill 1976), without similar consequences. The Black Death in 1348–50 only acted as a catalyst for a simultaneous rise in the frequency of warfare and of per capita incomes because there were so many European states and statelets that could fight each other. And fight they did: war became a near-constant feature of early modern Europe.

In Dan Brown's (2013) bestseller, *Inferno,* the chief villain is a geneticist about to unleash a diabolical virus. He points to the experience of the Black Death to suggest that population losses can be beneficial and that good economic times will follow. Our research suggests that this is misguided—sometimes, history offers little guidance for policy implications today. Most parts of the world have clearly escaped from Malthusian constraints; land-labor ratios no longer determine per capita income except in the poorest countries. Instead, human capital, institutions, and technology are key. This also means that the synergistic link between war, population losses, and higher incomes that we described is unlikely

to occur again. As a result of technological advances, war today is vastly more destructive. It potentially annihilates capital stock on a greater scale than it did in Europe before 1800. In addition, due to the complementarity between human capital and modern technology, the negative effects of population losses are much greater in modern wars. These changes in the nature of war and of production ensure that the silver lining caused by military conflict is much fainter today than it was in early modern Europe—effective resources per survivor will not increase as much after modern wars as they did when war was the "sport of kings."

References

Acemoglu, Daron, Simon Johnson, and James Robinson. 2005. "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American Economic Review* 95(3): 546–79.

Allen, Robert C. 2009a. "Agricultural Productivity and Rural Incomes in England and the Yangtze Delta, c.1620–c.1820." *Economic History Review* 62(3): 525–50.

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

Allen, Robert C., Jean-Pascal Bassino, Debin Ma, Christine Moll-Murata, and Jan Luiten van Zanden. 2011. "Wages, Prices, and Living Standards in China, 1738–1925: In Comparison with Europe, Japan, and India." *Economic History Review* 64(Issue Supplement S1): 8–38.

Anderson, Michael, and Ronald Lee. 2002. "Malthus in State Space: Macro Economic-Demographic Relations in English History, 1540 to 1870." *Journal of Population Economics* 15(2): 195–220.

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review* 101(5): 2003–41.

Bairoch, Paul, Jean Batou, and Pierre Chèvre. 1988. *La population des villes européennes, 800-1850: Banque de données et analyse sommaire des résultats [The population of European cities, 800–1850: Data bank and short summary of results.]* Geneva, Switzerland: Droz.

Biraben, Jean Noël. 1975. *Les hommes et la peste en France et dans les pays Européens et Méditerranéens.* Paris: Mouton.

Bosker, Maarten, Eltjo Buringh, and Jan Luiten van Zanden. 2008. "From Baghdad to London: The Dynamics of Urban Growth in Europe and the Arab World, 800–1800." CEPR Discussion Paper 6833.

Broadberry, Stephen, and Bishnupriya Gupta. 2012. "India and the Great Divergence: An Anglo-Indian Comparison of GDP per Capita, 1600–1871." Available at: http://www2.lse.ac.uk/economicHistory/whosWho/profiles/sbroadberry.aspx.

Brown, Dan. 2013. *Inferno.* Doubleday.

Chen, Shuo, and James Kai-sing Kung. 2012. "A Malthusian Quagmire? Maize, Population Growth, and Economic Development in China." Available at SSRN: http://ssrn.com/abstract=2102295.

Clark, Gregory. 2007. *A Farewell to Alms: A Brief Economic History of the World.* 1st edition. Princeton University Press.

Clodfelter, Micheal. 2002. *Warfare and Armed Conflicts: A Statistical Reference to Casualty and Other Figures, 1618–1990.* McFarland.

Comin, Diego, William Easterly, and Erick Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics* 2(3): 65–97.

Crafts, Nicholas, and Terence C. Mills. 2009. "From Malthus to Solow: How Did the Malthusian Economy Really Evolve?" *Advances in Historical Macroeconomics* 31(1): 68–93.

Cunningham, Andrew. 2000. *The Four Horsemen of the Apocalypse: Religion, War, Famine and Death in Reformation Europe.* Cambridge University Press.

Curtin, Philip D. 1998. *Disease and Empire: The Health of European Troops in the Conquest of Africa,* 1st edition. Cambridge University Press.

De Moor, Tine, and Jan Luiten van Zanden. 2010. "Girl Power: The European Marriage Pattern

and Labour Markets in the North Sea Region in the Late Medieval and Early Modern Period." *Economic History Review* 63(1): 1–33.

DeVries, Jan. 1984. *European Urbanization: 1500–1800.* London: Taylor & Francis.

Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies.* New York: Norton.

Dupâquier, Jacques. 1988. *Histoire de la population Française,* tome 1: *Des origines à la renaissance.* Paris: Presses Universitaires de France (PUF).

Ellman, Michael, and S. Maksudov. 1994. "Soviet Deaths in the Great Patriotic War: A Note." *Europe-Asia Studies* 46(4): 671–80.

Ferguson, Niall. 2002. *The Cash Nexus: Money and Power in the Modern World, 1700–2000.* New York: Basic Books.

Ferguson, Niall. 2006. *The War of the World: Twentieth-Century Conflict and the Descent of the West.* Harmondsworth: Penguin Press.

Galor, Oded. 2005. "From Stagnation to Growth: Unified Growth Theory." Chap. 4 in *Handbook of Economic Growth,* Vol. 1A., edited by Philippe Aghion and Steven N. Durlauf. Elsevier.

Hajnal, J. 1965. *European Marriage Patterns in Perspective.* Population in History. London: Arnold.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114(1): 83–116.

Hansen, Gary D., and Edward C. Prescott. 2002. "Malthus to Solow." *American Economic Review* 92(4): 1205–17.

Hobsbawm, Eric J. 1994. *Age of Extremes: The Short Twentieth Century, 1914–1991.* UK: Michael Joseph.

Hubert, Michel. 1998. *Deutschland im Wandel: Geschichte der Deutschen Bevölkerung seit 1815.* Franz Steiner Verlag.

Huntington, Samuel P. 1996. *The Clash of Civilizations and the Remaking of the World Order.* Simon and Schuster.

Jaques, Tony. 2007. *Dictionary of Battles and Sieges: A Guide to 8,500 Battles for Antiquity through the Twenty-first Century.* Greenwood Publishing Group.

Karaman, K. Kivanç, and Şevket Pamuk. 2010. "Ottoman State Finances in European Perspective, 1500–1914." *Journal of Economic History* 70(3): 593–629.

Kelly, Morgan. 2005. "Living Standards and Population Growth: Malthus was Right." University College Dublin working paper.

Knecht, R. J. 1996. *The French Wars of Religion, 1559–98,* 2nd edition. Longman.

Kohn, George Childs, ed. 1999. *Dictionary of Wars,* 3rd edition. Checkmark Books.

Landers, John. 2003. *The Field and the Forge: Population, Production, and Power in the Pre-Industrial West.* Oxford University Press.

Landers, John. 2005. "The Destructiveness of Pre-Industrial Warfare: Political and Technological Determinants." *Journal of Peace Research* 42(4): 455–70.

Landes, David S. 1999. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor.* New York: W. W. Norton.

Lee, James Z., and Feng Wang. 2001. *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700–2000.* Cambridge, MA: Harvard University Press.

Maddison, Angus. 2001. *The World Economy,* Vol. 1: *A Millennial Perspective.* Paris: OECD.

Maddison, Angus. 2007. *Contours of the World Economy 1–2030 AD: Essays in Macro-Economic History.* Oxford University Press.

Malthus, Thomas Robert. 1798. *An Essay on the Principle of Population.* London.

McEvedy, Colin, and Richard Jones. 1978. *Atlas of World Population History.* Puffin.

McNeill, William H. 1976. *Plagues and Peoples.* New York: Knopf.

Mitchell, Brian R. 1988. *British Historical Statistics.* Cambridge University Press.

Mokyr, Joel. 1992. *The Lever of Riches: Technological Creativity and Economic Progress.* New York: Oxford University Press.

Neutzner, Matthias. 2010. *Abschlussbericht der Historikerkommission zu den Luftangriffen auf Dresden zwischen dem 13. und 15. Februar 1945.* City of Dresden.

Nicolini, Esteban A. 2007. "Was Malthus Right? A VAR Analysis of Economic and Demographic Interactions in Pre-industrial England." *European Review of Economic History* 11(1): 99–121.

Nunn, Nathan, and Nancy Qian. 2011. "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment." *Quarterly Journal of Economics* 126(2): 593–650.

O'Rourke, Kevin H., and Jeffrey G. Williamson. 2002. "After Columbus: Explaining Europe's Overseas Trade Boom, 1500–1800." *Journal of Economic History* 62(2): 417–56.

Parker, Geoffrey. 1977. *The Dutch Revolt.* Penguin.

Piotrowski, Tadeusz. 1997. *Poland's Holocaust: Ethnic Strife, Collaboration with Occupying Forces and Genocide in the Second Republic, 1918–1947.* McFarland & Company.

Pomeranz, Kenneth. 2001. *The Great Divergence: China, Europe, and the Making of the Modern World Economy.* Revised. Princeton, NJ: Princeton University Press.

Putterman, Louis, and David N. Weil. 2010. "Post-1500 Population Flows and the Long-Run

Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics* 125(4): 1627–82.

Spolaore, Enrico, and Romain Wacziarg. 2009. "The Diffusion of Development." *Quarterly Journal of Economics* 124(2): 469–529.

Statistics Bureau of Japan. 2011. *Population Estimates of Japan 1920–2000.* http://www.e-stat.go.jp/SG1/estat/ListE.do?bid=000000090004&cycode=0.

Tallett, Frank. 1992. *War and Society in Early-Modern Europe, 1495–1715.* Routledge.

Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990–1990.* Cambridge, MA: Blackwell.

Tilly, Charles. 1992. *Coercion, Capital and European States: AD 990–1992.* Revised. Oxford: Wiley-Blackwell.

US Census Bureau. 2000. "Historical National Population Estimates: July 1, 1900 to July 1, 1999." Population Estimates Program, Population Division, US Census Bureau. http://www.census.gov/population/estimates/nation/popclockest.txt.

Voigtländer, Nico, and Hans-Joachim Voth. 2006. "Why England? Demographic Factors, Structural Change and Physical Capital Accumulation during the Industrial Revolution." *Journal of Economic Growth* 11(4): 319–61.

Voigtländer, Nico, and Hans-Joachim Voth. 2009. "Malthusian Dynamism and the Rise of Europe: Make War, Not Love." *American Economic Review* 99(2): 248–54.

Voigtländer, Nico, and Hans-Joachim Voth. 2013. "The Three Horsemen of Riches: Plague, War, and Urbanization in Early Modern Europe." *Review of Economic Studies* 80(2): 774–811.

Voigtländer, Nico, and Hans-Joachim Voth. Forthcoming. "How the West 'Invented' Fertility Restriction." *American Economic Review.*

Wells, H. G. (Herbert George). 1905. *A Modern Utopia.* London: Scribner's.

Wrigley, E. A. 1985. "Urban Growth and Agricultural Change: England and the Continent in the Early Modern Period." *Journal of Interdisciplinary History* 15(4): 683–728.

Wrigley, E. A., and R. S. Schofield. 1989. *The Population History of England 1541–1871.* Cambridge University Press.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 187–210

The Economics of Slums in the Developing World[†]

Benjamin Marx, Thomas Stoker, and Tavneet Suri

Urban populations have skyrocketed globally and today represent more than half of the world's population. In some parts of the developing world, this growth has more-than-proportionately involved rural migration to informal settlements in and around cities, known more commonly as "slums"—densely populated urban areas characterized by poor-quality housing, a lack of adequate living space and public services, and accommodating large numbers of informal residents with generally insecure tenure.[1] Worldwide, at least 860 million people are now living in slums, and the number of slum dwellers grew by six million each year from 2000 to 2010 (UN-Habitat 2012a). In sub-Saharan Africa, slum populations are growing at 4.5 percent per annum, a rate at which populations double every 15 years.

The global expansion of urban slums poses questions for economic research, as well as problems for policymakers. Some economists (Frankenhoff 1967; Turner 1969; World Bank 2009; Glaeser 2011) have suggested a "modernization" theory of

[1] Perhaps not surprisingly, the identification of slum inhabitants suffers from the lack of a consistent terminology—for example, "slums" and "squatter settlements" are used almost interchangeably, although tenure and ownership institutions vary greatly across informal settlements. UN-Habitat (2006) applies the notion of "slum household" to any household lacking access to improved water, improved sanitation, sufficient living area, durable housing, and secure tenure. Slum areas are generally thought of as geographic areas accommodating informal residents that combine several of these characteristics.

■ *Benjamin Marx is a PhD student at the Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Thomas Stoker is Gordon Y. Billard Professor in Management and Economics and Professor of Applied Economics, and Tavneet Suri is Maurice F. Strong Career Development Professor and Associate Professor of Applied Economics, both at the MIT Sloan School of Management, Cambridge, Massachusetts. Their email addresses are bmarx@mit.edu, tstoker@mit.edu, and tavneet@mit.edu.*

[†] To access the disclosure statements, visit
http://dx.doi.org/10.1257/jep.27.4.187

doi=10.1257/jep.27.4.187

slums: according to this thinking, slums are a transitory phenomenon characteristic of fast-growing economies, and they progressively give way to formal housing as economic growth trickles down and societies approach the later stages of economic development. Even if slum areas appear stable in the short- or medium-term, this argument holds, slum living only represents a transitory phase in the life cycle of rural migrants: the slum dwellers or their children eventually move into formal housing within the city, so that the benefits of migration into the slum get passed along from generation to generation.

But even if urban poverty is preferable to rural poverty, as apparently shown by the revealed preference of migrants, life in a slum is very difficult and often subsistence-level. Moreover, slums do not always seem to be a temporary phenomenon of migration to cities: in many countries slum areas have been growing for decades, and millions of households find themselves trapped in slums for generations. This might suggest that today's slums pose a problem of a different nature: because of multiple market and policy failures, acute governance and coordination problems that hinder investment, and unsanitary living conditions affecting the dwellers' human capital, life in the slum might constitute a form of poverty trap for a majority of their residents.

In this essay, we provide historical and contemporary facts to argue that the type of poverty observed in contemporary slums of the developing world is characteristic of that described in the literature on poverty traps. We document how human capital threshold effects, investment inertia, and a "policy trap" may prevent slum dwellers from seizing economic opportunities offered by geographic proximity to the city. We then discuss whether the basic assumptions of the "modernization" view hold: that is, whether there is a relationship between economic growth, urban growth, and slum growth in the developing world, and whether standards of living of slum dwellers are improving over time, both within slums and across generations. Finally, we discuss why standard policy approaches have often failed to mitigate the expansion of slums in the developing world. Our aim is to stimulate serious academic interest and to inform public debate on the essential issues posed by slums in the developing world.

A Contemporary Perspective on Slums

Slums are not, of course, a new phenomenon. They were a distinctive feature of European and US cities during the Industrial Revolution, and they persisted in some of these cities well into the twentieth century. The well-known slums of the past were often on the outskirts of dynamic economic growth, which both attracted migrants and offered them some access to economic opportunities. For example, the Whitechapel area of East London attracted a vast number of poor rural migrants during the eighteenth and nineteenth century due to the new factories and shops of that part of the city. The Hell's Kitchen area of New York City on the Hudson River side of Manhattan attracted immigrants in large part because of its proximity to docks and railroads, as well as to the growing city nearby. In the past, moderate to radical policy solutions were adopted to address the overcrowding and

Table 1
Two Lists of the Developing World's Largest Slums

UN-Habitat (2003)			***Davis (2006)***		
Name of slum	*City, Country*	*Population estimate*	*Name of slum*	*City, Country*	*Population estimate*
Dharavi	Mumbai, India	Over 500,000	Neza/Chalco/Izta	Mexico City, Mexico	4 million
Orangi	Karachi, Pakistan	Over 500,000	Liberatador	Caracas, Venezuela	2.2 million
Kibera	Nairobi, Kenya	400,000	El Sur/Ciudad Bolivar	Bogota, Columbia	2.0 million
Villa el Salvador	Lima, Peru	300,000	San Juan de Lurigancho	Lima, Peru	1.5 million
Ashaiman	Tema, Ghana	150,000	Cono Sur	Lima, Peru	1.5 million
			Ajegunle	Lagos, Nigera	1.5 million
			Sadr City	Baghdad, Iraq	1.5 million
			Soweto	Gautung, South Africa	1.5 million
			Gaza	Palestine	1.3 million
			Orangi	Karachi, Pakistan	1.2 million

Source: UN-Habitat (2003); Davis (2006).
Notes: UN-Habitat (2003) does not provide a comprehensive list of slums ranked by their population. The numbers given are mentioned in the report as part of individual case studies of slums.

unsanitary conditions in these types of areas. Examples of such policies include Baron Haussman's revamping of Paris in the 1860s–1870s, which involved altering more than half of the city's buildings and creating a sewage system and wide boulevards in lieu of slum neighborhoods. More recently, Singapore's compulsory savings scheme in the 1960s was used to finance the construction of public housing and to enable slum residents to purchase formal housing units at a subsidized rate.

Today, large slum settlements have disappeared in most advanced economies, but it is far from clear how comparable these historical examples are to the situations faced in the developing world. Some of today's slums are in countries experiencing rapid economic growth, such as China, but many slums are located in countries with slow or stagnant growth. The prevalence of slums is highest in sub-Saharan Africa, where slum dwellers represent 62 percent of the urban population (UN-Habitat 2012a): as of 2005, the three countries with the highest fraction of the urban population living in slums—Sierra Leone, Sudan, and the Central African Republic—are also located on that continent. Estimates of actual populations on a slum-by-slum basis across the developing world are few and far between and vary widely across sources. Table 1 displays two lists, giving a sense of the range of measurements that

can occur. The left-hand column lists five of the world's largest slums, with rough population estimates, as provided by UN-Habitat (2003). An unofficial ranking provided in Davis (2006) lists 30 slums with over 500,000 inhabitants, including ten in sub-Saharan Africa and nine in Latin America. The ten slums with largest population by this measure are shown in the right-hand columns of Table 1. Dharavi and Kibera, the first and third slums in the UN-Habitat (2003) list both appear further down in the Davis (2006) ranking with an estimated population of 800,000 each.

Of course, the fact that a lot of the global slum expansion takes place in poor economies does not invalidate a "modernization" or transitory view of slums. Since rural migration continues unabated even in those countries, urban productivity must be rising relative to rural productivity, either because capital accumulation and technological progress are concentrated in cities or because rural productivity is declining. The location decision of slum dwellers also indicates that standards of living in slums are somewhat preferable to those in the rural hinterlands. Glaeser (2011) provided evidence that the urban poor worldwide are on average richer and happier than their rural counterparts, and Chowdhury, Mobarak, and Bryan (2009) showed that seasonal urban migration in Bangladesh can generate welfare improvements for families of migrants.

However, these facts say little about the nature of poverty in slums—and whether it can be escaped via the competitive market forces offered by the city. There has been, in fact, a relative lack of empirical research work conducted in slums of the developing world to test this argument.[2] One reason is that data collection in slums is problematic due to a variety of factors, including safety issues for research fieldworkers, the high mobility and turnover rates of respondents, and the fact that target households are regularly absent from their dwellings. As a result, few studies have tried to document the degree of intergenerational social mobility of slum dwellers in the developing world. Here, we start by illustrating what life is like in slums and what characteristics seem to be common across slums, using recent surveys collected in slums of four countries: Bangladesh, India, Kenya, and Sierra Leone. We rely most heavily on the case of the Kibera slum in Nairobi, Kenya, where we have conducted extensive fieldwork over the past two years. Table 2 provides specific sources of data for these slum areas.

Slums as Poverty Traps

In this paper, we argue that slums may be poverty traps and are therefore neither temporary nor a short stop on the way to greater economic opportunities.

[2] A nonexhaustive list of recent empirical research conducted in slums includes Banerjee, Duflo, Glennerster, and Kinnan (2010) in Hyderabad; Banerjee, Pande, Vaidya, Walton, and Weaver (2011) in Delhi; El-Zanaty and Way (2004) in Cairo; Field (2007) in Peru; Galiani et al. (2013) in slums of three Latin American countries; Gulyani, Bassett, and Talukdar (2012) in Dakar and Nairobi; and Perlman (2010) in Rio de Janeiro. Gupta, Arnold, and Lhungdim (2009) analyze the NFSH-3 dataset on which we also rely.

Table 2
Description of Datasets

Country	*Locations*	*Source*	*Year*	*Number of slum households surveyed*
Bangladesh	Tongi, Jessore	SHAHAR Project/CARE-Bangladesh (Baseline census)	2000	26,830
Bangladesh	Tongi, Jessore	SHAHAR Project/CARE-Bangladesh (Baseline survey)	2000	1,120
India	Delhi, Meerut, Kolkata, Indore, Mumbai, Nagpur, Hyderabad, Chennai	National Family Health Survey (NFHS-3)	2005–2006	8,669
India	Hyderabad	Banerjee, Duflo, Glennerster, and Kinnan (2010)	2005	2,800
Kenya	Kibera (Nairobi)	Kenya National Bureau of Statistics (national census)	1999, 2009	64,588 (approx.)
Kenya	Kibera (Nairobi)	Marx, Stoker, and Suri (2013)	2012	31,765
Kenya	Kibera (Nairobi)	Marx, Stoker, and Suri (2013)	2012	1,093
Sierra Leone	Western Urban Area (Freetown)	UNICEF Sierra Leone SMART Survey	2010	789

Sources: The Bangladesh data is available from http://dvn.iq.harvard.edu/dvn/ and was collected in 2000 as part of the SHAHAR census and baseline surveys. The India data was collected with approximately 8,669 slum households in eight Indian cities as part of the 2005–2006 National Family Health Survey (NFSH-3). It is available from http://www.measuredhs.com/ (India Standard DHS, 2005–06). The Hyderabad dataset (Banerjee, Duflo, Glennerster, and Kinnan 2010) was collected as part of a randomized evaluation on the impact of microfinance. It is publicly available from the data repository of the Jameel Poverty Action Lab (J-PAL) at http://thedata.harvard.edu/dvn/dv/jpal. Our Kenya data come from two different sources. First, we collected one census and one household survey in Kibera, Nairobi's largest slum area. Second, we obtained from the Kenya National Bureau of Statistics (KNBS) two waves of complete micro census data collected in the same area in 1999 and 2009. The Sierra Leone data was collected as part of the 2010 UNICEF SMART Survey and is proprietary from UNICEF. We use these sources throughout unless otherwise indicated.

There is a wide literature on poverty traps, including a theoretical literature highlighting the specific mechanisms leading to poverty traps (for excellent definitions and reviews, useful starting points include Basu 2003, Matsuyama 2005, and Bowles, Durlauf, and Hoff 2006). The literature has also described spatial poverty traps, but mostly in rural settings (Jalan and Ravallion 2002; Golgher 2012). We argue that urban slums present a different challenge to communities and governments administering them, and that the very nature of life in the slums makes it difficult to achieve improvements in standards of living through marginal investments in housing, health, or infrastructure alone. We now discuss some of the mechanisms relevant to slum contexts that may lead to poverty traps.

Human Capital

Despite tremendous variations across slums, issues common to all slum settings are a lack of adequate living space, insufficient public goods provision, and the poor quality of basic amenities, all of which lead to extremely poor health and low levels of human capital.[3] In this section, we focus on the health aspects of human capital. Education, albeit another important component of human capital, is a less relevant metric for our argument given that universal free primary education laws have reduced disparities in access to education between rural and urban settings,[4] and that rural–urban differences in the quality of education are extremely difficult to measure.

In the Kibera area of Kenya, informal households reported in 2009 an average dwelling size of 1.17 habitable rooms (with average household size of 3.15), as opposed to 1.95 for urban households and 2.97 for rural households. For perspective, according to UN-Habitat (2006), a dwelling provides "sufficient living space" if each room is shared by no more than three individuals. In the Zimbabwe slum in Abidjan, population density was reported to be as high as 34,000 inhabitants per square kilometer (UN-Habitat 2003). As a point of comparison, Manhattan's population density was 26,924 per square kilometer in 2010 (US Census Bureau 2013).

Across slum settings, the adverse health effects of overcrowding are aggravated by poor access to water and sanitation facilities. Table 3 shows that the majority of slum dwellers across our datasets have no private latrine, and many use inferior-type latrines (such as an open space or traditional pit that is not connected to a sewage network), no source of private water, and no garbage collection (meaning that garbage is either left in a roadside ditch or burnt next to the household dwelling). These data are corroborated by a range of studies documenting the poor water access and overall hygiene of slum neighborhoods. For example, in Mumbai's Shiva Shakti Nagar slum, community taps are reportedly shared by 100 people on average (World Bank 2009). In a survey of slum dwellers in Delhi, Banerjee et al. (2011) found that the environment of 83 percent of toilet sites was infected with fecal or other waste matter.

Absent or deficient water and sewage systems translate into a broad range of health and sanitation issues, whether through direct exposure to bacterial agents, contaminated drinking water, or other channels. Duflo, Galiani, and Mobarak (2012) described the disease burden arising from the unsanitary living conditions in slums. In the slums of Tongi and Jessore in Bangladesh, 82 percent of respondents report any household member being sick in the past 30 days. In Kibera, 16 percent of our respondent households have at least one member chronically ill in the previous three months. In Sierra Leone, a country whose slums routinely experience cholera

[3] The World Bank (2009) argued that urban areas fare consistently better than rural areas worldwide along a variety of health indicators. In this section, we compare rural areas with slum areas specifically.

[4] Lopez (2007) shows evidence of this trend for Latin America. Hannum, Wang, and Adams (2008) study the case of China, where some urban–rural disparities remain, but the primary enrollment of 7–16 year olds in rural areas is nearly universal. Worldwide, school attendance rates have increased in the majority of low-income countries since 1990, and rural areas were the primary beneficiaries of this trend (World Bank 2009).

Table 3
Public Goods and Basic Amenities across Slums

	No private latrine	*Inferior latrine type*	*No private water source*	*No garbage collection*
Tongi (Dhaka)	70%	34%	81%	64%
Hyderabad	46%	43%	61%	NA
Kibera	NA	63%	92%	73%
Kolkata	75%	46%	57%	NA
Mumbai	78%	8%	12%	NA
All Indian slums NFSH-3	68%	49%	25%	NA

Source: Authors using data from the SHAKAR Project for Bangladesh; Marx, Stoker, and Suri (2013) for Kenya; and the 2005–2006 National Family Heath Survey (NFSH-3) for India.
Notes: In the **first column**, we report the fraction of slum households who share their latrine with other households. In the **second column**, we report the fraction of households using an open space or traditional pit as latrine. To compute this number, we combine "open air" and "septic tank/pit toilet" in Hyderabad; "traditional pit latrine," "bucket," and "bush or river or stream" in Kibera; "flush to septic tank," "flush to pit latrine," "flush to somewhere else," "flush, don't know where," "ventilated improved pit latrine," "pit latrine with slab," "pit latrine without slab/open pit," "no facility/bush/field," "composting toilet," and "dry toilet" in the 2005–2006 National Family Heath Survey (NFSH-3) data (Kolkata, Mumbai and "All Indian slums"). In the **third column**, we report the fraction of households who share their main drinking water source with other households. This combines "share restricted" and "shared unrestricted" in the Bangladesh data, "public tap" and "public water tank" in Kibera, "public tap/standpipe" and "tube well or boreholes outside the dwelling" in the NFSH-3 data. In the **fourth column**, we report the fraction of households whose garbage is not collected by a public or a private company. "NA" means "not available."

outbreaks, slum households exhibit poorer health outcomes than their rural counterparts. The prevalence of underweight, stunting, and wasting (acute malnutrition) is in fact greater in the slum outskirts of the capital Freetown than in rural areas nationwide, as children under five living in slums have significantly lower weight-for-age and weight-for-height indexes than children under five in rural areas.[5] Across cities in the developing world, there is some evidence that life expectancy is lower, and infant mortality higher among the urban poor than among comparable groups in rural and formal urban areas (Bradley, Stephens, Harpham, and Cairncross 1992).

Health and sanitation issues are rendered more problematic by the lack of provision of a social safety net in slums. Slum living involves a wide range of risks: in our Kibera data, 10 percent of households have experienced being evicted from their dwelling, and 4 percent report at least one death in the household in the past six months. In Bangladesh, 56 percent of respondents say they do not meet their basic needs of food, water, shelter, and healthcare; 48 percent do not feel safe in their house during bad weather; and of the households reporting one member being ill in the previous 30 days, 26 percent say that they could not afford to seek medical attention. In Hyderabad slums, where 70 percent of households classify themselves as "poor," only 12 percent report receiving any assistance from the government.

[5] Calculated by the authors using the UNICEF SMART Survey data for Sierra Leone.

A wide literature, both macroeconomic and microeconomic, documents the importance of health for income and of early health investments for longer-term outcomes: good reviews can be found in López-Casanovas, Rivera, and Currais (2005), Bleakley (2010), and Currie and Vogl (2013). In developing economies, there are large returns to health improvements and strong complementarities between investments in child health and child education (Miguel and Kremer 2004). Poor human capital and poor avenues for human capital investment in slum households may therefore lead to a lack of social mobility across generations of slum residents. Thus, this health data seems at odds with the "transitory" hypothesis, making it questionable whether slum inhabitants meet the "critical thresholds" in human capital required for the competitive forces of the labor market to come into play, as described theoretically in Azariadis and Drazen (1990). Slum dwellers may find themselves trapped in a low-skilled, low-income equilibrium as the continuous influx of rural migrants maintains wages at near-subsistence levels, hindering the investments in human capital that would be required to offset the adverse effects of slum living.

Investment Inertia

Slums not only seem trapped in a low-human-capital equilibrium, but they also exhibit dysfunctional institutions, low levels of physical capital, and poor access to developed services. Slums can be thought of as areas of depressed public and private investment where neither government nor broader society has managed to organize in a way that provides for widespread provision and maintenance of public goods (and we are defining "public good" broadly to include clean water, sanitation, garbage collection, a social safety net, and the legal infrastructure of property rights that allows for an effective market in land and housing). In this section, we describe five distinct phenomena that can lead to low investment in slums.

A first factor is the well-known informality of property rights intrinsic to slum areas. Without formal land titles, slum dwellers lack the incentives to improve the quality of their homes and neighborhoods. Informal settlements have typically emerged on vacant government land, which implies that the property rights over the land held by individuals living there are highly illiquid, although they may be enforceable locally. In Dakar and Nairobi, only 19 and 34 percent of owners respectively report that it is easy to transact housing in their area (Gulyani, Basset, and Talukdar 2012). De Soto (2000) popularized this argument, and suggested that a lower risk of eviction and tighter property rights on the land could unlock access to credit markets. More recently, Field (2005) and Galiani and Schargrodsky (2010) showed that formal titling could encourage investments in poor urban areas.

A second factor is the concurrence of overcrowding of slum areas and low marginal returns from small upgrading investments. It may therefore not be rational for slum dwellers to finance investments in housing or infrastructure. In addition, many upgrades may require rather large private investments. For example, Galiani et al. (2013) show how even simple improvements in housing in slums in Mexico, Uruguay, and El Salvador cost as much as $1,000 per household, although such investments do generate improvements in quality of life and safety. This situation

Table 4
Ownership Type of Main Dwelling

	Own	*Rent*	*Occupy*	*Other arrangements*
Tongi (Dhaka)	52%	41%	7%	0%
Hyderabad	68%	28%	3%	1%
Kibera	7%	92%	1%	1%
Kolkata	37%	56%	NA	8%
Mumbai	72%	26%	NA	2%

Source: Authors using data from the SHAKAR Project for Bangladesh; Marx, Stoker, and Suri (2013) for Kenya; and the 2005–2006 National Family Heath Survey (NFSH-3) for India.
Notes: "Occupy" refers to situations where the respondent lives in a dwelling without paying for rent and without holding a title on the land. "Other arrangements" include land given by the government in Hyderabad, land owned by a friend or relative in Kibera, and land obtained as part of an employment or any other arrangement in Mumbai and Kolkata. The "Other arrangements" category may include occupiers in Mumbai and Kolkata, although this is not clear from the survey questionnaire. "NA" means "not available."

stands in stark contrast to some of the problems that characterize rural poverty, where relatively cheap technologies can often lead to substantial improvements in income and welfare (for example, the results of "green revolution" technologies in agriculture). Not only are private investments in housing infrastructure low in slums, but Duflo, Galiani, and Mobarak (2012) have documented a low willingness-to-pay for improved public goods in poor urban areas. A "big push" approach would then seem necessary to address the lack of investment in slums, and to generate aggregate demand spillovers in the area of public goods and basic services—the seminal model described Murphy, Shleifer, and Vishny (1989) would justify this type of intervention.

A third, less well-known cause for low investment levels in slums could be the high rent premiums that dwellers must pay to live in close proximity to the city, and which reduce opportunities for savings accumulation. While slum dwellers are typically thought of as squatters occupying vacant public land, available evidence suggests that a large number of dwellers across slums are in fact rent-paying tenants, as shown in Table 4.

In a survey of city officials in Kibera (Marx, Stoker, and Suri 2013), we found that the two most common causes of landlord–tenant disputes in the slum were the tenant's inability to pay their rent and rent increases asked by the landlord. Looking at amounts paid in rent across different income brackets in Kibera, households in the poorest quantiles do not pay lower rents per square meter occupied (as shown in Table 5). This finding undermines the notion that rural migrants can pay a modest premium to live in close proximity to the city. Although rents for urban poor look rather low in amount (about 12 percent of consumption), most rural households (for example, 90 percent in Kenya) do not pay any rent. In 2008, the average monthly rent for rural households was 266 Kenyan shillings (about $3 US), or 1 percent of household consumption, as opposed to 3,303 Kenyan shillings ($39 US), or 10 percent

Table 5
Rent Prices across Consumption Quintiles in Kibera
(rent per month)

	1st quintile	*2nd quintile*	*3rd quintile*	*4th quintile*	*5th quintile*
Area rented per capita (square meters)	5.3	7.9	10.9	14.1	15.9
Rent per capita (Kenyan shillings)	310.9	435.2	488.6	666.0	1121.4
Rent per square meter (Kenyan shillings)	117.6	107.5	96.8	109.5	127.0

Source: Data from Marx, Stoker, and Suri (2013).
Notes: We trim the top 1 percent of the data for rents and area rented. The exchange rate was $1 = 85 Kenyan shillings. When looking at rents per capita in these data, it is hard to account for economies of scale in housing. Part of housing is a public good to the members of a household, and these rent figures are simply total rent paid divided by household size, not accounting for these economies of scale.

of consumption, for urban households (Jack and Suri forthcoming). In the Kibera slum, where food expenditure represents 61 percent of consumption, housing rents represent in fact almost a third of nonfood expenditure.[6]

A fourth set of factors that can cause low investment involve the extreme coordination failures and "governance gap" intrinsic to slum life. Widespread governance failures work against the prospects for the urban poor to find creative solutions to upgrade the quality of their neighborhoods (as envisioned in Turner and Fichter 1972). A large amount of anecdotal evidence suggests that allocation mechanisms in slums are inefficient and that private actors or bureaucratic entrepreneurs fill the governance space, as opposed to legitimate local governments or community representatives. For example, land and housing markets are often controlled by a handful of powerful or well-connected individuals: landlords, local bureaucrats, or gang members. Davis (2006) reported on the example of Mumbai, where it is claimed that 91 individuals control all vacant land.[7] The slums of Nairobi, where land permits on vacant land have been illegally awarded by the local administration since the 1970s, are a poster case. The Nairobi "slumlords" can often rely on the support of the local administration to settle rent disputes (Joireman 2011), and they may collude with local chiefs to discourage improvements in the housing infrastructure that could lead to more entrenched tenancy rights. In areas where chiefs are not able to enforce their authority, gangs sometimes fill the governance

[6] Across the developing world, poor households tend to spend a large share of their income on food, as discussed in this journal by Banerjee and Duflo (2007).
[7] Owning and renting out structures in slums can be immensely profitable to landowners: in Nairobi, Amis (1984) reported that annual returns on the housing capital stock in slums could be as high as 131 percent.

space to enforce rules of their own, levy taxes, and control expenditure and investments in their neighborhoods (Marx, Stoker, and Suri 2013). In other areas, the formal governance system is entirely absent and has been replaced by these other interests—an example is the role of drug cartels in the *favelas* of Rio de Janeiro (Ferraz and Ottoni 2013).[8]

A fifth potential contributor to low investment traps in slums comes from the well-known Todaro paradox (1976): slum living standards cannot be improved without generating an additional influx of rural migrants, which in turn depresses public and private investments in the existing settlements. This may give little incentive for the public sector to invest in infrastructure and public goods in slums. However, there is little rigorous evidence on the link between these "pull" factors and slum growth in the developing world, while "push" factors (such as overcrowding on the fertile lands and an overall decline in agricultural output) have been better documented. For instance, Lipton (1977) and Bates (1981) argued that the "urban bias" of policy and tax-based income transfers between peasants and city dwellers until the 1990s was a chief cause of rural–urban migration. The seminal model on the issue of rural–urban migration was that of Harris and Todaro (1970), who modeled the rural–urban wage gap as a driving force behind migration decisions. However, this work had little to say about locations decisions of migrants within cities, and we are not aware of any more recent theoretical attempt to model those location choices.[9]

The Policy Trap

In addition to being trapped in low-human-capital and low-investment equilibriums, slum areas are generally places of extreme policy neglect, well beyond the lack of public goods provision discussed above. This "policy trap" stems from political economy factors that are different from the market failures we just discussed.

First, the informal nature of slum neighborhoods implies that these areas are usually considered not eligible for urban planning or public upgrading projects. This may be for purely administrative reasons or because public investments could amount to more entrenched occupancy rights for the slum residents—an outcome that governments generally do not favor (Fox 2013). Over the past two decades, the few countries that made significant advances in the struggle against slum growth were those where political support was widespread for reducing the prevalence of slums and where a genuine political commitment was expressed for curbing slum expansion—for example, in Egypt and Mexico (UN-Habitat 2006).

Second, enumeration problems and the fact that slum populations are often (deliberately or mistakenly) undercounted distorts the weight assigned to slum areas

[8] Ferraz and Ottoni (2013) study the effects of a pacification program in the Rio *favelas* where military interventions were necessary to re-establish a police presence in the slums.

[9] For US cities, one canonical modeling approach to looking at location decisions in urban areas is to think of a city with a central business district and how housing, commerce, and schools are organized relative to that city center, along the lines of Alonso (1964), Muth (1969), and Mills (1972). But these sorts of models do not seem an apt description of the location decisions and outcomes in the cities of low-income countries.

in the political process. For example, Sabry (2010) showed how the undercounting of *ashwa'iyyat* (informal settlement) populations of Greater Cairo led to an undersampling of slum households in household surveys. In India, slum populations were comprehensively enumerated for the first time in 2001, but discrepancies in the state-level definitions of slums and the refusal of some states to validate the slum statistics resulted in "gross under-estimation/under-coverage of slum populations in the country" (Government of India 2011). Lack of representation can have dramatic consequences when issues of eviction are at stake. For example, the Makoko neighborhood in Lagos, Nigeria, one of the oldest slums in the world until its partial demolition in 2012, had not been covered by the country's last national census in 2007 (Babalola 2009). Not having accurate census data on slums is problematic for many reasons. The true population of the slum is unknown—for example, there has been controversy over the population of the Kibera slum, with estimates ranging from about 170,000 to over one million—and policy interventions are impossible to plan without accurate population numbers.[10]

Third, catering to the interests of the silent majority of slum dwellers might not even be in the best interest of the people in charge in the slum. As discussed above, planning or regulatory powers in slums often do not belong to legitimate governance bodies, but are usually split between a galaxy of private actors, landlords, chiefs and bureaucrats, and gangs. Conflicting interests between these actors, and policy conflicts between central government and municipal authorities could explain why "status quo" interests have often prevailed in slums (Fox 2013). In other words, maintaining high transaction costs and opaque governance mechanisms can be very beneficial to a minority of bureaucratic entrepreneurs willing to garner support in slum patronage politics or to extract rents from their informal control over the land (Amis 1984; Marx, Stoker, and Suri 2013). In line with this, Fox (2013) documents how members of Tanzania's ruling *Chama Cha Mapinduzi* (CMM) Party are involved in transactions on slum land markets. Syagga, Mitullah, and Karirah-Gitau (2002) found that 57 percent of landlords in Nairobi slums were public employees.

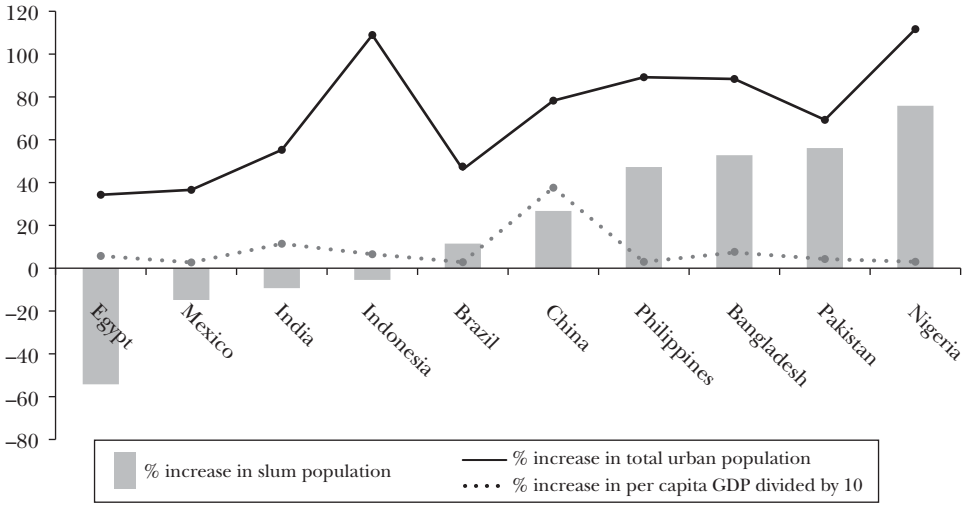
Slums and Economic Development

Slum Growth in a Cross-Country Perspective

The conceptualization of slums as places of poverty traps is at odds with a "modernization" view, which assumes that the prevalence of slums and urban poverty should decrease as markets develop and the forces of economic development come under way. Here we present some simple empirical facts on this hypothesis. In particular, is there a functional relationship between economic growth, urban

[10] Aside from the policy constraints that lack of data poses, the poor census implies that the sampling frame for any survey (national or specifically of the slum) cannot draw a representative random sample. It is therefore much harder to accurately collect socioeconomic indicators of conditions in a slum and understand how these conditions evolve over time.

Figure 1
Patterns in Urban and Slum Growth, 1990–2007



Source: Figure compiled by the authors using data from UN-Habitat's Global Urban Indicators online database (UN-Habitat 2013). The data on GDP per capita come from the World Bank's World Development Indicators online database.
Notes: Figure 1 compares slum growth and urban growth in the ten countries that had more than 10 million slum households in 1990. The countries are ranked by the percentage growth in their slum population from 1990 to 2007, shown by the bars. The dotted line shows for each country in the sequence the percentage increase in GDP per capita from 1990–2007 *divided by 10* (to fit on the same scale).

growth, and the prevalence of slums? Is there evidence that standards of living are improving within slums, and/or across generations of slum dwellers?

Over the past 20 years, countries that experienced fast economic growth are also the ones that achieved the most significant reductions in the proportion of urban households living in slums. In a cross-country regression framework, Arimah (2010) found that the prevalence of slums in any given country was significantly correlated with a variety of aggregate economic indicators, including GDP per capita (negatively), the debt stock and debt service, and inequality measured by the Gini coefficient (positively). However, cross-country correlations overlook widely heterogeneous situations, as rapid urbanization rates in developing countries are often not associated with fast economic growth. In fact, a number of the least developed countries have experienced a rapid growth of their urban population without experiencing much economic growth at all. Extreme rural poverty, natural disasters, and civil wars have been the main drivers of this "urbanization without growth." An example is given by the Democratic Republic of the Congo, where the population of the country's capital Kinshasa more than tripled in size between the beginning of the Mobutu regime (1965) and the end of the Second Congo War (2002).

Figure 1 compares slum growth and urban growth in the ten countries that had more than 10 million slum households in 1990. The countries are ranked

by the percentage growth in their slum population from 1990 to 2007, shown by the bars. These bars can be compared to overall urban population growth, shown by the dark solid line. A few countries experienced explosive urban growth with limited or no slum expansion (for instance, India, Indonesia, and Brazil), whereas in others (Pakistan and Nigeria), slum growth accounts for most of urban growth. The countries where urban population growth outstripped slum population growth have a declining share of urban population living in slums: between 1990 and 2009, this proportion decreased from 55 to 29 percent in India, from 44 to 29 percent in China, and from 51 to 23 percent in Indonesia (UN-Habitat 2012a).

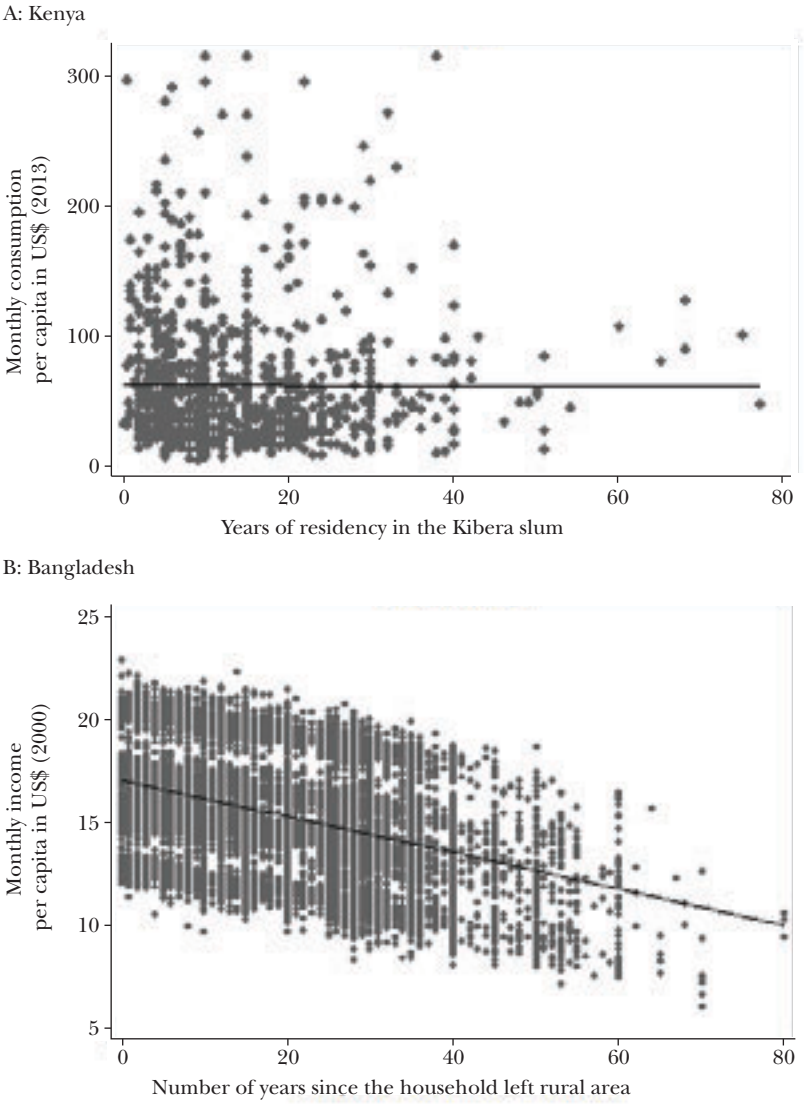
The dotted line in Figure 1 shows for each country in the sequence, the percentage increase in GDP per capita over 1990–2007 *divided by 10* (to fit on the same scale). Very roughly, per capita growth in GDP was similar between the set of countries where slum populations fell (Egypt, Mexico, and Indonesia), and the set of countries where most of urban growth was slum populations (the Philippines, Pakistan, and Nigeria). In fact, it appears that the connection between economic growth and slum growth across countries is quite diverse, without a uniform pattern. Below, we discuss how different policy choices may have contributed to these heterogeneous experiences.

An Intergenerational Perspective

In many slums in low-income countries, living standards do not seem to be improving over time. In Kibera, Kenya, census data suggest that living conditions have either deteriorated or at best stagnated over the 1999–2009 period, a period during which the economy as a whole grew at 3.5 percent per annum. The share of household heads with a primary education fell from 47 to 40 percent over this time; the number of rooms per capita held essentially the same at 0.68 in 1999, and 0.67 in 2009; and the share of those using a pit latrine (rather than the main sewer system) fell only slightly, from 82 percent in 1999 to 77 percent by 2009.

Of course, this simple analysis of living standards over a given period overlooks a fundamental selection problem: households that improved their condition over the period may no longer live in the slum, while other poor households may have migrated into the slum. The available evidence, overall, does not provide *prima facie* evidence of rapid changes in the composition of slums. In our Kibera data, respondent households have lived in the slum for 16 years on average, and incomes do not increase (nor decrease) with the duration of residency, as illustrated in Figure 2 (panel A). In the Bangladeshi settlements studied, income per capita correlates negatively and significantly with the total number of years that that household has spent in the slum and with the number of years since the households first left the countryside (Figure 2, panel B). Similarly, UN-Habitat (2003) reported that 41 percent of Kolkata slum dwellers had lived in slums for over 30 years, and more than 70 percent had lived in slums for over 15 years. In a case study of Bangkok slums, 60 percent of individuals were reportedly born in the same slum (UN-Habitat 2003).

Figure 2
Living Standards and Duration of Residency in Slums of Bangladesh and Kenya



Notes: The first scatterplot shows the result of a single variable regression of monthly consumption per capita on the number of years the household has lived in the slum from our Kibera survey data. The estimated slope is −0.017 (SE 0.15). In this dataset the average monthly consumption is USD 64. The second scatterplot presents the result of a regression of monthly income per capita (defined as the total household income in US$ (non PPP) divided by the household size) on the number of years since the household first left rural areas for Tongi and Jessore slums in Bangladesh, controlling for the age and education of the household head, the number of adults in the household, and district fixed effects. The average monthly consumption is US$15 (US$21 in 2012 equivalent). The slope of the fitted line is the coefficient of interest in that regression, and the estimated value of this coefficient is −0.055 (SE: 0.019). The negative sign of the slope is robust to removing these controls and to the inclusion of more controls.

To our knowledge, Perlman (2010) is one of the few studies to have attempted to address selection issues in surveys of slums by tracing slum respondents and their descendants over an extended period of time: 1969–2001, in the *favelas* of Rio de Janeiro. Her findings were somewhat ambiguous. She reported that a majority of individuals interviewed in 1969 and found again in 2001 were no longer living in the *favelas* (63 percent of the original interviewees, as well as 64 percent of their children and 68 percent of their grandchildren), and that the fraction that had remained in the favelas had better access to public services (including education) as well as improved household amenities. However, only a nonrandom 41 percent of the original sample (307 out of 750) could be accurately relocated, of which 22 percent were still alive and 19 percent were deceased. In addition, from a new random sample of 425 individuals in the same survey areas, she found considerably higher unemployment rates (51 percent, up from 32 percent in 1969) and a higher fraction of respondents with no income (23 percent, up from 17 percent in 1969).

One question that remains is why households that initially migrated to slums but did not experience welfare gains do not return to their province of origin. A dataset on two slums in Nairobi (APHRC 2013)[11] that collects information on where outmigrants go once they leave the slum provides some answers to that question. Between 2003 and 2007, 15 percent of slum residents moved locations. Of these, 26 percent moved to another slum, 4 percent moved to another location in the same slums, 22 percent moved to a nonslum area in Nairobi, and 41 percent moved to rural Kenya. The notion of slums being a poverty trap implies that households do not necessarily move there planning to stay, but instead get caught in a low-level equilibrium. The APHRC data does speak to this—very few slum residents moved, only about 15 percent, and one-third of these stayed in slum areas.

Empirically, little is known about the outside options that slum households have or contemplate at any given time. Although there is a small probability of success (for example, of finding consistent employment), perhaps the expected gains are large enough to make this decision individually rational. A related literature is that on entrepreneurship—many studies find that entrepreneurship does not pay (Hamilton 2000; Moskowitz and Vissing-Jørgensen 2002), but some of this may be due to nonpecuniary benefits of being self-employed (Hamilton 2000) or perhaps to the overconfidence that entrepreneurs have in their own skills (Bernardo and Welch 2001). In the literature on education, researchers also found that information provided to parents on the returns to education can have effects on enrollment rates of their children (Jensen 2010; Nguyen 2008). Issues of overconfidence and misinformation about individual success probabilities may also be at stake in slums—there is room for research on expectations and their role in migration decisions, especially in the context of slum populations.

[11] This data covers two slums, Korogocho and Viwandani, which form a Demographic Surveillance Site in Nairobi. The data is collected by the African Population and Health Research Center. A subset of the data is available at http://www.aphrc.org/.

Limitations of Past Approaches in Slum Policy

Ridding cities of slums may be considered an essential part of the development process; yet this has proved nearly impossible for policymakers in most emerging and developing economies. In a recent testimony of the problems encountered by ambitious slum policy, the government of India announced in 2009 the implementation of the *Rajiv Awas Yojana* (RAY) scheme to make the country "slum-free" within five years. Three months later, the timeframe of the program was extended to seven years (Aggarwal 2009). In 2011, India's Committee on Slum Statistics estimated that the total slum population in the country would still increase by 12 percent between 2011 and 2017 (Government of India 2011). By the end of 2012, the scheme was still in its infancy, and early implementation had been hampered by problems with land records, site selection, and the allocation of land (Kundu 2012). In this section, we discuss what policy approaches towards slums have been taken and why these approaches have been largely unsuccessful.

From "Benign Neglect" to "Aided Self-help"

Historically, outright evictions have been used as a primary policy instrument to reduce slum populations. A highly publicized example of a large-scale slum clearance scheme was "Operation Clean Up" implemented in 2005 in Harare, Zimbabwe, where 700,000 individuals lost their homes (Tibaijuka 2005). While little data has been collected on slum households evicted as part of these policies, it is quite clear that slum clearance does not address the roots of the slum problem. Hence a popular alternative to clearance, widely adopted in the 1960s and 1970s, was the deliberate neglect of expanding slum areas. As part of this approach, there were no policing of squatters, but also no provision of public services in informal settlements. In the "benign" interpretation, policymakers assumed that the market would "take care of it": slums provided much-needed low-cost housing to urban dwellers, who would eventually move into formal housing.

By the 1970s, however, it became clear that neither slum clearance nor "benign neglect" would address the continuous expansion of slums. Slum inhabitants were not being pulled away by improved economic prospects nor being pushed away from the less-desirable aspects of unregulated slum living. A new thinking emerged that the urban poor would find creative solutions to improve their livelihoods as long as basic improvements to the local environment could be performed by the government. Rather than resettling the squatters, governments should focus their efforts on providing basic infrastructure and improving sanitary conditions—like supplying safe water and facilitating waste disposal. Upon completion of these basic improvements, the slum residents would start investing in their own dwellings, and improvements in living standards would follow suit. This "aided self-help" paradigm, inspired by earlier experiences in Europe and the USSR (Harris 1999) and by the influential work of British architect John Turner (Turner and Fichter 1972), convinced the World Bank and other policy planners to reorient their policies towards a "slum upgrading" approach.

Compared to evictions or "benign neglect," slum upgrading seemed at the time to present great advantages. First, it was very cheap: a large upgrading project in Jakarta cost US$38 to US$120 per household, as opposed to the cost of building entirely new housing units for the slum dwellers (Werlin 1999). Second, it would be endorsed by the inhabitants, as local stakeholders and communities would view themselves as receiving an improved standard of living and would be involved in the maintenance of the new infrastructure.

By the early 1980s, slum upgrading had been included in numerous poverty alleviation programs across the developing world. Early evaluation results from upgrading projects conducted in Kolkata, Jakarta, and Manila seemed promising: for instance, mortality caused by waterborne disease was halved among beneficiaries in Kolkata, and investments in home improvements were doubled in Jakarta (Werlin 1999). Upgrading programs also seemed to increase the housing supply and the supply of labor by households (Keare and Parris 1982). However, the early enthusiasm for the upgrading approach began to dwindle by the late 1980s as slum areas continued to expand, and the basic infrastructure improvements appeared to be unsustainable. The maintenance systems of the upgraded public goods collapsed and environmental and health issues were on the rise. In Jakarta, where all slum areas had benefited from the upgrading program, 93 percent of the city's wells were contaminated with feces (Werlin 1999). Overall, even though upgrading programs were rarely submitted to rigorous evaluation,[12] the available evidence suggested that the upgrading projects of an "aided self-help" approach did not seem an adequate policy lever to transform slum conditions in a meaningful way.

The Land Titling Paradigm

The writings of Hernando de Soto (2000), despite some early criticism from the economics profession (Woodruff 2001) were influential in refocusing the slum debate towards issues of land tenure and land rights. De Soto argued that giving the poor property titles over their land would provide collateral for millions of poor urban households across the developing world. In de Soto's argument, property rights were the panacea, and encouraging slum dwellers to invest in improving their homes was still viewed as the Holy Grail of slum policy. The key driver of this encouragement would no longer be the provision of public goods by the government, but rather a reduction in the risk of eviction, resulting from the titling of land occupied by squatters. Thus, home investments would become safer for poor households, and slum households would become able to access credit markets to finance investments to create small businesses and educate their children—the ultimate engines of poverty reduction. Land titling would concomitantly increase the local tax base and enable municipalities to improve the provision of basic public goods.

[12] Slum upgrading programs until now rarely provided a natural experiment framework that could isolate causal effects (Field and Kremer 2006). A recent exception is the work of Galiani et al. (2013), who provide causal estimates of the impact of housing upgrading projects in slums of three Latin American countries.

In the past 20 years, national governments and the World Bank implemented urban land titling projects in more than 18 African, Asian, and Latin American countries (Durand-Lasserve, Fernandes, Payne, and Rakodi 2007).

The positive relationship between tenure security and investments in land has been well documented in the empirical literature for rural settings (for example, Banerjee, Gertler, and Ghatak 2002; Besley 1995; Goldstein and Udry 2008; Hornbeck 2010). However, few academic studies have looked at the impact of urban titling programs on the investment decisions made by households in slums. For urban households in Peru, Field (2005) found that land titling increases the rate of housing renovations (by about two-thirds), and Field (2007) showed that titling increased household labor supply (by 10 to 15 percent) by freeing resources that were previously used to protect household assets. Galiani and Schargrodsky (2010) showed that the allocation of formal land titles in Buenos Aires led to increased investments and education amongst households who benefited from the titling. On the policy side, however, by 2005 concerns began to emerge about how urban land titling programs were being widely promoted without much if any evidence on their effectiveness (Durand-Lasserve, Fernandes, Payne, and Rakodi 2007).

The major limitation of the land titling argument is that it assumes that a lack of formal titles implies weak or nonexistent property rights. However, there is no systematic evidence that land rights are always weakly enforced in slums. As one counterexample, Lanjouw and Levy (2002) have argued that informal rights can effectively substitute for formal titles in slum settlements in Ecuador. In Kibera, where all vacant land was formally reclaimed by the Kenyan government in 1969, the members of one ethnic group still claim land rights based on tenancy permits allocated by the British colonial authorities in the early twentieth century. The fact that most individuals recognized as landlords live outside the slum (Syagga, Mitullah, and Karirah-Gitau 2002) also implies that their informal rights over the renters are strongly enforced.

In a similar vein, analyzing the impact of two titling projects in Senegal and South Africa, Payne, Durand-Lasserve, and Rakodi (2008) found that "residents in most informal settlements in both [Senegal and South Africa] already enjoy *de facto* tenure security." Hence the impact of titling projects on individuals who already enjoy tenure security is, from the start, uncertain: Palmer (1998) pointed out how the effectiveness of land titling programs actually depended on whether they could achieve an increase in the total security that poor households enjoy. Durand-Lasserve, Fernandes, Payne, and Rakodi (2007) reviewed evidence from land titling projects that may have actually *decreased* tenure security (because they allowed for lawfully enforced evictions) in Afghanistan, Cambodia, Egypt, India, and Rwanda. Titling programs alone cannot be expected to lift households out of poverty and to overhaul existing social and economic dynamics within the slum, because existing systems of ownership act to preserve these dynamics. In fact, land titling is more likely to benefit the "slumlords" (whose informal ownership rights are often well-recognized locally) and hurt, at the bottom of the pyramid, the slum renters, either in the form of outright evictions or increased rents in the titled area.

These features have been documented in Senegal, for instance (Payne, Durand-Lasserve, and Rakodi 2008).

Ultimately, individual approaches such as upgrading and formal titling have largely failed to improve livelihoods in slums. Recognizing that the effect of titling projects and conventional land administration systems had been limited, UN-Habitat (2012c) recently advocated a "continuum of land rights" relying on participatory land enumeration and record-keeping to improve tenure security for the urban poor. This would ideally be included in a more holistic approach of slum policy. Countries that managed to curb the growth of slums, such as Brazil or Egypt, indeed appear to be those where slum policy relied on a combination of instruments—including efforts to increase the transparency and efficiency of land markets, to improve local governance, to increase public investments massively, and to increase the supply of cheap housing (UN-Habitat 2010a).

Conclusion

Slums represent a major policy challenge for developing economies in the twenty-first century. Adding migrants to the existing slum populations, UN-Habitat (2012b) estimated that 450 million new housing units would be needed in the next 20 years just to accommodate households in urgent need of housing. Yet the challenge of slums is not simply one of housing policy: a holistic approach is needed to address housing needs for rural migrants, health and sanitation issues, local governance, private savings and investments, and land market institutions. Both formal and informal systems of property rights may be necessary to curb the rapid growth of slum areas worldwide. In the absence of strong policy agendas similar to those adopted in Singapore or, more recently, in Brazil, it seems unlikely that slums will disappear in the foreseeable future, as implicitly assumed by a modernization view of the issue.

Overall, there has been very little theoretical and empirical economic research about how the public policy challenges posed by slums in low-income economies should be addressed. A research agenda on slums could focus on three distinct sets of methodological and policy questions. First, the methodological problems that hamper field research in slums should be addressed. In particular, efforts to enumerate slum populations and to track panel respondents over several generations of slum dwellers could be stepped up, and empirical methods used to deal with survey attrition in other contexts (Fitzgerald, Gottschalk, and Moffitt 1998) should be more consistently applied. This would allow for a better understanding of the most pressing issues faced by slum dwellers and a better integration of these dwellers in national political processes. Relatedly, there has been little effort to systematically study the movements into and out of slums, or to collect data to track the individuals who do move, even if those exiting are few and far between. Similarly, understanding the intergenerational correlation in incomes and other socioeconomic outcomes would be an important contribution.

Second, the returns to upgrading different types of public services should be identified and quantified, so that cost-effective projects and programs that provide discernable welfare gains for slum residents can be more consistently applied. The Abdul Lutif Jameel Poverty Action Lab (J-PAL) Urban Services Initiative (USI) has begun to promote such an agenda (Duflo, Galiani, and Mobarak 2012).

Third, given the stability in slums documented above and the notion that slums may be poverty traps of some sort, perhaps one policy direction would be the "big push." Several programs with mass investments or wholesale relocation of slum households into housing estates appear to have been successful. However, in many current slums in the developing world, this cannot be done without a political willingness to change governance dynamics in slum areas or, in particular, to deal with the actors who have taken over the governance of these areas in the absence of the government. The governance issues may perhaps be the most pressing since these informal actors have a strong presence in the slums and have large rents at stake should there be any change. They have strong incentives to maintain the status quo. Without changes to these institutions and a reversal of the lack of governance, it is unlikely that any attempts at any form of big push or coordinated investment will have the desired effects.

■ *The authors would like to thank David Autor, Chang-Tai Hsieh, Ulrike Malmendier, and Timothy Taylor for comments that greatly improved the paper. We thank the MIT Sloan School of Management for funding the Kibera surveys used in this paper. In addition, the authors acknowledge use of publicly available data for Bangladesh from the Supporting Household Activities for Health, Assets and Revenue Surveys; for India, from the National Family Health Survey (NFSH-3); and for Hyderabad, from the data repository of the Abdul Latif Jameel Poverty Action Lab (J-PAL). The authors are extremely grateful to the Kenya National Bureau of Statistics for access to their 1999 and 2009 Population Censuses and UNICEF for access to the 2010 UNICEF SMART Survey for Sierra Leone.*

References

Aggarwal, Smita. 2009. "Slum-free India Deadline Now 7 Years." *The Indian Express*, August 28. http://www.indianexpress.com/news/slumfree-india-deadline-now-7-yrs/508138/.

Alonso, William. 1964. *Location and Land Use: Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

Amis, Philip. 1984. "Squatters or Tenants: The Commercialization of Unauthorized Housing in Nairobi." *World Development* 12(1): 87–96.

APHRC. 2013. NUHDSS (Nairobi Urban Health and Demographic Surveillance Site) dataset. http://www.aphrc.org/ (accessed February 15, 2013).

Arimah, Ben C. 2010. "The Face of Urban Poverty. Explaining the Prevalence of Slums in Developing Countries." UNU-WIDER Working Paper no. 2010/30.

Azariadis, Costas, and Allan Drazen. 1990. "Threshold Externalities in Economic Development." *Quarterly Journal of Economics* 105(2): 501–26.

Babalola, Yemi. "Makoko Residents and Their Unwanted Guest." *Africa News Service*, May 1, 2009.

Banerjee, Abhijit V. and Esther Duflo. 2007. "The Economic Lives of the Poor." *Journal of Economic Perspectives* 21(1): 141–68.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2010. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." BREAD Working Paper 278.

Banerjee, Abhijit, Paul Gertler, and Maitreesh Ghatak. 2002. "Empowerment and Efficiency: Tenancy Reform in West Bengal." *Journal of Political Economy* 110(2): 239–80.

Banerjee, Abhijit, Rohini Pande, Yashas Vaidya, Michael Walton, and Jeff Weaver. 2011. "Delhi's Slum-Dwellers: Deprivation, Preferences and Political Engagement among the Urban Poor." International Growth Centre Conference proceedings (Growth Week 2011).

Basu, Kaushik. 2003. *Analytical Development Economics: The Less Developed Economy Revisited*. Cambridge, MA: MIT Press.

Bates, Robert H. 1981. *Markets and States in Tropical Africa: The Political Basis of Agricultural Policies*. Berkeley, CA: University of California Press.

Bernardo, Antonio E., and Ivo Welch. 2001. "On the Evolution of Overconfidence and Entrepreneurs." *Journal of Economics & Management Strategy* 10(3): 301–330.

Besley, Timothy. 1995. "Property Rights and Investment Incentives: Theory and Evidence from Ghana." *Journal of Political Economy* 103(5): 903–937.

Bleakley, Hoyt. 2010. "Health, Human Capital, and Development." *Annual Review of Economics* 2: 283–310.

Bowles, Samuel, Steven N. Durlauf, and Karla Hoff, eds. 2006. *Poverty Traps*. Princeton University Press.

Bradley, David, Carolyn Stephens, Trudy Harpham, and Sandy Cairncross. 1992. "A Review of the Environmental Health Impacts in Developing Country Cities." World Bank Urban Management Program Discussion Paper 11101.

Chowdhury, Shyamal, Ahmed Mushfiq Mobarak, and Gharad Bryan. 2009. "Migrating Away from a Seasonal Famine: A Randomized Intervention in Bangladesh." Human Development Research Paper 2009/41. http://hdr.undp.org/en/reports/global/hdr2009/papers/HDRP_2009_41.pdf

Currie, Janet, and Tom Vogl. 2013. "Early-Life Health and Adult Circumstance in Developing Countries." *Annual Review of Economics* 5: 1–36.

Davis, Mike. 2006. *Planet of Slums*. Verso.

de Soto, Hernando. 2000. *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Basic Books.

Duflo, Esther, Sebastian Galiani, and Mushfiq Mobarak. 2012. "Improving Access to Urban Services for the Poor: Open Issues and a Framework for Future Research Agenda." J-PAL Urban Services Review Paper. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab. http://www.povertyactionlab.org/publication/improving-access-urban-services-poor.

Durand-Lasserve, Alain, Edesio Fernandes, Geoffrey Payne, and Carole Rakodi. 2007. "Social and Economic Impacts of Land Titling Programmes in Urban and Peri-urban Areas: A Review of the Literature." Paper presented at the World Bank Urban Research Symposium, Washington DC, May 2007. http://www.bhamlive3.bham.ac.uk/Documents/college-social-sciences/government-society/idd/research/social-economic-impacts/social-economic-impacts-literature-review.pdf.

El-Zanaty, Fatma, and Ann A. Way. 2004. *Greater Cairo Slums: A Profile Based on the 2003 Egypt Demographic and Health Survey*. http://www.measuredhs.com/pubs/pdf/FA43/FA43.pdf.

Ferraz, Claudio, and Bruno Ottoni. 2013. "State Presence and Urban Violence: Evidence from the Pacification of Rio's Favelas." Unpublished paper.

Field, Erica. 2005. "Property Rights and Investment in Urban Slums." *Journal of the European Economic Association* 3(2–3): 279–90.

Field, Erica. 2007. "Entitled to Work: Urban Tenure Security and Labor Supply in Peru." *Quarterly Journal of Economics* 122(4): 1561–1602.

Field, Erica, and Michael Kremer. 2006. "Impact Evaluation for Slum Upgrading Interventions." World Bank, Doing Impact Evaluation Series No. 3.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources* 33(2): 251–99.

Fox, Sean. 2013. "The Political Economy of Slums: Theory and Evidence from Sub-Saharan Africa." LSE International Development Working Paper No. 13.146. http://www.lse.ac.uk/internationalDevelopment/pdf/WP/WP146.pdf.

Frankenhoff, Charles A. 1967. "Elements of an Economic Model for Slums in a Developing Economy." *Economic Development and Cultural Change* 16(1): 27–36.

Galiani, Sebastian, Paul Gertler, Ryan Cooper, Sebastian Martinez, Adam Ross, and Raimundo Undurraga. 2013. "Shelter from the Storm: Upgrading Housing Infrastructure in Latin American Slums." NBER Working Paper 19322.

Galiani, Sebastian, and Ernesto Schargrodsky. 2010. "Property Rights for the Poor: Effects of Land Titling." *Journal of Public Economics* 94(9–10): 700–729.

Glaeser, Edward. 2011. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier.* New York, NY: Penguin Press.

Goldstein, Markus, and Christopher Udry. 2008. "The Profits of Power: Land Rights and Agricultural Investment in Ghana." *Journal of Political Economy* 116(6): 981–1022.

Golgher, André Braz. 2012. "The Selectivity of Migration and Poverty Traps in Rural Brazil." *Population Review* 51(1): 9–27.

Government of India, Ministry of Housing and Urban Poverty Alleviation. 2011. "Report of the Committee on Slum Statistics/Census." http://mhupa.gov.in/W_new/Slum_Report_NBO.pdf.

Gulyani, Sumila, Ellen M. Bassett, and Debabrata Talukdar. 2012. "Living Conditions, Rents, and Their Determinants in the Slums of Nairobi and Dakar." *Land Economics* 88(2): 251–74.

Gupta, Kamla, Fred Arnold, and H. Lhungdim. 2009. *Health and Living Conditions in Eight Indian Cities.* National Family Health Survey (NFHS-3), India, 2005–06 Mumbai: International Institute for Population Sciences. http://pdf.usaid.gov/pdf_docs/PNADQ634.pdf.

Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *Journal of Political Economy* 108(3): 604–631.

Hannum, Emily, Meiyan Wang, and Jennifer Adams. 2008. "Urban–Rural Disparities in Access to Primary and Secondary Education Under Market Reform." In *One Country, Two Societies? Rural-Urban Inequality in Contemporary China*, edited by Martin King Whyte. Cambridge, MA: Harvard University Press.

Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review* 60(1): 126–142.

Harris, Richard. 1999. "Slipping through the Cracks: The Origins of Aided Self-help Housing, 1918–53." *Housing Studies* 14(3): 281–309.

Hornbeck, Richard. 2010. "Barbed Wire: Property Rights and Agricultural Development." *Quarterly Journal of Economics* 125(2): 767–810.

Jack, William, and Tavneet Suri. Forthcoming. "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution." *American Economic Review.*

Jalan, Jyotsna, and Martin Ravallion. 2002. "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China." *Journal of Applied Econometrics* 17(4): 329–46.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125(2): 515–48.

Joireman, Sandra F. 2011. *Where There is No Government: Enforcing Property Rights in Common Law Africa.* New York, NY: Oxford University Press.

Keare, Douglas H., and Scott Parris. 1982. "Evaluation of Shelter Programs for the Urban Poor: Principal Findings." World Bank Staff Working Papers 547.

Kundu, Amitabh. 2012. "The Challenges of Making India Slum-free." India at LSE blog. http://blogs.lse.ac.uk/indiaatlse/2012/10/24/the-challenges-of-making-indian-cities-slum-free-part-1/.

Lanjouw, Jean O., and Philip I. Levy. 2002. "Untitled: A Study of Formal and Informal Property Rights in Urban Ecuador." *Economic Journal* 112(482): 986–1019.

Lipton, Michael. 1977. *Why Poor People Stay Poor: Urban Bias in World Development.* Cambridge, MA: Harvard University Press.

Lopez, Nestor. 2007. "Urban and Rural Disparities in Latin America: Their Implications for Education Access." Background paper prepared for the Education for All Global Monitoring Report 2008 titled *Education for All by 2015: Will We Make It?* http://unesdoc.unesco.org/images/0015/001555/155540e.pdf.

López-Casasnovas, Guillem, Berta Rivera, and Luis Currais, eds. 2005. *Health and Economic Growth: Findings and Policy Implications.* Cambridge, MA: MIT Press.

Marx, Benjamin, Thomas M. Stoker, and Tavneet Suri. 2013. "There is No Free House: Ethnic Patronage and Property Rights in a Kenyan Slum." http://www.mit.edu/~tavneet/Marx_Stoker_Suri.pdf.

Matsuyama, Kiminori. 2005. "Poverty Traps." In *The New Palgrave Dictionary of Economics*, 2nd Edition, edited by L. Blume and S. Durlauf. Palgrave Macmillan.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1): 159–217.

Mills, Edwin S. 1972. *Studies in the Structure of the Urban Economy.* Baltimore, MD: Johns Hopkins Press.

Moskowitz, Tobias, and Annette Vissing-Jørgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92(4): 745–78.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97(5): 1003–1026.

Muth, Richard F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use.* University of Chicago Press.

Nguyen, Trang. 2008. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." January 23. http://www.povertyactionlab.org/sites/default/files/documents/Nguyen%202008.pdf .

Palmer, David. 1998. "Security, Risk and Registration." *Land Use Policy* 15(1): 83–94.

Payne, Geoffrey, Alain Durand-Lasserve, and Carole Rakodi. 2008. "Social and Economic Impacts of Land Titling Programmes in Urban and Peri-urban Areas: International Experience and Case Studies of Senegal and South Africa: Final Report." http://www.unrol.org/files/synthesis_report.pdf.

Perlman, Janice. 2010. *Favela: Four Decades of Living on the Edge of Rio de Janeiro.* New York, NY: Oxford University Press.

Sabry, Sarah. 2010. "How Poverty is Underestimated in Greater Cairo, Egypt." *Environment and Urbanization* 22(2): 523–41.

Syagga, Paul, Winnie Mitullah, and Sarah Karirah-Gitau. 2002. "Nairobi Situation Analysis Supplementary Study: A Rapid Economic Appraisal of Rents in Slums and Informal Settlements" (draft). Contribution to the Preparatory Phase (January–November 2002) of the Government of Kenya & UN-HABITAT Collaborative Nairobi Slum Upgrading Initiative.

Tibaijuka, Anna Kajumulo. 2005. "Report of the Fact-Finding Mission to Zimbabwe to Assess the Scope and Impact of Operation Murambatsvina by the UN Special Envoy on Human Settlements Issues in Zimbabwe Mrs. Anna Kajumulo Tibaijuka." United Nations. http://ww2.unhabitat.org/documents/ZimbabweReport.pdf.

Todaro, Michael P. 1976. "Urban Job Expansion, Induced Migration, and Rising Unemployment: A Formulation and Simplified Empirical Test for LDCs." *Journal of Development Economics* 3(3): 211–25.

Turner, John. 1969. "Uncontrolled Urban Settlement: Problems and Policies." In *The City in Newly Developing Countries: Readings on Urbanism and Urbanization,* edited by Gerald William Breese, 507–534. Englewood Cliffs, NJ: Prentice Hall.

Turner, John F. C., and Robert Fichter, eds. 1972. *Freedom to Build: Dweller Control of the Housing Process.* Collier Macmillan Ltd.

UN-Habitat. 2003. *The Challenge of Slums: Global Report on Human Settlements 2003.* London and Sterling, VA: Earthscan Publications.

UN-Habitat. 2006. *State of the World's Cities 2006/2007.* Nairobi: United Nations Human Settlements Programme.

UN-Habitat. 2010a. *State of the World's Cities 2010/2011.* Nairobi: United Nations Human Settlements Programme.

UN-Habitat. 2010b. *The Challenge of Slums: Global Report on Human Settlements 2003.* Revised and updated version (April 2010). http://www.unhabitat.org/downloads/docs/GRHS_2003_Chapter_01_Revised_2010.pdf.

UN-Habitat. 2012a. *State of the World's Cities 2012/2013: Prosperity of Cities.* Nairobi: United Nations Human Settlements Programme.

UN-Habitat. 2012b. *Sustainable Housing for Sustainable Cities: A Policy Framework for Developing Countries.* Nairobi: United Nations Human Settlements Programme.

UN-Habitat. 2012c. "Land Records for the Poor." Brief 1/2012. Nairobi: United Nations Human Settlements Programme.

UN-Habitat. 2013. Global Urban Indicators database. http://www.unhabitat.org/stats/Default.aspx (accessed February 15, 2013).

US Census Bureau. 2013. "New York County (Manhattan Borough), New York." State & County QuickFacts. http://quickfacts.census.gov/qfd/states/36/36061.html.

Werlin, Herbert. 1999. "The Slum Upgrading Myth." *Urban Studies* 36(9): 1523–34.

World Bank. 2009. *The World Development Report 2009: Reshaping Economic Geography.* Washington DC: The World Bank.

World Bank. 2013. "World Development Indicators." http://data.worldbank.org/data-catalog/world-development-indicators (accessed February 15, 2013).

World Health Organization. 2013. "Global Database on Child Growth and Malnutrition." http://www.who.int/nutgrowthdb/en/ (accessed February 15, 2013).

Woodruff, Christopher. 2001. "Review of de Soto's *The Mystery of Capital*." *Journal of Economic Literature* 39(4): 1215–23.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 211–218

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.

International Trade

The *World Investment Report 2013* from UNCTAD (the UN Conference on Trade and Development) has the theme "Global Value Chains: Investment and Trade for Development." "About 60 per cent of global trade, which today amounts to more than $20 trillion, consists of trade in intermediate goods and services that are incorporated at various stages in the production process of goods and services for final consumption." "For instance, even the relatively simple GVC [global value chain] of Starbuck's (United States), based on one service (the sale of coffee), requires the management of a value chain that spans all continents; directly employs 150,000 people; sources coffee from thousands of traders, agents and contract farmers across the developing world; manufactures coffee in over 30 plants, mostly in alliance with partner firms,

■ *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.27.4.211. doi=10.1257/jep.27.4.211

usually close to final market; distributes the coffee to retail outlets through over 50 major central and regional warehouses and distribution centres; and operates some 17,000 retail stores in over 50 countries across the globe. This GVC has to be efficient and profitable, while following strict product/service standards for quality. It is supported by a large array of services, including those connected to supply chain management and human resources management/development, both within the firm itself and in relation to suppliers and other partners. The trade flows involved are immense, including the movement of agricultural goods, manufactured produce, and technical and managerial services." At http://unctad.org/en/PublicationsLibrary/wir2013_en.pdf.

The *World Trade Report 2013* from the World Trade Organization has the theme "Factors Shaping the Future of World Trade." "Measured in gross terms, the dollar value of world merchandise trade increased by more than 7 per cent per year on average between 1980 and 2011, reaching a peak of US$ 18 trillion at the end of that period. Trade in commercial services grew even faster, at roughly 8 per cent per year on average, amounting to some US$ 4 trillion in 2011. . . . Developing economies only accounted for 34 per cent of world exports in 1980 but by 2011 their share had risen to 47 per cent, or nearly half of the total. . . . Improvements in transport, telecommunications and information technology, together with increased economic integration and greater trade openness, have resulted in higher levels of technological diffusion and increased mobility and accumulation of productive factors over time. As a result, countries have become less specialized in the export of particular products, and therefore more similar in terms of their export composition. Comparative advantage, or international differences in relative efficiencies among products, has become weaker over time in many countries." At http://www.wto.org/english/res_e/publications_e/wtr13_e.htm.

Arvind Subramanian and Martin Kessler discuss "The Hyperglobalization of Trade and Its Future." "This paper describes seven salient features of trade integration in the 21st century: Trade integration has been more rapid than ever (hyperglobalization); it is dematerialized, with the growing importance of services trade; it is democratic, because openness has been embraced widely; it is criss-crossing because similar goods and investment flows now go from South to North as well as the reverse; it has witnessed the emergence of a mega-trader (China), the first since Imperial Britain; it has involved the proliferation of regional and preferential trade agreements and is on the cusp of mega-regionalism as the world's largest traders pursue such agreements with each other; and it is impeded by the continued existence of high barriers to trade in services." Peterson Institute for International Economics, July 2013, Working Paper 13-6, http://www.iie.com/publications/wp/wp13-6.pdf.

The Bank of International Settlements has published the "Triennial Central Bank Survey," subtitled "Foreign Exchange Turnover in April 2013: Preliminary Global Results." "Trading in foreign exchange markets averaged $5.3 trillion per day in April 2013. This is up from $4.0 trillion in April 2010 and $3.3 trillion in April 2007. . . . The US dollar remained the dominant vehicle currency; it was on one side

of 87% of all trades in April 2013. The euro was the second most traded currency, but its share fell to 33% in April 2013 from 39% in April 2010. The turnover of the Japanese yen increased significantly between the 2010 and 2013 surveys. So too did that of several emerging market currencies, and the Mexican peso and Chinese renminbi entered the list of the top 10 most traded currencies." September 2013, http://www.bis.org/publ/rpfx13fx.pdf.

Lecture Notes

Alan Krueger spoke at the Rock and Roll Hall of Fame on "Rock and Roll, Economics, and Rebuilding the Middle Class." He used the music industry as a microcosm for technological trends that have led to greater income inequality in recent decades. "I want to highlight four factors that are important in generating a superstar economy. These are technology, scale, luck and an erosion of social norms that compress prices and incomes. All of these factors are affecting the music industry. . . . Technological changes through the centuries have long made the music industry a super star industry. Advances over time including amplification, radio, records, 8-tracks, music videos, CDs, iPods, etc., have made it possible for the best performers to reach an ever wider audience with high fidelity. And the increasing globalization of the world economy has vastly increased the reach and notoriety of the most popular performers. They literally can be heard on a worldwide stage. But advances in technology have also had an unexpected effect. Recorded music has become cheap to replicate and distribute, and it is difficult to police unauthorized reproductions. This has cut into the revenue stream of the best performers, and caused them to raise their prices for live performances. My research suggests that this is the primary reason why concert prices have risen so much since the late 1990s. In this spirit, David Bowie once predicted that 'music itself is going to become like running water or electricity,' and, that as a result, artists should 'be prepared for doing a lot of touring because that's really the only unique situation that's going to be left.' While concerts used to be a loss leader to sell albums, today concerts are a profit center.' June 12, 2013, at http://www.whitehouse.gov/blog/2013/06/12/rock-and-roll-economics-and-rebuilding-middle-class/.

Raghuram Rajan delivered the Andrew Crockett Memorial Lecture at the Bank of International Settlements, titled "A Step in the Dark: Unconventional Monetary Policy after the Crisis." "Two competing narratives of the sources of the crisis, and attendant remedies, are emerging. The first, and the better known diagnosis, is that demand has collapsed because of the high debt build up prior to the crisis. . . . But there is another narrative. And that is that the fundamental growth capacity in industrial countries has been shifting down for decades now, masked for a while by debt-fuelled demand. More such demand, or asking for reckless spending from emerging markets, will not put us back on a sustainable path to growth. Instead, industrial democracies need to improve the environment for growth. The first narrative is the standard Keynesian one, modified for a debt crisis. It is the one

most government officials and central bankers, as well as Wall Street economists, subscribe to, and needs little elaboration. The second narrative, in my view, offers a deeper and more persuasive view of the blight that afflicts our times." Rajan argues that central banks took the right actions during the financial crisis, but that the wisdom of the ultra-low interest rate policies in the aftermath of the crisis are not yet clear. "Churchill could well have said on the subject of unconventional monetary policy, 'Never in the field of economic policy has so much been spent, with so little evidence, by so few'. Unconventional monetary policy has truly been a step in the dark." June 23, 2013, at http://www.bis.org/events/agm2013/sp130623.htm.

Potpourri

An IMF staff team led by Bernardin Akitoby offers some thoughts on "Reassessing the Role and Modalities of Fiscal Policy in Advanced Economies." "The prevailing consensus before the crisis was that discretionary fiscal policy had a limited role to play in fighting recessions. The focus of fiscal policy in advanced economies was often on the achievement of medium- to long-run goals such as raising national saving, external rebalancing, and maintaining long-run fiscal and debt sustainability given looming demographic spending pressures. For the management of business cycle fluctuations, monetary policy was seen as the central macroeconomic policy tool. . . . While debate continues, the evidence seems stronger than before the crisis that fiscal policy can, under today's special circumstances, have powerful effects on the economy in the short run. In particular, there is even stronger evidence than before that fiscal multipliers are larger when monetary policy is constrained by the zero lower bound (ZLB) on nominal interest rates, the financial sector is weak, or the economy is in a slump. Earlier research often assumed that the impact of fiscal policy was similar across different states of the economy, but a number of recent empirical studies suggest that fiscal multipliers may be larger during periods of slack." September 2013, IMF Policy Paper at http://www.imf.org/external/np/pp/eng/2013/072113.pdf.

The Congressional Budget Office discusses "Rising Demand for Long-Term Services and Supports for Elderly People." "By 2050, one-fifth of the total U.S. population will be elderly (that is, 65 or older), up from 12 percent in 2000 and 8 percent in 1950. The number of people age 85 or older will grow the fastest over the next few decades, constituting 4 percent of the population by 2050, or 10 times its share in 1950. That growth in the elderly population will bring a corresponding surge in the number of elderly people with functional and cognitive limitations. . . . On average, about one-third of people age 65 or older report functional limitations of one kind or another; among people age 85 or older, about two-thirds report functional limitations. . . . The total value of long-term services and supports for elderly people, including the estimated economic value of informal (or donated) care, exceeded $400 billion in 2011 . . ." June 26, 2013, at http://www.cbo.gov/publication/44363.

Frank Levy and Richard J. Murnane consider the interaction between workers and machinery in "Dancing with Robots: Human Skills for Computerized Work." "On March 22, 1964, President Lyndon Johnson received a short, alarming memorandum from the Ad Hoc Committee on the Triple Revolution. The memo warned the president of threats to the nation beginning with the likelihood that computers would soon create mass unemployment: 'A new era of production has begun. Its principles of organization are as different from those of the industrial era as those of the industrial era were different from the agricultural. The cybernation revolution has been brought about by the combination of the computer and the automated self-regulating machine. This results in a system of almost unlimited productive capacity which requires progressively less human labor. Cybernation is already reorganizing the economic and social system to meet its own needs.' The memo was signed by luminaries including Nobel Prize winning chemist Linus Pauling, *Scientific American* publisher Gerard Piel, and economist Gunnar Myrdal (a future Nobel Prize winner). Nonetheless, its warning was only half right. There was no mass unemployment—since 1964 the economy has added 74 million jobs. But computers have changed the jobs that are available, the skills those jobs require, and the wages the jobs pay. For the foreseeable future, the challenge of "cybernation" is not mass unemployment but the need to educate many more young people for the jobs computers cannot do." Third Way, 2013, http://content.thirdway.org/publications/714/Dancing-With-Robots.pdf.

C. Robert Taylor and Diana L. Moss have written "The Fertilizer Oligopoly: The Case for Antitrust Enforcement," as a monograph for the American Antitrust Institute. "Collusive agreements between fertilizer producers on prices and market shares pepper the history of the global commercial fertilizer industry dating back to the 1880s. The underlying structure of the current global industry remains conducive to anticompetitive coordination—a landscape that undoubtedly prompted Wall Street Journal commentators to observe that fertilizer markets are so manipulated, 'they might make a Saudi prince blush,' and 'the global price sets a benchmark so American farmers pay essentially what the cartels dictate.' Indeed, the global industry is dominated by two government-sanctioned export associations in the U.S. (*PhosChem*) and Canada (*Canpotex*); a privately traded monopoly sanctioned and likely controlled by the Moroccan government (Office Chérifien des Phosphates (OCP)); and a cabal of three potash companies in the former Soviet Union (Belaruskali, Silvinit, and Uralkali, operating through their marketing cartel, Belarusian Potash Company (BPC)). . . . Damages from supra-competitive pricing of fertilizer likely amount to tens of billions of dollars annually, the direct effects of which are felt by farmers and ranchers. But consumers all over the world suffer indirectly from cartelization of the fertilizer industry through higher food prices, particularly low-income and subsistence demographics." American Antitrust Institute, 2013, at http://www.antitrustinstitute.org/~antitrust/sites/default/files/FertilizerMonograph.pdf.

C*ato Unbound* offers four essays on "The Political Economy of Recycling." In the lead essay, Michael Munger asks: "Recycling: Can It Be Wrong, When It

Feels So Right?" "There are two general kinds of arguments in favor of recycling. The first is that 'this stuff is too valuable to throw away!' In almost all cases, this argument is false, and when it is correct recycling will be voluntary; very little state action is necessary. The second is that recycling is cheaper than landfilling the waste. This argument may well be correct, but it is difficult to judge because officials need keep landfill prices artificially low to discourage illegal dumping and burning. Empirically, recycling is almost always substantially more expensive than disposing in the landfill. Since we can't use the price system, authorities resort to moralistic claims, trying to persuade people that recycling is just something that good citizens do. But if recycling is a moral imperative, and the goal is zero waste, not optimal waste, the result can be a net waste of the very resources that recycling was implemented to conserve." There are sharp and lively comments from Edward Humes, Melissa Walsh Innes, and Stephen Landsberg. June 2013, at http://www.cato-unbound.org/issues/june-2013/political-economy-recycling.

Laurie Simon Hodrick asks, "Are U.S. Firms Really Holding Too Much Cash?" "Five companies, General Electric, Microsoft, Google, Cisco, and Apple, account for 25 percent of the $1.27 trillion [in cash that US companies are holding], while 22 companies account for half of it. . . . [C]ash holdings are concentrated among highly profitable firms, many in the technology and health care sectors." "It is also important to recognize that the 'overseas' money owned by foreign subsidiaries need not be invested abroad, but instead can be held at U.S. banks, in U.S. dollars, or invested in U.S. securities. For example, according to SEC filings, $58 billion of Microsoft's total cash holding of $66.6 billion is held by foreign subsidiaries. Surprisingly, about 93 percent of Microsoft's cash held by foreign subsidiaries in 2012 was invested in U.S. government bonds, corporate bonds, and mortgage-based securities. The assets of Apple Operations International, Apple's Irish subsidiary, are managed in Reno, Nevada, by employees at one of its wholly owned subsidiaries, Braeburn Capital, according to a Senate report, with the funds held in bank accounts in New York." Stanford Institute for Economic Policy Research Policy Brief, July 2013. http://siepr.stanford.edu/?q=/system/files/shared/pubs/papers/briefs/PolicyBrief07_2013_1_v32.pdf.

Robert S. Goldfarb helps to stock the toolkit of teachers with "Shortage, Shortage, Who's Got the Shortage?" Goldfarb provides six categories of shortages, with multiple examples: 1) A Demand Deadline Enables a Short-Run Shortage (for example, popular Christmas toys); 2) Dynamic Shortages due to Lags in Supply or Demand (lag in number of trained nurses); 3) Market Prices Set by Suppliers Below Market-Clearing Levels (tickets to popular performances or athletic events); 4) Prices Set or Regulated by Government (and/or Quantity Regulation) (price of parking or toll roads); 5) Capacity Choice in the Face of "Regular" Variance in Demand (airlines and hospitals); and 6) Sudden Unexpected Supply Shocks (weather and agriculture). *Journal of Economic Education*, 2013, vol. 44, no. 3, pp. 277–97, at http://www.tandfonline.com/doi/full/10.108‰220485.2013.795461#.Ula8HH8lXSJ.

Discussion Starters

Richard V. Reeves, Isabel Sawhill, and Kimberley Howard discuss "The Parenting Gap." "Gaps in cognitive ability by income background open up early in life, according to research by Tamara Halle and her colleagues at Child Trends, a non-profit research center focused on children and youth. Children in families with incomes lower than 200 percent of the federal poverty line score, on average, one-fifth of a standard deviation below higher-income children on the standard Bayley Cognitive Assessment at nine months—but more than half a standard deviation below higher-income peers at two years. This is the social science equivalent of the difference between a gully and a valley. These early months are critical for developing skills in language and reasoning—and, of course, months in which parents play the most important role. Closing ability gaps in the first two years of life—pre-pre-K, if you like—means, by definition, closing the parenting gap. . . . Parents influence their child's fortunes right from their first breath, while pre-K is aimed at 4-year-olds. In child-development terms, four years is an eon. By the time pre-K kicks in, big differentials in test scores are already apparent. . . . In the last five years, the federal government has allocated $37.5 billion to Head Start—25 times as much as promised to home-visiting programs over the next five. This may not be the optimum ratio in terms of promoting greater mobility and opportunity. . . ." *Democracy*, Fall 2013, pp. 40–50. http://www.democracyjournal.org/30/the-parenting-gap.php.

Jacqueline Deslauriers tells of the great maple syrup heist in "Liquid Gold." "The Federation of Quebec Maple Syrup Producers was set up in 1966 to represent and advocate for producers—most of them dairy farmers who supplemented their income by tapping maple trees. By the 1990s, maple syrup output had grown rapidly, and by 2000 the industry was producing a surplus of between 1.3 and 2 million gallons a year. . . . Any output that cannot be sold must be transferred to the federation's reserve. Producers do not receive payment for this excess production until the federation sells it. . . . Maple syrup is sold from the reserve when current production does not meet the demand from authorized buyers. In 2009, after four dismal years of production, the global maple syrup reserve ran dry. Since then production has bounced back and the reserve is overflowing. . . . The $18 million theft was from one of three warehouses the federation uses to stash excess production and was discovered in mid–2012 during an audit of the warehouse contents. The warehouse, about 60 miles southwest of provincial capital Quebec City, was lightly guarded—in retrospect, perhaps, too lightly guarded. The thieves set up shop nearby, and over the course of a year, according to police, made off with roughly 10,000 barrels of maple syrup—about 323,000 gallons, or about 10 percent of the reserve." *Finance & Development*, June 2013, pp. 48–51. http://www.imf.org/external/pubs/ft/fandd/2013/06/deslauriers.htm.

John C. Williams explains that "Cash Is Dead! Long Live Cash!" "[S]ince the start of the recession in December 2007 and throughout the recovery, the value of U.S. currency in circulation has risen dramatically. It is now fully 42% higher than it was five years ago. . . . Over the past five years, cash holdings increased on average

about 7¼% annually, more than three times faster than the economy's growth rate over this period. At the end of 2012, currency in circulation stood at over $1.1 trillion, representing a staggering $3,500 for every man, woman, and child in the nation. . . . "As fears about the safety of the banking system spread in late 2008, many people became terrified of losing their savings. Instead, they put their trust in cold, hard cash. Not surprisingly, as depositors socked away money to protect themselves against a financial collapse, they often sought $100 bills. Such a large denomination is easier to conceal or store in bulk than smaller bills. Indeed, in the six months following the fall of the investment bank Lehman Brothers in 2008, holdings of $100 bills soared by $58 billion, a 10% jump." *2012 Annual Report*, Federal Reserve Bank of San Francisco, http://www.frbsf.org/files/2012_Annual_Report.pdf.

Michael Manville and Jonathan Williams advocate ending free parking placards for the disabled in "Parking without Paying." "The government isn't going to hand out free gasoline anytime soon, but at least 24 states and many local governments do distribute free parking passes, in the form of disabled placards. These placards not only grant access to spaces reserved for people with disabilities, but also let their holders park free, often for unlimited time, at any metered space. Nor are placards difficult to get. In California, for example, doctors, nurses, nurse practitioners, optometrists and chiropractors can all certify people for placards, for everything from serious permanent impairments to temporary conditions like a sprained ankle. We recommend that cities and states limit or eliminate free parking for disabled placards. We believe the payment exemption has high costs and few benefits. It harms both the transportation system and the environment, and offers little help to most people with disabilities." *Access*, Spring 2013, pp. 10–13, http://www.uctc.net/access/42/access42_parkingwoutpaying.shtml.

Ben A. Minteer and Leah R. Gerber propose "Buying Whales to Save Them." "Under this plan, quotas for hunting of whales would be traded in global markets. But again, and unlike most 'catch share' programs in fisheries, the whale conservation market would not restrict participation in the market; both pro- and antiwhaling interests could own and trade quotas. . . . Conservation groups, for example, could choose to buy whale shares in order to protect populations that are currently threatened; they could also buy shares to protect populations that are not presently at risk but that conservationists fear might become threatened in the future." "Despite the widely acknowledged failure of the IWC [International Whaling Commission] moratorium to curtail unsustainable whaling, the whale conservation market idea has proved to be wildly controversial within conservation and antiwhaling circles. . . . Many critics of the idea are also plainly not comfortable with the ethics of putting a price on such iconic species—that is, with using contingent market methods for what they believe should be a categorical ethical obligation to preserve whales. On the other hand . . . the vulnerable status of many whale populations and the failure of the traditional regulatory response to halt unsustainable harvests call for a more innovative and experimental approach to whale policy, including considering unconventional proposals, such as the whale conservation market." *Issues in Science and Technology*, Spring 2013, http://www.issues.org/29.3/minteer.html.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 219–222

Correction: The Composition and Drawdown of Wealth in Retirement

James Poterba, Steven Venti, and David Wise

In extending our research on issues related to those in this paper, we have discovered an error in our calculations of the expected present discounted value of life-contingent payout streams in our paper "The Composition and Drawdown of Wealth in Retirement," published in the Fall 2011 issue of this journal.[1] This error affected entries in Tables 1 and 2 of the paper and the associated discussion. The error overstated the wealth-equivalent value of both Social Security and defined benefit pensions. The corrected values, shown in the tables below, indicate that the average capitalized value of Social Security for all retirement-age households is $204,264 rather than $341,556. The mean capitalized value of defined benefit pension payouts is $99,147. Taken together, Social Security and defined benefit pension wealth represent 34.9 percent of the mean value of household wealth, not 46 percent as initially reported. The share of home equity, which was previously reported as 16.8 percent, correspondingly rises to 20.2 percent and

[1] We are grateful for the research assistance of Ben Sprung-Keyser, whose work on a related project helped us to uncover this error.

■ *James Poterba is Mitsui Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Steven Venti is DeWalt Ankeny Professor of Economic Policy, Dartmouth College, Hanover, New Hampshire. David Wise is John F. Stambaugh Professor of Political Economy, Kennedy School of Government, Harvard University, Cambridge, Massachusetts. Poterba is President, Venti is a Research Associate, and Wise is Area Director for Health and Retirement Programs and Program Director for Aging, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are poterba@mit.edu, Steven.Venti@dartmouth.edu, and david_wise@harvard.edu.*

http://dx.doi.org/10.1257/jep.27.4.219 doi=10.1257/jep.27.4.219

Table 1
Balance Sheets for Households Aged 65–69 in 2008

Asset category	*Percent of households with positive balance*	*Mean holding (dollars)*	*Share of total wealth (percent)*
All households			
Net worth	**99.2**	**870,908**	100.0
Social Security	88.2	**204,264**	**23.5**
Defined benefit pension	42.1	**99,147**	**11.4**
Non-annuitized wealth	90.8	567,496	**65.2**
Financial assets	86.7	132,484	**15.2**
Personal retirement accounts	52.2	121,137	**13.9**
Housing and other real estate	81.3	271,605	**31.2**
Single-person households			
Net worth	**98.8**	**530,556**	100.0
Social Security	86.6	**134,006**	**25.3**
Defined benefit pension	38.0	**62,555**	**11.8**
Non-annuitized wealth	84.4	333,996	**63.0**
Financial assets	82.3	83,082	**15.7**
Personal retirement accounts	36.4	47,074	**8.9**
Housing and other real estate	67.8	188,813	**35.6**
Married couples			
Net worth	**99.5**	**1,148,873**	**100.0**
Social Security	89.6	**261,645**	**22.8**
Defined benefit pension	45.5	**129,033**	**11.2**
Non-annuitized wealth	96.0	758,196	**66.0**
Financial assets	90.3	172,830	**15.0**
Personal retirement accounts	65.1	181,625	**15.8**
Housing and other real estate	92.3	339,222	**29.5**

Source: Authors' tabulations using Health and Retirement Study, Wave 9, 2008. Two components of net worth, business assets (mean value $45,966 for all households) and debt (–$3,697) are included in net worth and non-annuitized wealth, but are not in any of the subcategories (financial assets, personal retirement accounts, or housing and other real estate). The sum of these three subcategories therefore does not equal non-annuitized wealth.
Note: Numbers in bold have changed.

the share of wealth in primary homes, secondary homes, and other real estate is 31.1 percent rather than 25.9 percent as previously reported.

The corrected calculations suggest that one household in five has defined benefit pension wealth of at least $169,400 (in contrast to $238,500 in our earlier analysis), and that a household at the 30th percentile of the Social Security wealth distribution has $126,800 (not $214,500) of such wealth, while a household at the 90th percentile has $384,800 (not $643,100). The median net worth values at the 70th and 90th percentiles of the net worth distribution are $911,900 (not $1.1 million) and $1,826,400 (not $2.1 million), respectively.

Table 2
Distribution of Wealth Components for Households Aged 65–69 in 2008
(in 1,000s)

Percentile	*Net worth*	*Social Security*	*Denfined benefit pension*	*Non-annuitized wealth*	*Financial assets*	*Personal retirement account assets*	*Housing & other real estate*
All households							
10	**127.3**	0.0	0.0	0.1	0.0	0.0	0.0
30	**289.3**	**126.8**	0.0	71.8	2.0	0.0	42.0
50	**548.2**	**187.4**	0.0	221.7	15.0	5.0	120.0
70	**911.9**	**227.8**	**83.0**	518.0	70.0	75.0	229.5
90	**1,826.4**	**384.8**	**329.6**	1,274.0	358.0	347.0	585.0
Single-person households							
10	**97.4**	0.0	0.0	0.0	0.0	0.0	0.0
30	**176.2**	**98.2**	0.0	14.0	0.4	0.0	0.0
50	**295.8**	**136.7**	0.0	100.0	5.0	0.0	60.0
70	**544.1**	**177.2**	**51.2**	272.0	34.0	10.1	150.0
90	**1,094.1**	**230.1**	**206.2**	892.0	240.0	124.0	392.0
Married households							
10	**240.9**	0.0	0.0	24.7	0.0	0.0	12.0
30	**509.2**	**195.3**	0.0	158.0	6.0	0.0	90.0
50	**769.1**	**284.0**	0.0	357.0	27.8	35.0	170.0
70	**1,234.1**	**342.6**	**116.1**	755.7	107.0	137.0	300.0
90	**2,224.2**	**425.5**	**440.4**	1,677.8	459.2	464.0	725.0

Source: Authors' tabulations using 2008 (Wave 9) Health and Retirement Study; see Table 1 and text for further description.
Note: Numbers in bold have changed.

■ *Poterba is a trustee of CREF and of the TIAA-CREF mutual funds; TIAA-CREF is a provider of retirement services and annuity products.*

References

Poterba, James, Steven Venti, and David Wise. 2011. "The Composition and Drawdown of Wealth in Retirement." *Journal of Economic Perspectives* 25(4): 95–118.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 223–226

Correspondence

Patents and the Dissemination of Inventions

In the Winter 2013 issue, Michele Boldrin and David K. Levine ("The Case Against Patents," pp. 3–22) conclude that patents *cannot* promote the dissemination of technology and technical information. They argue (p. 9) that if an inventor were reasonably certain that others would not learn the details of an invention during the period of patent protection, then patenting would be unnecessary because competitors would be unable to make the invention and appropriate any of the market from its inventor. That is, a patent provides only one benefit to the inventor: additional years during which to sell an invention without competition. Similar analysis is frequently employed in the legal and economics literature: for examples, see Bessen and Maskin (2009, especially footnote 31); Ohlin 2005; Kultti, Takalo, and Toikka (2007); and in this journal, Posner (2005, p. 66).

In contrast, we believe that a well-functioning patent system can make a valuable contribution to the general welfare by facilitating rapid dissemination of innovation. The Boldrin–Levine argument assumes, on the contrary, that inventors intend to manufacture and market their own inventions and therefore must keep the details of their inventions out of the hands of competitors. This assumption may be true in some circumstances. However, here are four examples of circumstances in which an inventor can benefit from voluntary disclosure of his invention to others—provided he can be assured that the invention cannot be stolen and copied. Generally these situations have a common theme: The inventor requires assistance to bring his invention to customers, and such collaboration compels disclosure of the invention. After all, much of the research output of universities is commercialized in precisely this way. Indeed, society benefits when tasks like innovation, manufacturing, and marketing are performed by those who are most efficient at doing so—not necessarily the inventor.

1. The inventor needs partners to produce and sell the invention. In general, it is much more efficient for an inventor to partner with a firm that has superior manufacturing and selling capabilities. Even in the earliest stages of negotiation between the inventor and, for instance, a potential manufacturer, the inventor may need to disclose the invention. Confidentiality agreements will not overcome the many pitfalls of this situation, so without a patent, the inventor would risk having his invention copied without compensation. A similar issue arises when an inventor requires financing from investors, such as venture capitalists. Such investors are unlikely to part with their money until the inventor discloses pertinent information about the invention.

2. The inventor intends to license the use of the patent to others. An inventor may lack the necessary capital, infrastructure, or operational expertise to compete in the market for the product based on his invention. For instance, universities routinely invent novel technology, which they then license to industry. Without patent protection, inventors are rightly reluctant to reveal the details of their inventions—for fear that they will be copied by potential licensees.

3. The inventor sells the invention in one market, but others are more efficient at selling in other markets. For example, a firm that develops, manufactures, and sells low-cost personal computers may invent components that also are appropriate for high-end web server computers—a different product, with different customers. Without patent protection, an inventor has no incentive to disclose this invention to others or license it to other firms that could serve other markets. As a result, the other markets will not be served at all—or only will be served much later on.

4. The inventor wishes to exchange technology with a competitor. Technology transfer is a large and growing practice in many high-tech industries. Consider an ordinary laptop, which is constructed and operated with technology incorporating thousands of inventions made by hundreds of different inventors over many years. These inventions are covered by literally thousands of different patents held

by many different proprietors. For instance, an estimated 90,000-plus patents related to microprocessors were held by more than 10,000 parties in 2002 (US Federal Trade Commission 2003, p. 9). Thus, laptops are essentially dense bundles of mutually beneficial innovations that work together to offer consumers a beneficial product or service. Patent protection allows each firm to show the other the specifications of its technology and ultimately exchange the rights to make use of each other's inventions through licensing agreements. This yields products that are superior to any that could be produced entirely by a single firm.

In all of these situations, patents are indispensable because other firms and potential sources of finance cannot usefully discuss an invention without first knowing the details of its production and utilization. A patent reveals the invention while protecting against unauthorized duplication. Thus, patents directly facilitate the dissemination of innovations because they allow inventors to carry out the transfer of information. One extreme and revealing example of the gains from dissemination of an invention is the invention of obstetric forceps, which can save lives during childbirth. Unfortunately, the inventor of this tool kept it secret so he could profit from its use without competition. It remained a family secret for more than 100 years and, consequently, benefited only the small number of women who were patients of that family of doctors (Dunn 1999).

Granted, inefficiencies in the current implementation of the US patent system leave much to be desired and arguably damage the general interest. In particular, current patenting rules offer little incentive for clear and full explanation of the workings and utilization of inventions in the descriptions provided and made public in patent documents. However, new incentives could be incorporated into the patenting process to reward full and comprehensible disclosure. Indeed, if implemented correctly, the patent system also facilitates the voluntary exchange of inventions through the leasing or sale of patent rights and in this way speeds the replacement of obsolete products and technology.

Dean Alderucci
Stern School of Business, New York University

William J. Baumol
Berkley Center for Entrepreneurship and Innovation, Stern School of Business, New York University
Princeton University

References

Bessen, James, and Eric Maskin. 2009. "Sequential Innovation, Patents, and Imitation." *RAND Journal of Economics* 40(4): 611–35.

Boldrin, Michele, and David K. Levine. 2013. "The Case Against Patents." 27(1): 3–22.

"The Disclosure Function of the Patent System (Or Lack Thereof)." 2005. *Harvard Law Review* 118(6): 2007–2028.

Dunn, Peter M. 1999. "The Chamberlen Family (1560–1728) and Obstetric Forceps." *Archives of Disease in Childhood—Fetal and Neonatal Edition* 81(3): F232–F234.

Kultti, Klaus, Tuomas Takalo, and Juuso Toikka. 2007. "Secrecy versus Patenting." *RAND Journal of Economics* 38(1): 22–42.

Posner, Richard A. 2005. "Intellectual Property: The Law and Economics Approach." *Journal of Economic Perspectives* 19(2): 57–73.

US Federal Trade Commission. 2003. *To Promote Innovation: The Proper Balance of Competition and Patent Law and Policy*, October. http://xml.coverpages.org/FTC-InnovationReport2003.pdf.

Response from Michele Boldrin and David K. Levine:

First let us say that nowhere do we conclude nor do we argue "that patents *cannot* promote the dissemination of technology and technical information." We argue that the simple and standard argument according to which patents do promote the dissemination of technology and technical information is incorrect. Our main argument does not revolve around the dissemination of information, but is rather that while patents may in some particular circumstances be useful to promote innovation, even in those cases, the "collateral damages" they produce generate social costs arguably larger than their benefits. With respect to dissemination of information, we showed in Boldrin and Levine (2004) that the situation is complex and that patents may help or hinder the dissemination of information.

We do not have the impression that Alderucci and Baumol disagree with this. Rather their main point is that under some circumstances—when it may be advantageous both individually and socially to involve third parties—patents may promote the dissemination of technology and technical information, and we do not disagree with them about this. However, we would add that, even without patents, information is frequently shared with third parties—in addition to the evidence we mention below, the incentives that give rise to information sharing without patents are discussed, for example, in Anton and Yao (1994), Henry and Ponce (2009), Boldrin and Levine (2010), Ponce (2011), and Bessen and Nuvolari (2013) among others. The models and practical cases discussed in this literature address, as a matter of fact, pretty much all the four cases suggested by Aldarucci and Baumol, showing that their primarily theoretical conclusion is far from obvious.

Ultimately the issue of information dissemination—like every issue involving patents—is empirical and quantitative. Here we would point to the two most well-established pieces of systematic evidence we know of. Kahn and Sokoloff (2001) show that, in the nineteenth century under US patent law, there was widespread sharing of information. Nuvolari (2004) shows, also in the nineteenth century, that in the absence of the use of patents there was widespread sharing of information. Taken together, these two studies suggest that the presence or absence of patents may not be terribly important for information sharing. Or, if we probed the matter more deeply both theoretically and empirically, as have Bessen and Nuvolari (2013), we would find that when new technologies compete with old, patents hinder rather than help information sharing; and Bessen and Nuvolari document this historically in steam engines, power looms, and the steel industry.

Alderucci and Baumol give one anecdote to support the idea that information sharing is encouraged by patents: the invention by the Chamberlen family of obstetric forceps. As the Chamberlens kept the invention secret for roughly a century, exactly how and when the forceps were invented are not known. It is not the case, however, that they were unwilling to sell the secret to a third party: they attempted explicitly to do so in 1670 in France where it does not appear that they held a patent. The reason they were unable to sell the device had nothing to do with patents: that demonstration was a fiasco and the patient died. Moreover, with the passage of the Statute of Monopolies in 1624, the Chamberlens did have recourse to patents in England, where they resided. Nevertheless, they continued to keep the invention secret. The basic history of the Chamberlens as described here appears to be widely agreed upon, and can be found in Dunn (1999) or the Wikipedia article on Peter Chamberlen. This anecdote, then, supports our main point: availability of patents, per se, does not necessarily foster the dissemination of innovations.

There is a broader point about both secrecy and the incentive to innovate that should be understood: regardless of whether a successful invention can be or is patented, a successful invention generally increases the demand for the inventor's services—and this tendency is reinforced if the invention is widely known and used. Another anecdote (told in Loudon 1989) about obstetric forceps in the mid-nineteenth century makes the point: "[A]ny obstetrician worth his salt had to have a pair of forceps to his name." And the more widely used the forceps, the greater the reputation of the obstetrician.

Michele Boldrin
Washington University in St Louis, St. Louis, Missouri
Federal Reserve Bank of St. Louis

David K. Levine
Washington University in St Louis, St. Louis, Missouri
Federal Reserve Bank of St. Louis

References

Anton, James J., and Dennis A. Yao. 1994. "Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights." *American Economic Review* 84(1): 190–209.

Bessen, James, and Allesandro Nuvolari. 2013. "Diffusing New Technology without Dissipating Rents: Some Historical Case Studies of Knowledge Sharing." Unpublished paper, Boston University School of Law.

Boldrin, Michele, and David K. Levine. 2004. "Rent Seeking and Innovation." *Journal of Monetary Economics* 51(1): 127–60.

Boldrin, Michele, and David K. Levine. 2010. *Against Intellectual Monopoly*, Cambridge University Press.

Dunn, Peter M. 1999. "The Chamberlen Family (1560–1728) and Obstetric Forceps." *Archives of Disease in Childhood—Fetal and Neonatal Edition* 81(3): F232–F234.

Henry, Emeric, and Carlos Ponce. 2009. "Waiting to Imitate: On the Dynamic Pricing of Knowledge." CEPR Discussion Paper DP7511. Available at SSRN: http://ssrn.com/abstract=1507507.

Kahn, B. Zorina, and Kenneth L. Sokoloff. 2001. "History Lessons: The Early Development of Intellectual Property Institutions in the United States." *Journal of Economic Perspectives* 15(3): 233–46.

Loudon, Irvine. 1989. Review of "The Obstetric Forceps: A Short History and Descriptive Catalogue of the Forceps in the Museum of the Royal College of Obstetricians and Gynaecologists," by Bryan M. Hibbard. *Medical History* 33(1): 141–42. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1035787/.

Nuvolari, Allesandro. 2004. "Collective Invention during the British Industrial Revolution: The Case of the Cornish Pumping Engine." *Cambridge Journal of Economics* 28(3): 347–63.

Ponce, Carlos J. 2011. "Knowledge Disclosure as Intellectual Property Rights Protection." *Journal of Economic Behavior & Organization* 80(3): 418–34.

To be considered for publication in the Correspondence section, letters should be relatively short—generally less than 1,000 words—and should be sent to the journal offices at jep@jepjournal.org. The editors will choose which letters will be published. All published letters will be subject to editing for style and length.

Journal of Economic Perspectives—Volume 27, Number 4—Fall 2013—Pages 227–228

Notes

For additional announcements, check out the continuously updated JEP online Bulletin Board *(http://www.aeaweb.org/bulletinboard.php). Calls for papers, notices of professional meetings, and other announcements of interest to economists should be submitted to Ann Norman, at jep@jepjournal.org, in one or two paragraphs containing the relevant information. These will be posted at the JEP online Bulletin Board. Given sufficient lead time (at least one month before an issue goes online), we will also print a shorter, one-paragraph version of your notice in the "Notes" section of the* Journal of Economic Perspectives. *We reserve the right to edit material received.*

Call for Sessions and Papers for the **January 2015 American Economic Association Annual Meeting** in Boston, MA. Members wishing to give papers or organize complete sessions for the program for the meetings in Boston are invited to submit proposals electronically to Professor Richard H. Thaler via the American Economic Association website http://www.aeaweb.org/Annual_Meeting/.

To be considered, individual paper proposals (with abstracts) and up to two *Journal of Economic Literature* bibliographic codes in rank order should be submitted by **April 1, 2014**. The deadline for complete session proposals is **April 15, 2014**. *At least one author of each paper must be an AEA member.* All authors of papers on a complete session must join the AEA if the session is selected for the program.

Proposals for complete sessions have historically had a higher probability of inclusion (35–40 percent) than papers submitted individually (10–15 percent). Individual paper contributors are strongly encouraged to use the AEA's EconHarmony website to form integrated sessions. Proposals for a complete session should be submitted only by the session organizer. Sessions normally contain three or four papers.

Please make certain your information is complete before submission. No changes will be accepted until a decision is made about inclusion on the program (usually in July). Econometric studies or highly mathematical papers are not appropriate for sessions sponsored by the AEA; such papers should be submitted to the Econometric Society. Do not send a complete paper. The Association discourages multiple proposals from the same person, and under no circumstances should the same person submit more than two proposals.

Some of the papers presented at the annual meeting are published in the May *American Economic Review* (the *Papers & Proceedings*). The President-elect includes at least three contributed sessions (12 papers) from among those submitted in response to this Call for Sessions and Papers.

The **2014 Annual Meeting** will take place in Philadelphia, PA, on January 3–5, 2014. The headquarters hotel will be the Philadelphia Marriott Downtown. Registration is now open, please go to https://conf.aeaweb.org/.

AEA Continuing Education Program is held immediately after the Annual Meeting in January. The program aims to help mid-career economists and others maintain the value of their human capital. It is tailored primarily to faculty at liberal arts colleges and teaching-oriented state universities who may have fewer research opportunities than colleagues at universities with PhD programs. The lecturers are leading scholars who also are excellent expositors. The focus is on content to help improve teaching and research.

The three topics for January 2014 in San Diego are: Cross Section Econometrics (Alberto Abadie, Harvard University, and Joshua Angrist, MIT), Education and the Economy (Susan Dynarski, University of Michigan, and Brian Jacob, University of Michigan), and Economic Growth (Oded Galor, Brown University, and David Weil, Brown University). For more information go to http://www.aeaweb.org/cont_education/.

Election Results. The American Economic Association is pleased to announce the election results for the 2013 Election for Officers of the American

Economic Association for 2014: Richard H. Thaler (president-elect), David Card and N. Gregory Mankiw (vice-presidents), and Dora L. Costa and Guido W. Imbens (members).

The Fourth Annual AEA Conference on Teaching and Research in Economic Education (CTREE) will be held May 28–30, 2014, at the Washington Marriott at Metro Center. The conference is hosted by the AEA Committee on Economic Education in cooperation with the *Journal of Economic Education*. Plenary talks will be given by Alan Blinder (Princeton), Kenneth G. Elzinga (UVA), Cecilia Rouse (Princeton), and other speakers to be announced. Submissions will be accepted via the AEA online submission system: http://www.aeaweb.org/committees/AEACEE/Conference/2014/submissions/. Submissions may be of individual papers, complete sessions of three or four papers, workshops, or panels. We especially encourage submissions of completed sessions. **The submission deadline is December 1, 2013.** Questions about submissions should be sent to Gail Hoyt (ghoyt@email.uky.edu).

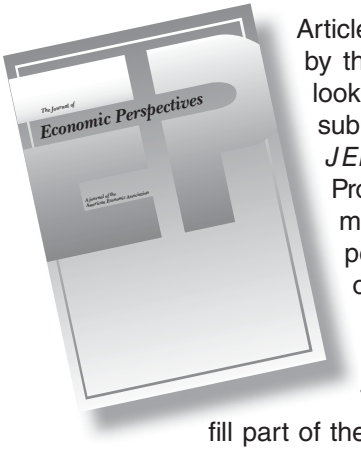
Call for Papers. The 18th **International Conference on Macroeconomic Analysis and International Finance** will be held 29–31 May 2014, at the University of Crete, Rethymno, Crete, Greece. Keynote Speakers will be Alberto Alesinia (Harvard), Fabio Canova (EUI), Stephen G. Hall (Leicester), Yannis Ioannides (Tufts), and Roberto Rigobon (MIT). The conference solicits papers in all areas of macroeconomic theory and policy and international finance. All papers must be submitted electronically through http://www.soc.uoc.gr/macro/. **The deadline for submission is December 31, 2013.** For further information, please contact: Professor Athanasios P. Papadopoulos, University of Crete, at appapa@econ.soc.uoc.gr.

Call for Papers. The 28th **Annual Conference of the European Society for Population Economics (ESPE)** will take place on June 18–21, 2014, at the University of Minho, Braga, Portugal. The aim of the conference is to facilitate the exchange of research ideas and results across a range of fields, including the economics of the household, labor economics, public economics, demography, and health economics. Papers must be submitted electronically using the online submission form: http://editorialexpress.com/conference/ESPE2014/. **The submission deadline is February 1, 2014.** For more details, please visit the conference web page: www.especonferences.org.

Call for Proposals. Proposals are invited for individual presentations on the implications of hydraulic fracturing at the community level, as part of **"The Implications of Hydraulic Fracturing for Creating Sustainable Communities,"** one of the SUNY Conversations in the Disciplines, to be held April 10–11, 2014 at Binghamton University. Please note that the purpose of this CID is not to either promote or discourage hydraulic fracturing but to understand the impacts, both positive and negative, and provide communities with the knowledge and tools that they need to decide what is in their own best interest. Presentations should be 15–20 minutes each. There will also be a poster session. To submit a proposal, email Tom Sinclair, Associate Professor of Public Administration (sinclair@binghamton.edu), **by December 15, 2013**. For details, see: http://www.isecoeco.org/implications-of-hydraulic-fracturing-for-creating-sustainable-communities/.

Harberger Prize for Retrospective Analysis. The ***Journal of Benefit–Cost Analysis*** awards the 2013 Arnold Harberger Prize for Retrospective Analysis to David Greenberg, Victoria Deitch, and Gayle Hamilton for their synthesis of welfare-to-work studies. Honorable mention was awarded to Chris Rohlfs for his study of the benefits and costs of the military draft. These and other articles may be downloaded from the *JBCA* website at www.degruyter.com/jbca.

CONSIDERATIONS FOR THOSE PROPOSING TOPICS AND PAPERS FOR JEP



Articles appearing in the journal are primarily solicited by the editors and associate editors. However, we do look at all unsolicited material. Due to the volume of submissions received, proposals that do not meet *JEP*'s editorial criteria will receive only a brief reply. Proposals that appear to have *JEP* potential receive more detailed feedback. Historically, about 10–15 percent of the articles appearing in our pages originate as unsolicited proposals.

Philosophy and Style

The *Journal of Economic Perspectives* attempts to fill part of the gap between refereed economics research journals and the popular press, while falling considerably closer to the former than the latter. **The focus of *JEP* articles should be on understanding the central economic ideas of a question, what is fundamentally at issue, why the question is particularly important, what the latest advances are, and what facets remain to be examined.** In every case, articles should argue for the author's point of view, explain how recent theoretical or empirical work has affected that view, and lay out the points of departure from other views.

We hope that most *JEP* articles will offer a kind of intellectual arbitrage that will be useful for every economist. For many, the articles will present insights and issues from a specialty outside the readers' usual field of work. For specialists, the articles will lead to thoughts about the questions underlying their research, which directions have been most productive, and what the key questions are.

Articles in many other economics journals are addressed to the author's peers in a subspecialty; thus, they use tools and terminology of that specialty and presume that readers know the context and general direction of the inquiry. By contrast, **this journal is aimed at all economists, including those not conversant with recent work in the subspecialty of the author.** The goal is to have articles that can be read by 90 percent or more of the AEA membership, as opposed to articles that can only be mastered with abundant time and energy. Articles should be as complex as they need to be, but not more so. Moreover, the necessary complexity should be explained in terms appropriate to an audience presumed to have an understanding of economics generally, but not a specialized knowledge of the author's methods or previous work in this area.

The *Journal of Economic Perspectives* is intended to be scholarly without relying too heavily on mathematical notation or mathematical insights. In some

cases, it will be appropriate for an author to offer a mathematical derivation of an economic relationship, but in most cases it will be more important that an author explain why a key formula makes sense and tie it to economic intuition, while leaving the actual derivation to another publication or to an appendix.

JEP does not publish book reviews or literature reviews. Highly mathematical papers, papers exploring issues specific to one non-U.S. country (like the state of agriculture in Ukraine), and papers that address an economic subspecialty in a manner inaccessible to the general AEA membership are not appropriate for the *Journal of Economic Perspectives*. Our stock in trade is original, opinionated perspectives on economic topics that are grounded in frontier scholarship. If you are not familiar with this journal, it is freely available on-line at <http://e-*JEP*.org>.

Guidelines for Preparing JEP Proposals

Almost all *JEP* articles begin life as a two- or three-page proposal crafted by the authors. If there is already an existing paper, that paper can be sent to us as a proposal for *JEP*. However, given the low chances that an unsolicited manuscript will be published in *JEP*, no one should write an unsolicited manuscript intended for the pages of *JEP*. **Indeed, we prefer to receive article proposals rather than completed manuscripts.** The following features of a proposal seek to make the initial review process as productive as possible while minimizing the time burden on prospective authors:

- Outlines should begin with a paragraph or two that precisely states the main thesis of the paper.
- After that overview, an explicit outline structure (I., II., III.) is appreciated.
- The outline should lay out the expository or factual components of the paper and indicate what evidence, models, historical examples, and so on will be used to support the main points of the paper. The more specific this information, the better.
- The outline should provide a conclusion
- Figures or tables that support the article's main points are often extremely helpful.
- The specifics of fonts, formatting, margins, and so forth do not matter at the proposal stage. (This applies for outlines and unsolicited manuscripts).
- Sample proposals for (subsequently) published *JEP* articles are available on request.
- For proposals and manuscripts whose main purpose is to present an original empirical result, please see the specific guidelines for such papers below.

The proposal provides the editors and authors an opportunity to preview the substance and flow of the article. For proposals that appear promising, the editors provide feedback on the substance, focus, and style of the proposed article. After the editors and author(s) have reached agreement on the shape of the article (which may take one or more iterations), the author(s) are given several months to submit a completed first draft by an agreed date. This draft will receive detailed comments from the editors as well as a full set of suggested edits from *JEP*'s Managing Editor. Articles may undergo more than one round of comment and revision prior to publication.

Readers are also welcome to send e-mails suggesting topics for *JEP* articles and symposia and to propose authors for these topics. If the proposed topic is a good fit for *JEP*, the *JEP* editors will work to solicit paper(s) and author(s).

Correspondence regarding possible future articles for *JEP* may be sent (electronically please) to the assistant editor, Ann Norman, at <anorman@JEPjournal.org>. Papers and paper proposals should be sent as Word or pdf e-mail attachments.

Guidelines for Empirical Papers Submitted to JEP

The *JEP* is not primarily an outlet for original, frontier empirical contributions; that's what refereed journals are for! Nevertheless, *JEP* occasionally publishes original empirical analyses that appear uniquely suited to the journal. In considering such proposals, the editors apply the following guidelines (in addition to considering the paper's overall suitability):

1. The paper's main topic and question must not already have found fertile soil in refereed journals. *JEP* can serve as a catalyst or incubator for the refereed literature, but it is not a competitor.
2. In addition to being intriguing, the empirical findings must suggest their own explanations. If the hallmark of a weak field journal paper is the juxtaposition of strong claims with weak evidence, a *JEP* paper presenting new empirical findings will combine strong evidence with weak claims. The empirical findings must be robust and thought provoking, but their interpretation should not be portrayed as the definitive word on their subject.
3. The empirical work must meet high standards of transparency. *JEP* strives to only feature new empirical results that are apparent from a scatter plot or a simple table of means. Although *JEP* papers can occasionally include regressions, the main empirical inferences should not be regression-dependent. Findings that are not almost immediately self-evident in tabular or graphic form probably belong in a conventional refereed journal rather than in *JEP*.

The American Economic Association Awards Recipients

John Bates Clark Medal

The John Bates Clark medal was awarded biennially from 1947–2009 to that American economist under the age of forty who is judged to have made the most significant contribution to economic thought and knowledge. From 2010 forward, the Clark Medal has been awarded annually. The Medal winner is:

RAJ CHETTY
Harvard University

Distinguished Fellows

The recognition of Distinguished Fellows began in 1965. Past presidents of the Association are Distinguished Fellows. Additional Distinguished Fellows may be elected, but not more than four in any one calendar year from economists of high distinction in the United States and Canada. The following economists are Distinguished Fellows for 2013:

HAROLD DEMSETZ
University of California, Los Angeles

STANLEY FISCHER
Bank of Israel

JERRY HAUSMAN
Massachusetts Institute of Technology

PAUL JOSKOW
Massachusetts Institute of Technology

American Economic Association
www.vanderbilt.edu/AEA/

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 following Statement of Ownership, Management and Circulation is provided in accordance with the requirements, as contained in 39 U.S.C. 3658. The *Journal of Economic Perspectives* is owned, managed and published by the American Economic Association, a nonprofit scientific organization, located at 2014 Broadway, Suite 305, Nashville, Davidson County, TN 37203-2425. The Editor is Professor David H. Autor, *Journal of Economic Perspectives*, Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive E52-371, Cambridge, MA 02142-1347. During the preceding twelve months the average number of copies printed for each issue was 8,742; the average paid circulation, 7,316; the average free distribution, 111; the average number of copies distributed 7,427. Corresponding figures for May 2013, the issue published nearest to filing date: total number of copies printed, 7,998; total paid circulation,6,666; total free distribution, 137; total distribution, 6,803.

The Journal of
Economic Perspectives

Fall 2013, Volume 27, Number 4

Symposia

The First 100 Years of the Federal Reserve

Economics and Moral Virtues

Articles

AMERICAN ECONOMIC ASSOCIATION