

the CATO JOURNAL

AN INTERDISCIPLINARY JOURNAL OF PUBLIC POLICY ANALYSIS

ARTICLES

- Andreas Hoffmann
and Gunther Schnabl** Adverse Effects of Unconventional Monetary Policy
- Stephen Kirchner** The G20 and Global Governance
- Russell S. Sobel
and J. R. Clark** Interest Group Activity and Government Growth:
A Causality Analysis
- Richard E. Wagner** The Peculiar Business of Politics
- William T. Gavin** A Critique of Proposals to Raise the Fed's Inflation
Target
- Mark A. Calabria** Behavioral Economics and Fed Policymaking
- Brian J. L. Berry,
Jayshree Bihari,
and Euel Elliott** The Limits of Knowledge and the Climate Change
Debate
- Austin P. Johnson** The Impact of Corporate Tax Policy on State Trade
Flows
- Adam G. Byrne** Removing Direct Contribution Limits: Free Speech
and Disclosure
- Joshua C. Hall
and Shree B. Pokharel** Barber Licensure and the Supply of Barber Shops:
Evidence from U.S. States
- Tain-Jy Chen
and Ying-Hua Ku** Rent Seeking and Entrepreneurship:
Internet Startups in China
- Igor Lukšić
and Milorad Katnić** The Making of a State: Transition in Montenegro
- John B. Taylor** Interest on Reserves and the Fed's Balance Sheet

BOOK REVIEWS

**Richard Morrison • Nick Zaiac • Brad Stapleton
Michael F. Cannon • Randal J. Meyer**

the CATO JOURNAL

VOLUME 36 NUMBER 3 • PUBLISHED BY THE CATO INSTITUTE • FALL 2016

Editor
JAMES A. DORN

Managing Editor
TOM CLOUGHERTY

Book Review Editor
TREVOR BURRUS

EDITORIAL BOARD

TERRY L. ANDERSON
Property and Environment
Research Center

CHARLES W. BAIRD
California State University,
East Bay

RANDY E. BARNETT
Georgetown University School
of Law

JOHN H. COCHRANE
University of Chicago

KEVIN DOWD
Durham University

RICHARD A. EPSTEIN
University of Chicago

JAGADEESH GOKHALE
University of Pennsylvania

JAMES GWARTNEY
Florida State University

STEVE H. HANKE
Johns Hopkins University

RANDALL S. KROZNER
University of Chicago

DEEPAK LAL
University of California,
Los Angeles

STEPHEN MACEDO
Princeton University

ANTONIO MARTINO
University of Rome, Luiss

JEFFREY A. MIRON
Harvard University

KEVIN M. MURPHY
University of Chicago

GERALD P. O'DRISCOLL JR.
Cato Institute

SAM PELTZMAN
University of Chicago

JUDY SHELTON
Fredericksburg, Virginia

VERNON L. SMITH
Chapman University

JOHN B. TAYLOR
Stanford University

ROLAND VAUBEL
Mannheim University

RICHARD E. WAGNER
George Mason University

LAWRENCE H. WHITE
George Mason University

LELAND B. YEAGER
Auburn University

THE CATO JOURNAL (ISSN 0273-3072) is published in the winter, spring/summer, and fall by the Cato Institute, 1000 Massachusetts Ave., N.W., Washington, D.C. 20001-5403. Annual subscriptions are \$22 for individuals and \$50 for institutions. Single copies are \$8. Two-year and three-year subscriptions are \$38 and \$55, respectively, for individuals, and \$85 and \$125, respectively, for institutions. Foreign subscribers should add \$5 per year for regular delivery and \$10 per year for airmail delivery.

Correspondence regarding subscriptions, changes of address, acquiring back issues, advertising and marketing matters, and so forth should be addressed to the Publications Department. All other correspondence, including requests to quote or reproduce material, should be addressed to the editor.

Unsolicited manuscripts cannot be returned and will be acknowledged only if accompanied by a stamped, self-addressed envelope. There is no submission fee. Articles should focus on specific public policy issues, be scholarly, but also be intelligible to the interested lay reader. Accepted manuscripts must conform to the CATO JOURNAL's style requirements. A style sheet is available on request.

Authors may submit their papers electronically, preferably in MS Word, to the editor (jdorn@cato.org), along with an abstract of no more than 250 words.

Unsolicited book reviews cannot be returned and will be acknowledged only if they are accepted for publication. They should conform to the above requirements and should provide an in-depth analysis of major studies of public policy. Reviews should range from 800 to 1,000 words.

The views expressed by the authors of the articles are their own and are not attributable to the editor, the editorial board, or the Cato Institute.

THE CATO JOURNAL is listed in *Cabell's Directory*, *Journal of Economic Literature*, *Current Contents/Social and Behavioral Sciences*, *America: History and Life*, *Historical Abstracts*, *International Political Science Abstracts*, and the ABI/INFORM database.

Printed in the United States of America.

Copyright © 2016 by the Cato Institute.
All rights reserved.

ARTICLES

Andreas Hoffmann and Gunther Schnabl
Adverse Effects of Unconventional Monetary Policy 449

Stephen Kirchner
The G20 and Global Governance 485

Russell S. Sobel and J. R. Clark
Interest Group Activity and Government Growth:
A Causality Analysis 507

Richard E. Wagner
The Peculiar Business of Politics 535

William T. Gavin
A Critique of Proposals to Raise the Fed's Inflation Target 557

Mark A. Calabria
Behavioral Economics and Fed Policymaking 573

Brian J. L. Berry, Jayshree Bihari, and Euel Elliott
The Limits of Knowledge and the Climate Change Debate 589

Austin P. Johnson
The Impact of Corporate Tax Policy on State Trade Flows 611

Adam G. Byrne
Removing Direct Contribution Limits: Free Speech and Disclosure 625

Joshua C. Hall and Shree B. Pokharel
Barber Licensure and the Supply of Barber Shops:
Evidence from U.S. States 647

Tain-Jy Chen and Ying-Hua Ku
Rent Seeking and Entrepreneurship: Internet Startups in China 659

Igor Lukšić and Milorad Katnić
The Making of a State: Transition in Montenegro 689

John B. Taylor
Interest on Reserves and the Fed's Balance Sheet 711

BOOK REVIEWS

**Markets without Limits: Moral Virtues and
Commercial Interests**
Jason Brennan and Peter M. Jaworski
Reviewed by Richard Morrison 721

Zoning Rules! The Economics of Land Use Regulation
William A. Fischel
Reviewed by Nick Zaiac 725

**The Unquiet Frontier: Rising Rivals, Vulnerable Allies,
and the Crisis of American Power**
Jakub J. Grygiel and A. Wess Mitchell
Reviewed by Brad Stapleton 731

**Side Effects and Complications: The Economic
Consequences of Health-Care Reform**

Casey B. Mulligan

Reviewed by **Michael F. Cannon**

734

**The Law of the Land: A Grand Tour of Our
Constitutional Republic**

Akhil Reed Amar

Reviewed by **Randal J. Meyer**

739

ADVERSE EFFECTS OF UNCONVENTIONAL MONETARY POLICY

Andreas Hoffmann and Gunther Schnabl

Following the recent waves of financial crises in the advanced economies and a prolonged period of low interest rates, major parts of the world economy are experiencing low growth, a rise in financial volatility, and low rates of inflation. Specifically, in Japan and in large parts of the eurozone the crises persist. In large parts of the world, unconventional monetary policies—that is, ultra-low interest rates and large-scale asset purchases, also known as “quantitative easing” (QE)—are seen as important determinants of employment and growth. The announced exit from unconventional monetary policy in the United States, where growth appears more robust, has clouded the growth perspectives of many emerging market countries. The Chinese growth engine, which was a main driver of world growth during the 2000s, has begun to stutter—and emerging market corporate bond markets have come under pressure.

Macroeconomists have identified several reasons for the recent wave of financial crises in the advanced economies. One strand of literature explains financial crises as result of a random or exogenous shock, amplified by the irrationality of human action (Keynes 1936, De Grauwe 2011), asymmetric information and financial constraints

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Andreas Hoffmann is Assistant Professor at the Institute of Economic Policy (IEP) at Leipzig University. Gunther Schnabl is Professor of Economics and Director of IEP. The authors thank Jim Dorn for valuable comments and the Friedrich August von Hayek-Foundation and Jackstädt Foundation for financial support.

(Bernanke, Gertler, and Gilchrest 1996). Another strand of literature suggests that a savings glut—caused by a higher saving propensity of the aging populations in Germany, China, and Japan—has contributed to a fall in (natural) interest rates in advanced economies (Bernanke 2005, Summers 2014, von Weizsäcker 2014).

On the contrary, assessments (implicitly) based on the Taylor rule suggest that overly expansionary monetary policies during the 2000s sowed the seeds for financial exuberance and therefore the current crisis (Taylor 2007; Jorda, Schularick, and Taylor 2015). Adrian and Shin (2008), Brunnermeier and Schnabel (2014), as well as Hoffmann and Schnabl (2008, 2011, 2014), have shown that overly expansionary monetary policy can contribute to financial market bubbles that lead to crisis. Selgin (2014), Selgin, Lastrapes, and White (2012), as well as Howden and Salerno (2014), see public central banks at the root of macroeconomic instability.

Depending on the view of the very roots of the crisis, policy recommendations point in different directions. One side emphasizes the need for unconventional monetary policy to stabilize the financial system—for example, by easing collateral constraints to maintain growth and employment (Draghi 2014, Bernanke 2014). In contrast, the other side sees ultra-low interest rate policy and QE as major sources of distortions and bubbles. These critics demand a timely exit from unconventional monetary policy to prevent further distortions caused by boom and bust in the financial markets.

This article contributes to the second strand of literature. We discuss the developments during the last three decades against the backdrop of the monetary overinvestment theories of Wicksell (1898), Mises (1912), and Hayek (1929, 1937). In particular, we elaborate on channels through which ultra-low interest rate policies can contribute to a decline in investments and growth in the world economy.

Monetary Overinvestment Theories and Boom-and-Bust in Financial Markets

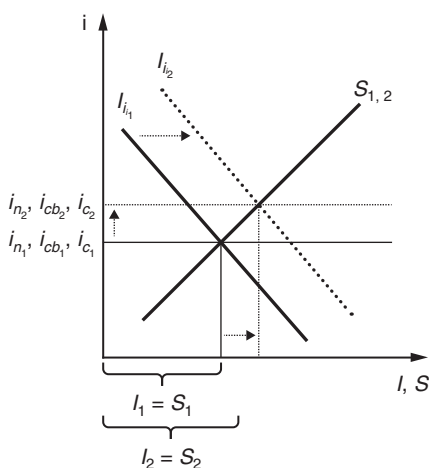
In order to model boom-and-bust cycles based on the overinvestment theories of Wicksell, Mises, and Hayek we distinguish between four types of interest rates. First, the internal interest rate i_t reflects the (expected) returns of (planned) investment projects. Second, Wicksell's natural interest rate i_n is the interest rate that balances the

supply (saving) and demand (investment) of capital.¹ Third, the central bank's policy interest rate i_{cb} shall represent the interest rate that commercial banks are charged by the central bank for refinancing operations. Fourth, we define the capital market interest rate i_c as the interest rate set by the private banking (financial) sector for credit provided to private enterprises. For simplicity, we shall assume that central banks simply set a policy interest rate and that the capital market interest rate equals this policy interest rate (see Hoffmann and Schnabl 2011).

According to the monetary overinvestment theories, an economy is in equilibrium when the natural rate of interest equals the policy interest rate—that is, planned saving (S) equals planned investment (I). An economic upswing starts when positive expectations due to an important innovation raise the internal interest rate of investment, bringing about a rise in investment demand at given interest rates. In Figure 1, this corresponds to a rightward shift of the investment curve from I_{i_1} to I_{i_2} . The natural rate of interest rises along from i_{n_1} to i_{n_2} . Credit demand in the economy rises.

¹Wicksell and Hayek have different concepts of the natural interest rate. According to Wicksell's work, the deviation of the central bank interest rate (capital market interest rate) from the natural rate of interest (which guarantees goods market equilibrium) disturbs the equilibrium between ex ante saving (S) and investment (I) plans, bringing about inflationary ($I > S$) or deflationary ($S > I$) processes. During an inflationary credit boom, the supply of goods cannot satisfy the additional demand for goods at given prices. Therefore, Wicksell's natural rate of interest is the interest rate at which inflation is zero (or at the target level). In contrast to Say, in Wicksell's framework money is not neutral; additional money affects decisions of economic agents. Wicksell also interpreted the natural rate of interest as real rate of return or marginal efficiency of new investment projects. Building on Wicksell's inflation theory, Mises and Hayek aimed to explain business cycles caused by the deviation of the central bank interest rate (capital market interest rate) from the natural rate of interest. They attribute the main role in the creation of cycles to central banks and the private banking sector. In contrast to Wicksell, Hayek emphasized the importance of the intertemporal alignments of plans of producers and consumers, explaining malinvestment or overinvestment phenomena as mismatches between the production structure and consumer preferences. The natural interest rate is the interest rate that aligns saving and consumption preferences with the production structure over time. A fall in the central bank interest rate (capital market interest rate) below the natural interest rate causes a cumulative inflationary process, creating distortions in the production structure that later make an adjustment necessary (unless the central bank keeps on inflating credit at an ever-increasing pace and artificially prolongs the credit boom).

FIGURE 1
EQUILIBRIUM

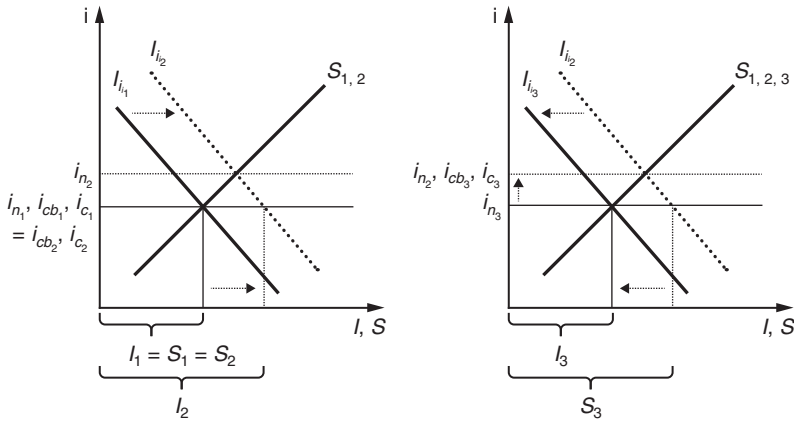


If the central bank increases the policy interest rate from i_{cb_1} to i_{cb_2} , assuming a perfect interest rate transmission to credit markets, planned saving and investment in the economy will stay in equilibrium ($S_2 = I_2$). If, however, the central bank does not raise the policy interest rate, ($i_{n_1} = i_{cb_1} = i_{c_2} < i_{n_2}$) as shown in Figure 2, relatively low interest rates will give rise to an unsustainable monetary overinvestment boom. Holding policy rates too low (for too long) will be referred to as a Type 1 error in monetary policy.

To market participants, a rise in credit to the private sector at constant interest rates signals that saving activity of households increased. Additional investment projects aim to satisfy the expected rise in future consumption. As planned household saving did not actually increase, an unsustainable disequilibrium between ex ante saving and investment $S_2 < I_2$ at $i_{c_2} < i_{n_2}$ arises. In the following, additional investments of some enterprises trigger additional investments of other enterprises (cumulative upward process). As soon as capacity limits are reached and employment is high, wages and prices rise.

At first, rising prices signal additional profits and therefore trigger a further increase in investment. There may be spillovers to financial markets. Increases in expected profits of companies are typically associated with rising stock prices. Given relatively low interest rates on deposits, shares are attractive. When stock prices

FIGURE 2
BOOM AND BUST



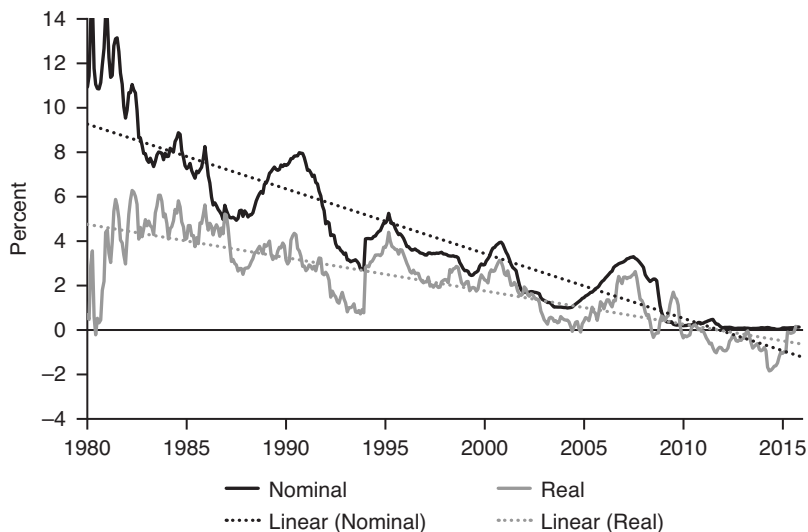
move upward, trend-followers will provide extra momentum such that “the symptoms of prosperity themselves finally become . . . a factor of prosperity” (Schumpeter 1911: 226). Consumption is fuelled by rising stock prices via the wealth channel, which leads, with a lag, to an increasing price level.

The boom turns bust when the central bank increases the central bank interest rate to stem inflation (Mises 1912; Hayek 1929, 1937). Then investment projects with an internal interest rate below the risen natural interest rate turn out unprofitable. The fall in investment of some firms will depress investment of other firms as expected returns fall. When stock (and other asset prices) burst, balance sheets of firms and banks worsen, bringing about further disinvestment (cumulative downward process). Wages fall and unemployment rises. The investment curve shifts back from I_{i_2} to I_{i_3} .

In this situation, the central bank should cut the central bank interest rate to contain the downward spiral. Based on the monetary overinvestment theories, we consider holding policy interest rates at a high plateau during the downturn a Type 2 monetary policy error. Figure 2 shows that when the policy interest rate is above the natural interest rate ($i_{cb_3} = i_{c_3} > i_{n_3}$), credit supply is restricted further such that ex ante saving is higher than investment ($S_3 > I_3$).

According to Mises (1949: 572), “The wavelike movement affecting the economic system, the recurrence of periods of boom which are followed by periods of depression is the unavoidable outcome of

FIGURE 3
SHORT-TERM INTEREST RATES IN G3 ECONOMIES
(ARITHMETIC AVERAGES)



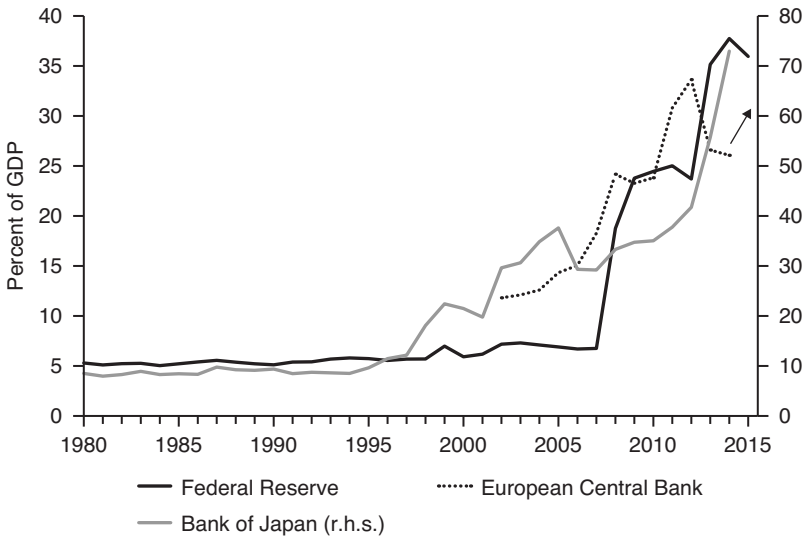
NOTE: The graph shows the arithmetic average of interest rates of the G3 economies, namely the United States, Japan, and Germany up to 1998; later dates include the interest rate of the eurozone.

SOURCE: International Monetary Fund via Datastream, 2016.

the attempts, repeated again and again, to lower the gross market rate of interest by means of credit expansion.” Hoffmann and Schnabl (2008, 2011, 2014) outline that the spate of boom and bust cycles in different parts of the world since the 1980s can be understood based on monetary overinvestment theories.

They make, however, one clear distinction: central banks have tended to hold policy interest rates too low during periods of economic upswing, fueling booms in financial markets (i.e., Type 1 errors in monetary policy). During financial crises, however, central banks have slashed interest rates decisively to stabilize the economy and prevent Type 2 errors in monetary policy. Specifically, in the so-called Jackson Hole consensus, U.S. central bankers agreed that central banks do not have sufficient information to spot bubbles, but should react swiftly in times of financial turmoil (Blinder and Reis 2005). Consequently, we observe (in cycles) a downward trend in nominal and real interest rates in the large economies (Figure 3).

FIGURE 4
G3 CENTRAL BANK ASSETS AS PERCENTAGE OF GDP



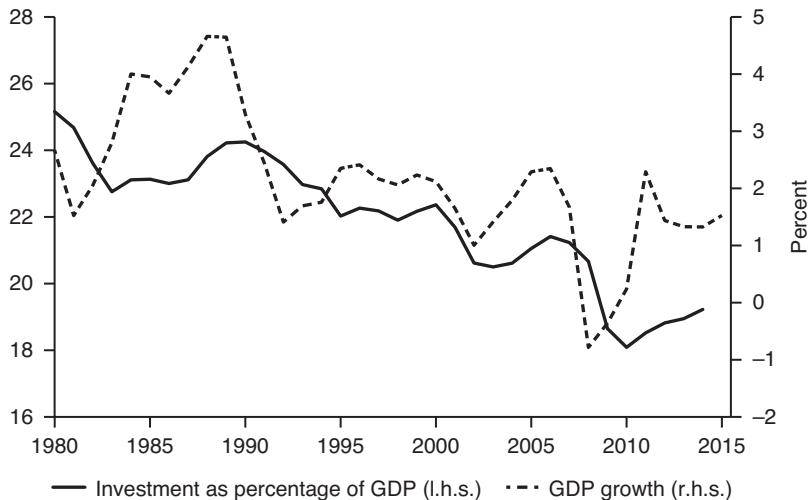
SOURCES: World Economic Outlook, European Central Bank, and Eurostat, 2016.

Once interest rates have reached the zero-bound (in Japan since 1999 and the United States and Europe since 2008), central bank balance sheets have been inflated more aggressively to prevent a meltdown of the financial sector by pushing down the interest rate on the long end of the yield curve (Figure 4). The discussions on tapering and the long-delayed increase in interest rates (for the first time after nine years) by the Federal Reserve, signal that an exit from such low interest rate policies is a difficult endeavor because large distortions have emerged and politically it is difficult to end ultra-low rates on government debt (see Buiter 2010).

Negative Growth Effects of Low Interest Rate Policies

Although the drop in interest rates and the dramatic expansion of central bank balance sheets had a stabilizing effect on financial markets and employment, investment and growth can be dampened in the medium and long term. This is most evident in Japan, where an exuberant financial market boom (the so-called Bubble Economy

FIGURE 5
G3 REAL GROWTH AND INVESTMENT AS SHARE OF GDP
(ARITHMETIC AVERAGES)



SOURCE: International Monetary Fund via Datastream, 2016.

from 1985 until 1989) and the following crisis (2.5 decades now lost) set in around 15 years earlier than in Europe and the United States. But along with investments, real growth is also declining on average for all three G3 countries (Figure 5). In the following, the cause of the decline in growth dynamics is identified as the creeping nationalization of lending, declining (marginal efficiency of) investments, as well as financial and real wage repression promoted by the redistributive effects of monetary expansion.

All three effects can be linked to the unconventional monetary policies in the major industrialized countries.² In this sense, we

²The link between monetary policies and the stylized facts of macroeconomic trends is based on the selective choice of particularly significant links. Empirical studies of causal relationships between monetary policy and, for example, developments in financial markets usually work on the assumption that national monetary policies affect national financial markets. This is obvious, although not necessarily true in globalized financial markets (Hoffmann and Schnabl 2014, Hoffmann 2014). An expansionary monetary policy in one country can—depending on the difficult-to-comprehend dynamics of financial markets, national specifics, and national regulatory arrangements—also affect any other segment of the international capital markets.

present a counter-hypothesis to theoretical constructs that assume that nominal and real interest rates have fallen to or below zero due to exogenous factors. According to Weizsäcker (2014), Bernanke (2005), and Summers (2014), aging societies in individual industrialized countries (including China) and growing retirement savings are causing a savings glut (particularly in Japan, China, and Germany). This coincides with declining investment activity due to weak technological progress.

The result is a declining (what they call) natural real interest rate (r), which involves an increasing probability of financial market bubbles, while product markets remain in equilibrium with $I(r) = S(r)$. Similarly, Laubach and Williams (2015) suggest that the fall in trend GDP growth rates triggered a decline in the natural rate of interest (in the United States). Therefore, this literature suggests that central bank interest rates have to decline further to match the ever-falling natural rate and guarantee goods market equilibrium. In contrast, we maintain that unconventional monetary policies—especially the ultra-low interest rates—are responsible for the decline in investment and growth and therefore the decline of the natural interest rate as for instance defined by Laubach and Williams (2015).

Creeping Nationalization of Lending

Asymmetric monetary policy and, since the major crises, the (almost) free and quasi-unlimited provision of central bank liquidity to commercial banks have incentive effects. First, new excesses are encouraged on the financial markets (see Adrian and Shin 2008, Brunnermeier and Schnabel 2014). The bursting of these asset price bubbles causes more banks to record book losses on assets. The portfolio of bad loans increases sharply. Entire financial sectors run into trouble and threaten to lead to a credit crunch (Ishikawa and Tsutsui 2005).³ Since the banks suffer high book losses on their balance sheets, their equity shrinks. This forces them to restrict lending to (high-risk) companies or for new investment projects. The crisis in the banking sector sparks a crisis in the corporate sector, in turn leading to a rise in unemployment.

³Since the low interest rate policy in Japan has continued the longest, the effects on the financial sector are most evident there, as reflected in research on the impact of a zero interest rate policy (ZIRP) on the financial and corporate sectors. This is why Japan, in particular, is used as a case study.

In order to counteract a credit crunch originating from the supply side of the lending market, Posen (2000) proposes interest rate reductions and extensions of a central bank's balance sheet. Banks receive "fresh money" from the central bank, which aims to enable them to extend corporate loans. A spillover of the crisis from the financial to the corporate sector should be prevented, facilitating the economic recovery. By cutting interest rates and purchasing assets, the central banks contribute to minimizing book losses on assets, or even bringing about a significant shift into positive territory. The portfolio of bad loans (as a proportion of the balance sheet total) does not continue to grow, or it diminishes. This curbs contagion effects in the financial markets.

However, unconventional monetary policy during a crisis can lead to an implicit nationalization of money and credit markets. In money markets, market structures change during a crisis because of rising distrust among banks. Interbank lending of commercial banks is substituted by borrowing from the central bank. A zero interest rate policy (ZIRP) perpetuates this situation, because it drives profit margins in the money markets down to a minimum (McKinnon 2012). Banks with excess liquidity no longer have any incentive to act as a supplier in the money markets. Even if banks requiring liquidity were to offer higher interest rates in order to create a supply, offering high interest rates signals higher risk. The requested loan is not granted.⁴ As a result, the private supply of money is also substituted in the long term by the money supplied by the central bank. Banks with excess liquidity invest with the central bank.⁵

In the lending markets supported by the banks, too, ZIRP contributes to market shrinkage (Schnabl 2015). ZIRP amounts to a subsidy for companies that are traditionally aggregated demanders on the lending market. Especially for large companies that can issue their own securities, in many cases borrowing costs drop. The demand and willingness to pay for shares (equity) rise because alternative forms of investment, such as bank deposits and government bonds, bear low interest. Then the prices of shares and securities rise. The low cost of obtaining capital gives rise to additional profit for large companies,

⁴This represents a market failure according to Akerlof (1970).

⁵In line with our article, David Malpass (2015) argues that "the zero interest rate freezes interbank markets and allocates credit away from the economy's growth engines—new businesses."

which becomes visible in the form of increasing corporate savings.⁶ The demand for loans declines, and companies tend to purchase more of their own shares, which increases the profit per share and therefore, in many cases, the bonus payments of the upper management.

The banks, which are more strictly regulated and need to amass more equity following the banking crisis, have an incentive to restrict lending to higher-risk companies.⁷ If the larger, less risky companies withdraw from the loan portfolios of commercial banks, then the average risk in the banks' loan portfolios increases. Loans to comparatively high-risk small and medium-sized enterprises have to be restricted. Then loans to the private sector can be substituted by loans to the public sector, if national debt increases during the crisis. Unlike companies' investment risks, the default risks of the state are implicitly guaranteed by the central bank if it signals additional purchases of government bonds. This process is favored by the Basel capital adequacy rules, which do not set out equity reserves for the purchase of government bonds.⁸

Under the constraint of ZIRP, the hope that a rapid recapitalization of banks will prevent a credit crunch—and thus also the creeping nationalization of banks—may be in vain. The reason is that incentive structures for substituting loans to nonfinancial corporations by providing loans to the public sector remain unchanged as long as public debt rises. Since the state has no savings, it must obtain the capital needed to recapitalize the banks by issuing government bonds. The banks can use the additional lending potential generated by their recapitalization to purchase these government bonds, which are issued to finance recapitalization. In such cases, lending is not extended to companies.

⁶It is therefore difficult to provide sound empirical evidence for the hypothesis of the global liquidity glut (e.g., Bernanke 2005), because the assumed structural increase in household savings cannot be observed in any of the countries with surplus savings (overinvestments). The increase in aggregate savings surpluses in these countries (relative to investments) is rather due to the increase in corporate savings (especially resulting from declining financing costs) and the fall in investments.

⁷In Japan, these are mainly the relatively economically weak SMEs, and in Europe primarily companies in the Southern European crisis states.

⁸A prerequisite for this process is that government debt increases as it happened in Japan and the United States. In Europe, the more or less effective restrictions on government debt constitute an impediment to the substitution of credit to private sector by credit to the public sector.

A further reason lending does not increase during the crisis is that banks face tighter regulation and capital requirements as a response to the crisis. In general, the banks' requirements of loan collateral are pro-cyclical.⁹ They decrease during times of prosperity and increase in a downturn. Although the central bank lowers the interest rate during a downturn and provides additional funds, the banks increase their requirements for loan collateral. New loans are not awarded, despite cuts in the interest rate, if the value of the collateral falls. Tighter regulation enhances this effect in the downswing. In contrast, existing (possible bad) loans are extended, because banks fear that defaulting enterprises erode banks' (reported) capital base. Commercial banks tend to clandestinely relax their requirements of loan collateral for already outstanding credit during a crisis.

In this way, banks become dependent on the state via two channels. First, if returns in the traditional banking business shrink,¹⁰ banks depend on the supply of free liquidity from the central bank. Any major hike in the key interest rate would cause the banking sector to falter. In Japan's case, Caballero, Hoshi, and Kashyap (2008) coined the term "zombie banks." Second, the banking sector tends to replace loans to the private sector by loans to the public sector.

Drop in Investments and Their Marginal Efficiency

Traditional banking involves accepting deposits with a positive rate of return and lending that capital, in the form of loans, to businesses and households at higher interest rates. Banks fulfil an intermediary function in which they examine the future returns on investments. Projects with higher expected returns are financed at a given interest rate. By contrast, projects with lower expected returns (where the probability of default is high) are (in the best case) rejected. The banking sector thus plays a crucial role in the allocation function of interest, separating investment projects with higher expected returns from those with lower expected returns.

⁹Following Kiyotaki and Moore (1997), two criteria apply to lending: the expected return r and the security z . For a given central bank interest rate i_{cb} , all projects where $r < i_{cb}$ or $z < Z$ is true are ruled out, where Z is the required minimum loan collateral.

¹⁰In Japan, for example, during the crisis the lending-deposit rate spread fell from approximately 3.5 percentage points to approximately 0.5 percentage points (Schnabl 2015).

If the banking system is no longer subject to strict budget constraints, then the allocation function of interest rates is undermined. In this case, rescue measures implemented during the crisis through ultra-low rates prevent or delay the structural adjustment process during crisis as stressed by Schumpeter (1911). To remain in the market banks in trouble disguise their difficult situations by prolonging loans for investment projects with low or negative returns. In Japan's case Sekine, Kobayashi, and Saita (2003) talk of forbearance lending: banks continue to provide irrecoverable loans, thus keeping themselves and (potentially) insolvent companies alive. Peek and Rosengren (2005) also associate Japan's ZIRP with a misallocation of capital in the credit sector, which keeps companies with poor profit prospects alive ("evergreening").

Thus, the constant supply of cheap liquidity by the central bank can affect the quality of the loan portfolio. Investments which would not have been financed at Wicksell's natural rate of interest continue to be financed.¹¹ Tying capital up in traditional structures restricts the financing possibilities for innovative new investments. The average efficiency of investments decreases. In the sense of Leibenstein (1966), "X-inefficiency" emerges. If enterprises can expect that cheap credit will be provided without tight conditions with respect to profitability, this expectation discourages the pursuit of innovation and cost savings.

János Kornai (1986) spoke of "soft budget constraints" in the case of companies in Central and Eastern European planned economies. Since unemployment was politically undesirable, nonprofitable companies were kept alive by supplying liquidity via a state-controlled banking system. Quian and Xu (1998) showed that such soft-budget constraints made it harder to select profitable from unprofitable projects as the ex post competition-driven selection mechanism is undermined. Instead, banks rely on ex ante screening, which is costly and less effective, putting a drag on innovative potential of the economy.¹² Indeed, Caballero, Hoshi, and Kashyap (2008) showed for

¹¹In contrast to Weizsäcker (2013) and Summers (2014), we do not attribute the fall in growth rates to exogenous increases in saving rates but to monetary policy mistakes. Therefore, we do not assume that the natural rate of interest fell structurally toward or below zero due to preference shifts.

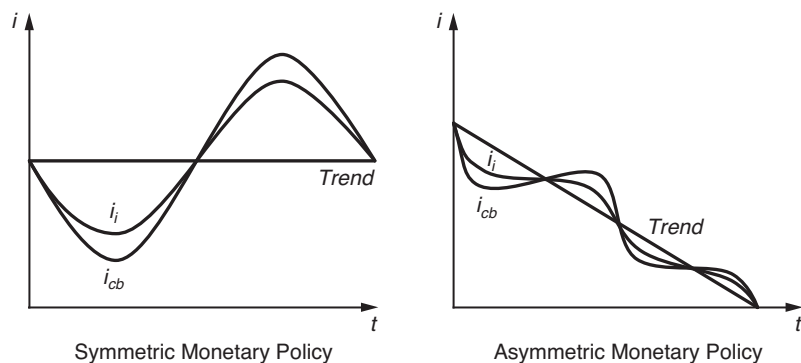
¹²See Maskin and Xu (2001) on a discussion of soft-budget constraint theories related to the transition process of planned economies toward market economies.

Japanese companies that under ZIRP company profits became dependent on cheap loans. Although the expansionary Japanese monetary policy has successfully protected jobs, the average productivity of firms has dropped considerably.

Similar developments seem to take place in other industrialized countries, in particular since the advent of ZIRP. Barnett et al. (2014) demonstrate that since 2007 the United Kingdom has seen a significant drop in productivity growth among businesses. Cardarelli and Lusinyan (2015) show for the United States that total factor productivity has dropped significantly since the turn of the millennium. Gopinath et al. (2015) provide empirical evidence for the Southern European countries since the outbreak of the European debt and financial crisis.

In Figure 6, we model the relationship—derived from monetary overinvestment theories—between the central bank interest rate and the internal interest rate, which can also be interpreted as the marginal efficiency of investments. In the monetary overinvestment theories of Wicksell (1898) and Hayek (1929), the central bank interest rate fluctuates around the natural rate of interest. During times of prosperity, refinancing conditions being too favorable causes additional investment projects with low expected returns to be financed. The marginal and average efficiency of investments decreases. During a downturn and crises, investment projects with low internal rates of return are cancelled. The marginal and average efficiency of investments increases (left side of Figure 6).

FIGURE 6
 SYMMETRIC VERSUS ASYMMETRIC MONETARY POLICY



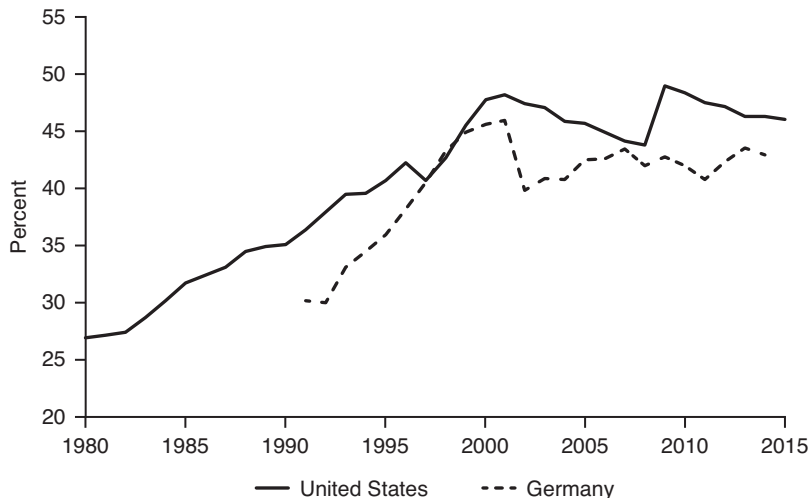
However, if the markets expect a gradually declining interest rate level due to an established asymmetric monetary policy path (like in Figure 3), then the average internal rate of interest necessary to ensure the repayment of loans will also drop. The declining trend in central bank interest rates leads on the one hand to a partial or total absence of the structural adjustment process during the crisis. The marginal efficiency of investments falls during the boom before the crisis, and remains largely constant during the crisis. If during the crisis the interest rate is lowered again to below the precrisis level, then the average and marginal efficiency of investments continue to decrease (right part of Figure 6).

We explain the reduction in investments in fixed assets (as shown in Figure 5) by incentive effects of an asymmetric monetary policy on different types of investments. During financial crises, an asymmetric monetary policy constitutes an implicit insurance mechanism, because an abundance of central bank liquidity is quickly provided to stabilize the financial markets (Jackson Hole Consensus). Interest rate cuts and an unconventional monetary policy stop or even reverse the fall in asset prices. Even if prices fall in some assets classes (for instance, Japanese stocks), prices are driven upward in other asset classes (for instance, U.S. stocks), making it possible to compensate for the losses in the asset classes affected by the crisis.

In contrast, returns on real investments fall relative to the investments in the financial markets due to at least three reasons. (1) As discussed earlier, the marginal efficiency of investment is likely to decline. (2) In contrast to financial investment, there is no public insurance mechanism for the risks of individual innovations, product lines, or new production processes. (3) In addition, uncertainty grows. With growing amplitudes of boom-and-crisis cycles in the financial markets, long-term investment decisions in the real sector tend to be associated with growing uncertainty. The growing uncertainty/risk reduces the expected return of real investments.

Different expected returns on investments in fixed assets and financial investments create an incentive for companies to substitute speculative financial investments for real investments. In the original monetary overinvestment theories, too low interest rate policies contribute to unsustainable changes in the structure of the economy. The durable consumer goods and capital goods sectors expand at the expense of nondurable consumer goods sectors, signaling an intertemporal reallocation of funds in favor of projects with higher expected

FIGURE 7
FINANCIAL ASSETS RELATIVE TO NONFINANCIAL
CORPORATE ASSETS



SOURCES: Deutsche Bundesbank and Federal Reserve Bank of St. Louis, 2016.

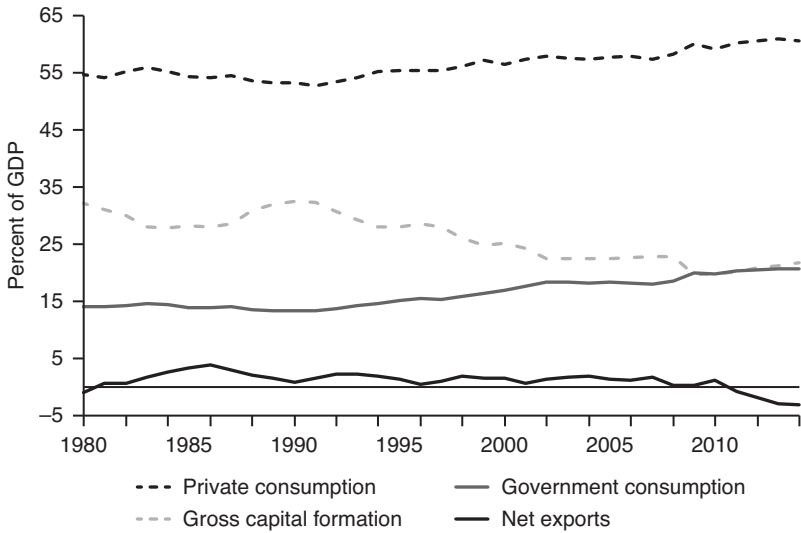
returns in later periods (vertical malinvestments). In recent boom-and-crisis cycles, the economy's structure saw further shifts from the goods market sector to the financial sector, which sees disproportionately high growth during the boom (horizontal or sectorial malinvestments) (see White 2012). The boundary between the goods market sector and the financial sector is blurred. For example, in the course of a speculative boom in real estate, growth in the real estate sector (construction) may be attributed to either the real or financial sector.¹³ If monetary policy behaves asymmetrically, then during the crisis there will be no structural adjustments in the expanded financial sector.

This blurring contributes to the fact that investments in physical capital become less significant for companies in relation to (speculative) investments in the financial markets. Accordingly, the proportion of financial assets in relation to nonfinancial corporate assets has risen steadily since the 1980s. Figure 7 shows this trend for Germany and the United States, where it can be observed until the turn of the millennium.¹⁴ From a private-sector perspective, it is

¹³Real estate is created as a speculative investment. Once the bubble bursts, many properties are left vacant.

¹⁴Since then, there is a tendency to use cash holdings for stock purchases.

FIGURE 8
DEMAND COMPONENTS OF JAPANESE GDP, 1980–2014



SOURCE: Cabinet Office, Government of Japan, 2016.

true that the average return on financial investments will seem relatively high if potential losses are counteracted by the central bank. In aggregate, however, the *ex post* returns need to be adjusted for possible state subsidies. This is for example the case when banks are recapitalized using public money, or the costs of bailouts implemented under monetary policy become visible—for example, in the form of higher inflation or the recapitalization of the central bank. From a macroeconomic perspective, returns on speculative investments in the financial markets are therefore significantly lower, or even negative.

If the financial crisis is transformed into a sustained, creeping crisis in which there is no limit to the central bank's government bond purchases, this may result in the substitution of public investments and/or government consumption for private investment. After the Japanese bubble burst in December 1989, numerous Keynesian economic stimulus programs were implemented. The construction of highways, bridges, high-speed railway lines, and public buildings bolstered growth particularly in Japan's low-growth provinces outside of Tokyo, Kansai, and Aichi (Yoshino and Mizoguchi 2010). Figure 8 shows that since the Japanese bubble burst, gross investment in

Japan as a share of GDP declined from 32 percent in 1990 to 20 percent in 2011. In the same period, government spending as a share of GDP rose from 13 percent to 21 percent. If we assume that public investments have a lower marginal efficiency than private investments, then the average efficiency of investments further decreases.

Redistributive Effects and Real Wage Repression

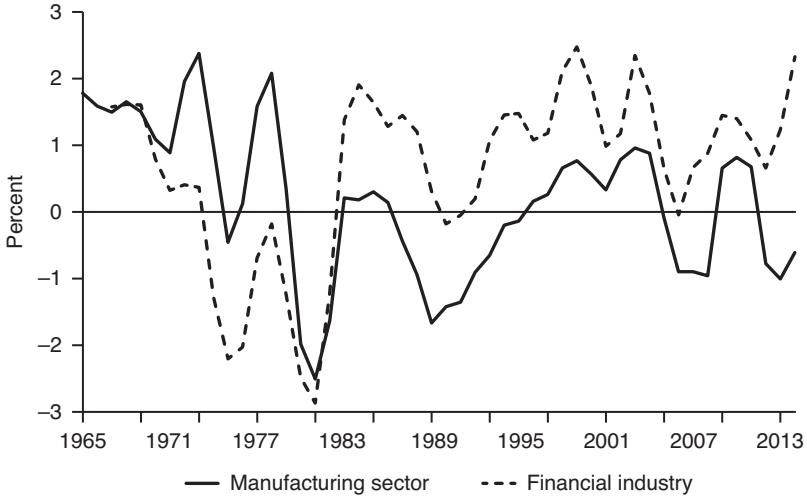
The prolonged periods of low interest rates in Japan and in the United States have also had a negative impact on consumption, because they have lowered the income of broad sections of the population. Such distributional effect has an absolute and a relative aspect. In absolute terms, with the marginal efficiency of investment declining, productivity gains also gradually decline. Therefore, the scope for real wage increases is gradually becoming smaller. This implies that in a scenario of zero productivity gains a growing real income of one social group has to come along with a declining real income of another social group.

Low interest rate policies can, for instance, favor higher-income groups, because the extra liquidity created by the central banks is initially available to the banks and other financial institutions (the so-called primary dealers) (Cantillon Effect). Following Cantillon (1931), banks benefit not only from an increase in the lending business under favorable refinancing conditions. They can also acquire stocks, real estate, and securities at lower prices. If via purchases of stocks, real estate, and securities the additional financial means provided by the central bank make their way further into other areas of the economy, then real estate, stock, and security prices increase for the next buyers. This results in redistributive effects in favor of the financial institutions, in the form of valuation gains. Furthermore, the financial sector intermediates a growing volume of capital market transactions, for which firms obtain rising commissions due to rising asset prices.

Such redistribution effects in favor of the financial sector are for instance visible in the United States.¹⁵ Figure 9 shows that until the mid-1980s the income of industrial sector workers grew faster than in the financial sector. However, since the mid-1980s employees in

¹⁵The world's largest financial market is in the United States, which is why the data can be assumed to be particularly telling.

FIGURE 9
WAGE DEVELOPMENT IN THE U.S. FINANCIAL AND
MANUFACTURING SECTORS



SOURCE: International Monetary Fund via Datastream, 2016.

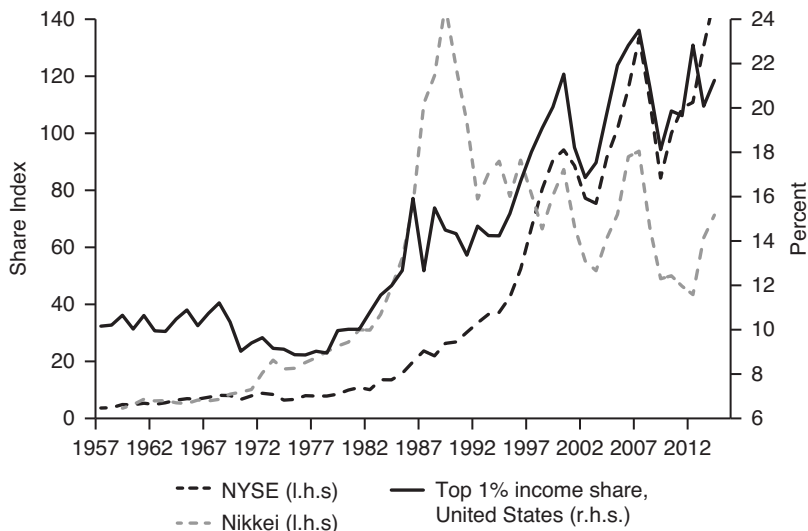
the financial sector benefitted from higher wage increases. This even applies to periods of financial market crises, during which industrial workers' wages declined more sharply than salaries in the financial sector. It is likely that financial sector executives tend to benefit more than other employees from financial institutions' windfall profits, because one-off dividends due to exceptional performance (bonuses) are more common at this level.

Figure 10 shows the potential redistributive effects between individual income groups for the United States. It is based on the assumption that monetary policy has more of an effect on asset prices than on goods prices. The left axis shows price trends on the U.S. and Japanese stock markets (NYSE and Nikkei, respectively), while the right axis plots the share of the top 1 percent of incomes as a proportion of the total income of the U.S. economy (including income from capital). There has been a clear correlation between the two variables since the mid-1990s. The development of the Japanese stock index suggests that the redistributive effects of speculative waves on globalized financial markets do not necessarily stop at national borders.

The highest income groups in the United States appear to have already benefitted from the speculative bubble in the Japanese stock market in the late 1980s. Since 1987, when Alan Greenspan took office as chairman of the Federal Reserve and initiated a monetary policy aimed primarily at stabilizing the financial markets, the share of the top 1 percent of incomes in the United States has risen from around 13 percent to nearly 22 percent of total income. Similar developments can be observed in other industrialized countries.

Hayek (1944: 153) argued that “with every grant of complete security to one group the insecurity of the rest necessarily increases. If you guarantee to some a fixed part of a variable cake, the share left to the rest is bound to fluctuate proportionally more than the size of the whole.” In recent years, the increasing concentration of income at higher income levels has mainly worked through the following channels to the (relative) lower income of the middle or lower classes.

FIGURE 10
 STOCK PRICES (U.S., JAPAN) AND U.S.
 INCOME DISTRIBUTION



NOTES: NYSE index, 2010 = 100; Nikkei, 2010 = 50.

SOURCE: OECD, Main Economic Indicators, The World Top Incomes Database.

An asymmetric monetary policy geared toward stabilizing asset prices amounts to a guarantee of security for high-income groups if they hold disproportionately large shares of the total assets. In this scenario, middle- and lower-income groups have to bear the risk of boom-and-crisis cycles in the financial markets.¹⁶ If average growth is low, zero, or negative, then the absolute income gains of higher-income groups must be associated with absolute income losses among middle- and lower-income groups. The redistributive channels are manifold and often arbitrary. Among other things, the returns on low-risk investments such as fixed-income savings, which are often held by the middle class, are lowered toward zero nominally—and into negative territory in real terms.¹⁷ Figure 11 shows for Japan how, following the implementation of ZIRP, domestic incomes from capital gains, rentals, and dividends dropped sharply after the Japanese bubbles burst at the end of 1989, which indicates financial repression.¹⁸

In addition, real wage repression can occur when a crisis undermines the bargaining power of employees. Since the financial crises (and growing public expenditure triggered by exuberant boom phases in the financial markets) drive national debt upward, reduced spending flexibility during a crisis puts pressure on wages in the public sector. The signalling effects of public wage agreements and gloomy business expectations cause public austerity to be followed by wage moderation in the private sector. Wages are driven down especially in those segments of the labor market where qualifications and bargaining power are low.

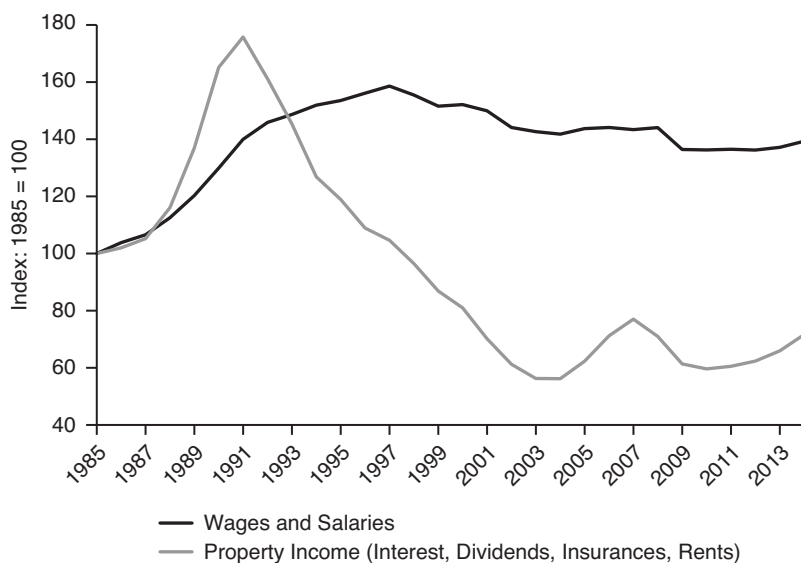
As shown in Figure 11, the average real wage level has fallen steadily since the Japanese financial market crisis (1998). In Europe, too, financial repression and real wage repression—the latter currently with the exception of Germany, where real wages have up to

¹⁶Under certain circumstances, middle-income groups tend to hold more low-risk financial assets, because they perceive investments in the asset markets to be high risk.

¹⁷For more on financial repression, see Hoffmann and Zemanek (2012).

¹⁸Income components in Figure 11 refer to domestic variables, and as such this does not include comparatively high-risk investments abroad (which are riskier due to exchange rate risks for example), such as in the U.S. stock market. As a result, some Japanese investors may have generated high returns through financial investments abroad.

FIGURE 11
REAL WAGE AND FACTOR INCOME IN JAPAN



SOURCE: Economic and Social Research Institute, Cabinet Office, Government of Japan.

the outbreak of the European financial and debt crisis stagnated—are increasingly becoming a reality for large parts of their societies. Real wage repression and financial repression, in turn, can be seen as important determinants of weak private demand among a large proportion of the population, which is anticipated by enterprises by lower investment.¹⁹ As a result, the negative redistributive and real wage effects widely ascribed to consumer price inflation are also achieved indirectly, without consumer price inflation, via boom-and-crisis cycles in the financial markets.

Hysteresis and Growth Effects

An asymmetric monetary policy has self-reinforcing effects when it favors an increase in government debt, thereby undermining the independence of monetary policy. In addition, inflation targets

¹⁹The negative demand effect of declining real incomes is partially offset by declining saving rates of households, particularly the younger generations.

become less effective if additionally created central bank liquidity flows mainly into the financial sector. This may in turn further promote the creeping nationalization of the banking and corporate sector, which can dampen productivity and growth in the long term.

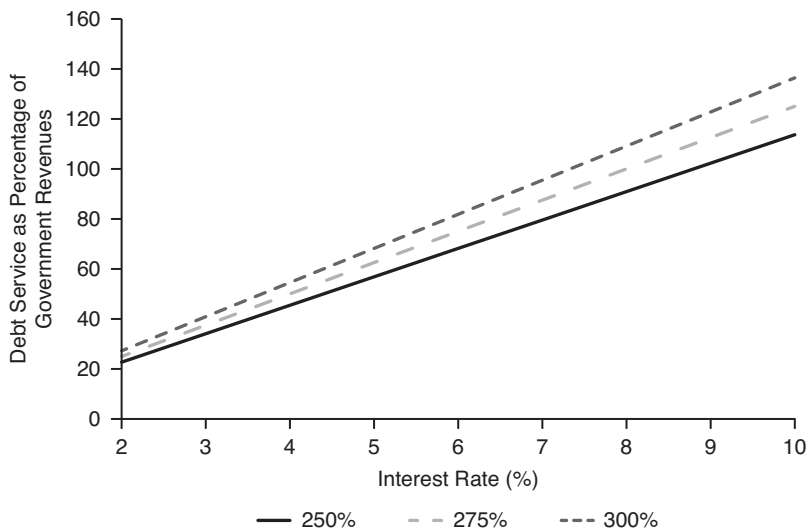
Hysteresis Effects Due to Rising National Debt

An asymmetric monetary policy can be self-reinforcing if it encourages an increase in national debt. There are two transmission channels. First, an increase in tax revenues during boom periods in the financial markets can encourage increased public spending, for example on more social benefits, higher wages for workers in the public sector, or ambitious construction projects. If the bubble bursts, these additional expenditure commitments are difficult to revise—despite falling tax revenues. Second, additional costs usually arise during a crisis as a result of the recapitalization of financial institutions and countercyclical stimulus packages. Structurally declining interest rates on government bonds make a higher debt level possible, because the burden of interest for any given stock of government debt shrinks.

However, the more national debt has grown, the stronger the pressure is on central banks to keep interest rates low. This can result in persistence of ZIRP and QE, because of high government debt levels. Once an interest rate of (almost) zero has been reached and government debt is at a record level, it is politically difficult to raise the central bank interest rate. Indeed, the restrictive monetary policy stance would potentiate the burden of interest on the revolving debt stocks for three reasons: (1) the negative economic impact of a tightening of monetary policy would lead to further tax revenue losses, (2) new stimulus packages would be needed to counteract a rise in unemployment, and (3) new instability in the financial sector would force further recapitalizations or the nationalization of banks. All the effects together would drive up even further not only the level of government debt, but also the risk premiums on government bonds.

Figure 12 shows a simulation of the interest burdens of a tightening of monetary policy for the Japanese government budget. Since the Japanese bubble burst in 1989, national debt as a share of GDP has risen from 60 percent to roughly 250 percent. The simulation assumes a current average interest rate of 2 percent on Japanese government bonds. At this level of interest, a good 20 percent of

FIGURE 12
SIMULATION OF INTEREST RATE PAYMENTS OF THE
JAPANESE CENTRAL GOVERNMENT



SOURCES: International Monetary Fund via Datastream; authors' calculations.

Japanese government income is spent on interest payments. The simulation is based on the assumption that an end to the unconventional monetary policy would lead to an increase in interest rates on government bonds.

If the returns on government bonds were to rise to an average of 4 percent, then 40 percent of the national budget would need to be spent on debt service. This would severely restrict the Japanese state's financial clout. If they were to rise to more than 5 percent, which was far from unheard of in the 1980s, the Japanese government would be financially incapacitated. If a tightening of monetary policy were to cause national debt to rise further still, this would also result in a further multiplication of interest burdens. A scenario where the government would have to spend 80 percent of its budget on interest burdens (debt levels at 300 percent of GDP and an average interest rate of 6 percent on government bonds) is not unrealistic. It is therefore not surprising that Japan's Abe administration has kept a tight rein on the once independent central bank under Governor Kuroda.

But even an independent central bank may have little incentive to raise interest rates with no pressure coming from the government. Many central banks have accumulated risky assets on their balance sheets as a result of unconventional monetary policy (e.g., by purchasing asset-backed securities and government bonds). If interest rate increases meant that value adjustments became necessary on the securities held on their balance sheets, this would deplete their own equity. The central bank would have to rely on recapitalization by the state, in turn undermining its financial independence. This is currently most obvious in Japan, where the bubble burst earlier and national debt has risen to the highest level among industrialized countries.

Hysteresis Effects Due to Interrupted Monetary Policy Transmission

The established model generations assume that monetary policy affects the price level by changing interest rates and steering expectations (Taylor 1993, Woodford 2003).²⁰ Following Woodford (2003), by applying a rule-based monetary policy in a fiat money system it is possible to achieve a specific inflation target without paying particular attention to monetary aggregates. With this in mind, an increasing number of central banks have established specific inflation targets, communicated their monetary policy decisions to the financial markets, and paid ever-decreasing attention to monetary aggregates. One key assumption for this model is that in the medium term monetary policy decisions are reflected in changes in the measured rates of inflation.

However, the relationship between short-term interest rates and inflation and, once rates have reached zero, the relationship between (the communication of) long-term low interest rates and an expansion of central bank balance sheets by purchasing (long-term) assets, does not have the assumed stability. The policy of low interest rates

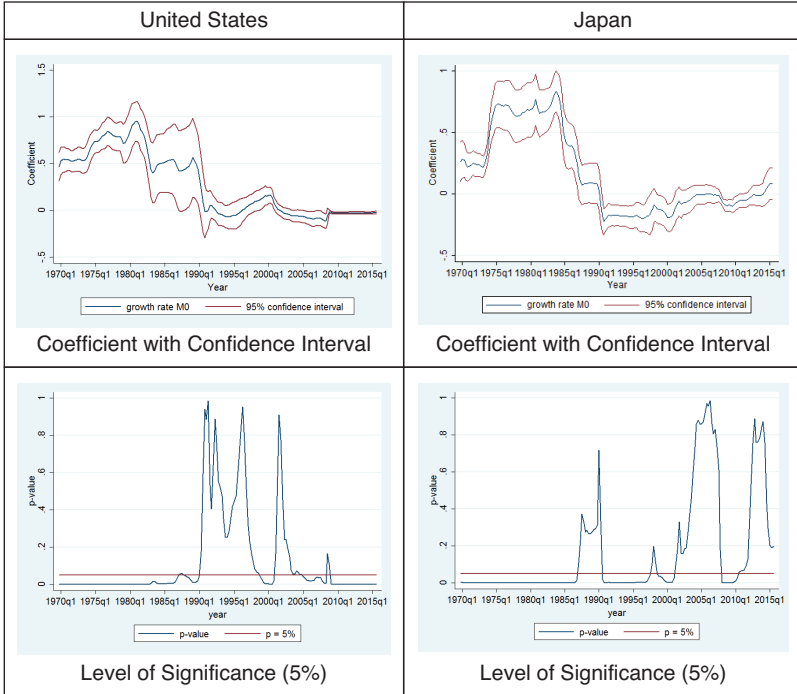
²⁰Monetary policy according to Woodford (2003) is similar to Wicksell ([1898] 2005) in that interest rates are used to control inflation. Woodford (2003) calls his models “neo-Wicksellian.” However, one considerable difference is that the model framework by Woodford (2003) does not require monetary aggregates, whereas according to Wicksell ([1898] 2005) these play an important role in the transmission of changes in interest rates to inflation via credit creation. For a detailed analysis of the similarities and differences between Wicksell and Woodford, see Barbaroux (2007).

also persists because, given the changing monetary policy transmission, interest rate cuts and the expansion of central bank balance sheets through the purchase of assets no longer go along with an increase in consumer price inflation.

The theoretical basis of Figure 13 is the quantity equation that links an increase in the monetary base beyond growth in the quantity of goods with an increase in the price level.²¹ Based on backward-looking 10-year windows, it estimates rolling coefficients that model the effect of growth in the monetary base (minus real growth) on inflation. Sufficiently long time series are available for the United States and Japan. For the United States, it can be observed that in the 1970s the assumed relationship between the monetary base and consumer price inflation is positive and statistically significant. This relationship has become less pronounced in the course of the 1980s and is no longer detectable from the 1990s. By contrast, a statistically significant negative relationship seems to emerge from the turn of the millennium. This could mean that the expansion of the Federal Reserve's balance sheet leads to a drop in the inflation rate.

²¹According to Friedman (1970: 24), "Inflation is always and everywhere monetary phenomenon." He assumes a stable long-term negative relationship between the key interest rate and the monetary base. If the central bank extends its monetary base by purchasing assets at a money market interest rate above zero, the volume of central bank money offered to the banking sector increases. From a theoretical perspective, if there is a constant demand for money the interest rate must fall in order to meet the condition of equilibrium in the money market. In monetary policy practice, the monetary base consists of several components (autonomous factors, standing facilities, and open market operations), of which only open market operations are controlled directly by the central bank. The reserves of commercial banks at the central bank are largely independent of interest rates in the short term. What are decisive are the demand from banks for minimum reserves and the holding of excess reserves, which depend on factors such as the uncertainty of payment flows or characteristics of the payment system. However, in the medium term the economic cycle plays a role in the development of the monetary base, if it is associated with a change in bank lending. If for example the central bank lowers the interest rate (as an operational objective of monetary policy), which leads to increased bank lending under normal conditions, then the demand for the monetary base provided by the central bank increases. Since the money market interest rate has reached zero in many industrialized nations, the monetary base, or the size of a central bank's balance sheet, has become a direct instrument of monetary policy. Monetary policy is based on expanding the central bank balance sheet by purchasing (all kind of) assets. The money market interest rate is close to zero, or even below zero, while the purchase of assets as part of the unconventional monetary policy puts pressure on interest rates for long-term investments, including government bonds.

FIGURE 13
 ROLLING COEFFICIENTS FOR THE EFFECT OF THE
 MONETARY BASE ON INFLATION



NOTES: First row: coefficients of rolling regression of inflation (backward weighted 10-year window) on the growth rate of the monetary base (minus real growth); second row: p-value of the rolling regression.
 SOURCE: International Monetary Fund, quarterly data.

For Japan too, there is a positive statistically significant relationship between growth in the monetary base and inflation during the 1960s and 1970s. This relationship collapses in the 1980s. After that, the correlation between the monetary base and inflation is negative, sometimes even at a statistically significant level.

Specifically, given the stark balance sheet expansions of the Fed since October 2008, we would have expected increases in bank lending and inflation. However, the Fed undermines its policy with interest on reserves held at the central bank (see Selgin 2015) as well as macroprudential regulation (Dorn 2015). When reserves are held at the central bank and do not increase M2, inflation is unlikely to

pick up. Further, if unconventional monetary policy causes a rise in asset prices, wealth effects can cause a rising demand for consumer goods because some people feel richer. However, there is a delay in the increase in consumer prices, meaning there is a time lag before the inflation target is reached. If redistributive effects mostly cause a rise in demand for mainly luxury goods, which are not included or are highly underrepresented in the predefined consumer basket, then substitutions between the various groups of products result in more inflation. The monetary policy transmission toward higher inflation is, however, delayed even further.

It is possible that the relationship between the monetary base and inflation is delayed to an extent that inflation will not rise noticeably until a considerable bubble has already built up in one or another segment of the asset markets. If the central bank then raises its interest rate in an effort to curb the looming inflation, the bubble will burst. The outbreak of the crisis dampens the risks of inflation once more, while this expansionary monetary policy crisis therapy sows the seeds of new asset price bubbles.

Even more, unconventional monetary policy can lead to disinflation—as measured in the usual consumer price indices. There are at least four reasons for this. First, in many countries central bank interest rate cuts and the expansion of the monetary base were/are often associated with excesses in real estate markets. Boom periods in real estate markets are usually accompanied by booms in the construction industry, as the demand for real estate increases. The impacts on consumer price indices are low. Although prices for new rentals rise, housing market regulations dampen any transmission from rising real estate prices to average rental rates. The construction boom does however create additional capacity, which dampens rental rates in the long term after the bubble bursts.

Second, low interest rates are in many cases paired with boom phases in the stock markets, which makes it easier to raise capital. Bearing in mind the significant increase in global competition following the entry of China and many Central and Eastern European countries into the world economy, the declining cost of capital is likely to have contributed to price reductions in the product markets. This has contributed to low consumer price inflation.

Third, financial institutions can use the additional liquidity to purchase government bonds, meaning government spending will

continue to grow. A shift in demand from private to public is not reflected by the established consumer price indices.

Fourth, the distributional effects of boom-and-crisis cycles in the financial markets indirectly bring about income repression for major parts of the population. This dampens consumption among those sections of the population whose consumption habits are modelled in the consumer price indices of central banks.

Growth Effects of Persistently Expansionary Monetary Policy

According to the neoclassical growth theory, growth is explained by the accumulation of capital toward a long-term equilibrium between investment and depreciation (steady-state economy). The steady state is based on the assumption of a declining marginal efficiency of capital when the stock of capital increases (Solow 1956, Swan 1956). Only through innovation and technological progress, which can also be interpreted as increasing productivity, can growth be positive in the long term (Solow 1957). In this context, an asymmetric monetary policy can affect growth dynamics if it has a negative impact on innovation and productivity.

The implicit nationalization of the banking sector causes productivity gains to fall in the corporate sector (zombie firms as described by Caballero, Hoshi, and Kashyap 2008; “evergreening” according to Peek and Rosengren 2005; and “soft budget constraints” according to Kornai 1986). Leibenstein (1966) regards incentives and motivation as major factors in a concept of efficiency which goes beyond allocative efficiency (assuming constant production costs in different types of markets such as polypolies and monopolies). He assumes that businesses do not realize all possible efficiency gains and that production costs rise when competition is limited.

A form of Leibenstein’s (1966) X-inefficiency can arise when asymmetrical monetary policy results in the creation and cementation of structural distortions.²² Liquidity and loans are provided independently of efficiency criteria, causing the average productivity of zombie firms supported by zombie banks to decline. Loan provision to new dynamic enterprises becomes more restricted. A reduced pace of innovation, which according to Hayek (1968) is triggered by

²²On the impact of credit booms on the allocation of labor and productivity dynamics, see also Borio et al. (2015).

lower levels of competition, may have an equally negative impact on productivity.

By shifting and tying resources to sectors with low or negative productivity gains, in the context of the Solow-Swan model a negative allocative effect is created which results from declining average productivity (defined as output per unit of labor). From the perspective of companies, average costs will rise, *ceteris paribus*. At the macroeconomic level, fewer goods and services are produced with a constant amount of labor. Since declining output also entails a decrease in savings per worker, this results in an additional negative growth effect because households make fewer savings available for investment.

Another determinant favoring lower growth is declining household savings and, coupled with this, declining investments, which result from reduced incentives for people to save. The transmission channel from monetary policy toward reduced savings activity is financial repression, which drives down returns on low-risk investments.²³ Real household savings fall, meaning real investments also fall and, in turn, production opportunities increase less—or even decrease, depending on the level of depreciation.²⁴ Once depreciations exceed gross investment, the result is a downward spiral of declining returns on capital, households saving less, declining investments, and a declining output. The foundation of prosperity dwindles.

Economic Policy Implications

We have argued that unconventional monetary policies, especially ultra-low interest rates, in the large advanced countries can discourage investment and lead to adverse distributional effects. We suggest that both factors are reflected in declining economic growth and political dissatisfaction of increasing shares of the societies. In many countries, this process has led to growing political polarization. Consequently, policies that aim at curtailing the negative side-effects of unconventional monetary policy—such as minimum wages, financial regulation, rent controls, and taxation of

²³Following Rothbard (1962), the monetary marginal returns of capital, which are defined as a discounted monetary product of capital, decline.

²⁴Similar reasoning can be found in McKinnon (1973) and Shaw (1973), who identified financial repression as a major obstacle to growth in developing countries.

higher income classes—add further distortions and an additional drag on growth.

Our interpretation of the current secular stagnation as the outcome—and not the origin—of the ultra-low interest rates is in line with Hayek (1929, 1937, 1944). He described the events leading up to the Great Depression and the following stagnation as an outcome of ultra-low interest rates, intervention spirals, and creeping nationalization in response to the crisis. To stop the resulting vicious circle of policy interventions and declining growth a timely exit from the ultra-low interest rates is necessary to reconstitute the allocation and signalling function of the interest rate as well as the principle of liability in financial markets. By gradually, and irrevocably, raising interest rates, growth could be restored via (at least) five transmission channels.

First, risk would be priced based on market forces, and incentives for financial market speculation would be reduced. The resulting cleansing process would free capital and labor for real investment, previously bound in sectors with low productivity. The increasing interest rates would provide an incentive for more household saving to finance growing investment. The marginal and average efficiency of investments would increase again. Aggregated saving and investments as well as innovation would be strengthened.

Second, growing debt-servicing costs would force governments to consolidate their spending by pushing forward structural reforms. Parts of the public economic activity would have to be privatized, which would likely contribute to an increasing average productivity of previously public expenditure. By substituting public consumption and investment by private investment, the average efficiency of investment should increase.

Third, rising interest rates and fiscal consolidation would force banks to restore their traditional business model. The banking sector would return to its very task to finance investment projects with the highest expected returns (instead of buying government bonds). This would provide incentives to (large) enterprises to come up with more innovative ideas and investment projects. A higher degree of X-efficiency would be reached.

Fourth, productivity gains would allow real wages to grow again. This would be even more the case for the middle- and low-income classes if the redistribution effects of boom and bust in financial markets would be eliminated. A growing purchasing power of broad

parts of the society would help fully use the newly created capacities. Growing income levels would contribute to higher tax revenues for the state, which could be used to reduce debt.

Fifth, and probably most important, political polarization would be contained. The political pressure toward regulation, price and rent controls, and redistribution of wealth would be eased. A higher degree of economic freedom would help—in the spirit of Hayek (1944, 1968)—to create sustainable growth and to secure welfare for all parts of the society. To which extent the exit from ultra-low interest rates is politically feasible or desirable hinges on the awareness of the electorate about the negative implications of this very policy. This article suggests that an end of ultra-low interest rate policies is a prerequisite for a return to a sustainable growth path.

References

- Adrian, T., and Shin, H. S. (2008) “Liquidity, Monetary Policy, and Financial Cycles.” *Current Issues in Economics and Finance* 14 (1): 1–7.
- Akerlof, G. (1970) “The Market for Lemons: Quality Uncertainty and the Market Mechanism.” *Quarterly Journal of Economics* 84 (3): 488–500.
- Barbaroux, N. (2007) “Woodford and Wicksell: A Cashless Economy or a Moneyless Economy Framework?” Paper presented at the History of Economics Society 2007 Annual Conference, Fairfax, Va. (June).
- Barnett, A.; Chiu, A.; Franklin, J.; and Sebastián-Barriel, M. (2014) “The Productivity Puzzle: A Firm-Level Investigation into Employment Behavior and Resource Allocation over the Crisis.” Bank of England Working Paper No. 495.
- Bernanke, B. (2005) “The Global Saving Glut and the U.S. Current Account Deficit.” Sandridge Lecture, Virginia Association of Economists, Richmond, Va.
- _____. (2014) “The Federal Reserve: Looking Back, Looking Forward.” Speech at the Annual Meeting of the American Economic Association, Philadelphia (January 3).
- Bernanke B. S.; Gertler M.; and Gilchrist S. (1996) “The Financial Accelerator and the Flight to Quality.” *Review of Economics and Statistics* 78: 1–15.

- Blinder, A., and Reis, R. (2005) “Understanding the Greenspan Standard.” Princeton University Department of Economics Working Paper No. 88.
- Borio, C.; Kharroubi, E.; Upper, C.; and Zampolli, E. (2015) “Labor Reallocation and Productivity Dynamics: Financial Causes, Real Consequences.” BIS Working Paper No. 534 (December).
- Brunnermeier, M., and Schnabel, I. (2014) “Bubbles and Central Banks: Historical Perspectives.” Unpublished manuscript.
- Buiter, W. H. (2010) “Reversing Unconventional Monetary Policy: Technical and Political Considerations.” In M. Balling, J. M. Berk, and M. O. Strauss-Kahn (eds.), *The Quest for Stability: The Macro View*, 23–43. Vienna: SUERF, The European Money and Finance Forum.
- Caballero, R.; Hoshi, T.; and Kashyap, A. (2008) “Zombie Lending and Depressed Restructuring in Japan.” *American Economic Review* 98 (5): 1943–77.
- Cantillon, R. (1931) *Abhandlung über die Natur des Handels im allgemeine (Treatise on the Nature of Trade in General)*. Jena, Germany: Fischer.
- Cardarelli, R., and Lusinyan, L. (2015) “U.S. Total Factor Productivity Slowdown: Evidence from the U.S.” IMF Working Paper No. 15/116.
- De Grauwe, P. (2011) “Monetary Policy and Animal Spirits.” *Economic Theory* 47 (2–3): 423–57.
- Dorn, J. A. (2015) “The Pitfalls of the Federal Reserve’s Zero Interest Rate Policy.” *Caixin Online* (October 15).
- Draghi, M. (2014) “The Path to Recovery and the ECB’s Role.” Speech at the Symposium on Financial Stability and the Role of Central Banks, organized by the Deutsche Bundesbank, Frankfurt am Main (February 27).
- Friedman, M. (1970) “The Counter-Revolution in Monetary Theory.” Wincott Memorial Lecture. Institute of Economic Affairs Occasional Paper No. 33.
- Gopinath, G.; Kalemli-Özkan, Ş.; Karabarbounis, L.; and Villegas-Sanchez, C. (2015) “Capital Allocation and Productivity in South Europe.” Unpublished manuscript.
- Hayek, F. A. ([1929] 1976) *Geldtheorie und Konjunkturtheorie*. Reprint. Salzburg: Philosophia Verlag.
- _____. ([1937] 1989) *Monetary Nationalism and International Stability*. Reprint. New York: Augustus M. Kelley.

- _____ ([1944] 2007) *The Road to Serfdom*. Text and Documents, In B. Caldwell (ed.), *The Collected Works of F. A. Hayek*, Vol. 2. University of Chicago Press.
- _____ ([1968] 1996) “Der Wettbewerb als Entdeckungsverfahren.” Reprint. In Hrsg: K. Leube: *Die österreichische Schule der Nationalökonomie*, 119–37. Band II: von Hayek bis White. Vienna: Manz’sche Verlags- und Universitätsbuchhandlung.
- Hoffmann, A. (2014) “Zero-Interest Rate Policy and Unintended Consequences in Emerging Markets.” *The World Economy* 37: 1367–87.
- Hoffmann, A., and Schnabl, G. (2008) “Monetary Policy, Vagabonding Liquidity and Bursting Bubbles in New and Emerging Markets: An Overinvestment View.” *The World Economy* 31: 1226–52.
- _____ (2011) “A Vicious Cycle of Manias, Crises and Asymmetric Policy Responses: An Overinvestment View.” *The World Economy* 34: 382–403.
- _____ (2014) “Monetary Policy in Large Industrialized Countries, Emerging Market Credit Cycles, and Feedback Effects.” CESifo Working Paper No. 4723.
- Hoffmann, A., and Zemanek, H. (2012) “Financial Repression and Debt Liquidation in the U.S. and the Euro Area.” *Intereconomics: Review of European Economic Policy* 47 (6): 344–51.
- Howden, D., and Salerno, J. T., eds. (2014) *The Fed at One Hundred: a Critical View on the Federal Reserve System*. Berlin: Springer.
- Ishikawa, D., and Tsutsui, Y. (2005) “Has a Credit Crunch Occurred in Japan?” RIETI Discussion Paper No. 06-E-012.
- Jorda, O.; Schularick, M.; and Taylor, A. (2015) “Betting the House.” *Journal of International Economics* 96 (1): 2–18.
- Keynes, J. M. (1936) *The General Theory of Employment, Interest, and Money*. London: Macmillan.
- Kiyotaki, N., and Moore, J. (1997) “Credit Cycles.” *Journal of Political Economy* 105 (2): 211–48.
- Koo, R. (2003) *Balance Sheet Recession: Japan’s Struggle with Uncharted Economics and Its Global Implications*. Singapore: John Wiley and Sons.
- Kornai, J. (1986) “The Soft Budget Constraint.” *Kyklos* 39 (1): 3–30.

- Laubach, T., and Williams, J. (2015) “Measuring the Natural Rate of Interest Redux.” Federal Reserve Bank of San Francisco Working Paper No. 2015-16.
- Leibenstein, H. (1966) “Allocative vs. ‘X-Inefficiency.’” *American Economic Review* 56 (3): 392–415.
- Malpass, D. (2015) “Lower Inflation: Another Chance for Government Mistakes.” *Forbes* (February 9).
- Maskin, E., and Xu, C. (2001) “Soft Budget Constraint Theories: From Centralization to the Market.” *Economics of Transition* 9 (1): 1–27.
- McKinnon, R. (1973) *Money and Capital in Economic Development*. Washington: Brookings Institution.
- _____ (2012) “Zero Interest Rates in the United States Provoke World Monetary Instability and Constrict the U.S. Economy.” SIEPR Policy Brief, Stanford (August).
- Mises, L. von (1912) *Die Theorie des Geldes und der Umlaufmittel*. Leipzig, Germany: Duncker und Humblot.
- _____ ([1949] 1998) *Human Action*. 4th ed. Irvington-on-Hudson, N.Y.: Foundation for Economic Education.
- Peek, J., and Rosengren, E. (2005) “Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan.” *American Economic Review* 95 (4): 1144–66.
- Posen, A. (2000) “The Political Economy of Deflationary Monetary Policy.” In R. Mikitani and A. Posen (eds.), *Japan’s Financial Crisis and Its Parallels to U.S. Experience*, 194–208. Washington: Institute for International Economics.
- Quian, Y., and Xu, C. (1998) “Innovation and Bureaucracy under Soft and Hard Budget Constraints.” *Review of Economic Studies* 65 (1): 151–164.
- Rothbard, M. ([1962] 2009) *Man, Economy, and State*. 2nd ed. Auburn, Ala.: Ludwig von Mises Institute.
- Schnabl, G. (2015) “Monetary Policy and Structural Decline: Lessons from Japan for the European Crisis.” *Asian Economic Papers* 14 (1): 124–50.
- Schumpeter, J. ([1911] 1934) *The Theory of Economic Development*. Cambridge, Mass.: Harvard University Press.
- Sekine, T.; Kobayashi, K.; and Saita, Y. (2003) “Forbearance Lending: The Case of Japanese Firms.” *Bank of Japan Institute for Monetary and Economic Studies* 21 (2): 69–92.

- Selgin, G. (2014) "Operation Twist-the-Truth: How the Federal Reserve Misrepresents Its History and Performance." *Cato Journal* 34 (2): 229–63.
- _____ (2015) "Interest on Reserves, Part I." *Cato at Liberty* (December 17). Washington: Cato Institute.
- Selgin, G.; Lastrapes, W.; and White, L. H. (2012) "Has the Fed Been a Failure?" *Journal of Macroeconomics* 34 (3): 569–96.
- Shaw, E. (1973) *Financial Deepening in Economic Development*. New York: Oxford University Press.
- Shimizu, Y. (2000) "Convoy Regulation, Bank Management and the Financial Crisis in Japan." In R. Mikitani and A. Posen (eds.), *Japan's Financial Crisis and Its Parallels to U.S. Experience*, 57–99. Washington: Institute for International Economics.
- Solow, R. (1956) "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70 (1): 65–94.
- _____ (1957) "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39 (2): 312–20.
- Summers, L. (2014) "U.S. Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound." *Business Economics* 49 (2): 65–73.
- Swan, T. (1956) "Economic Growth and Capital Accumulation." *Economic Record* 32 (2): 334–61.
- Taylor, J. (1993) "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39: 195–214.
- _____ (2007) "Housing and Monetary Policy." NBER Working Paper No. 13682.
- Weizsäcker, C. von (2014) "Public Debt and Price Stability." *German Economic Review* 15 (1): 42–61.
- White, W. (2012) "Ultra Easy Monetary Policy and the Law of Unintended Consequences." Federal Reserve Bank of Dallas Working Paper No. 126.
- Wicksell, K. ([1898] 2005) *Geldzins und Güterpreise*. Reprint. Munich, Germany: FinanzBuch Verlag.
- Woodford, M. (2003) *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, N.J.: Princeton University Press.
- Yoshino, N., and Mizoguchi, T. (2010) "The Role of Public Works in the Political Business Cycle and the Instability of the Budget Deficits in Japan." *Asian Economic Papers* 9 (1): 94–112.

THE G20 AND GLOBAL GOVERNANCE

Stephen Kirchner

The Group of 20 sees itself as “the premier forum for international economic cooperation” (G20 2009b). This article examines its evolution and performance, and member countries’ compliance with G20 summit commitments.

The G20 evolved as a response to the shortcomings of its predecessor, the G7/8. Yet its creation allowed member countries to avoid confronting many of the problems that arose out of the earlier forum. The best defense of the G20 is that it is the only institution of its type, but it still consumes scarce political and diplomatic capital, sometimes to the detriment of the policy objectives to which it is notionally committed.

In this article, I compare data on members’ compliance with G20 summit commitments to proxy measures of the quality of domestic policies and institutions. While the proxies predict G20 compliance, it turns out that G20 compliance has no power to predict subsequent changes in domestic policies and institutions. The main implication of this data is that international economic and political cooperation is a symptom, not a cause, of domestic policies and institutions. Improvement in domestic policies makes the best contribution to advancing the G20 agenda, but such improvements do not appear to depend on the G20 process.

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Stephen Kirchner is an economist with the Australian Financial Markets Association and a Senior Fellow at the Fraser Institute. The views expressed are his own. An earlier version of this article was delivered at the Annual General Meeting of the International Council of Securities Associations held in Mumbai, India, from April 15 to 17, 2015.

I also consider some of the G20's major initiatives in relation to economic policy coordination and financial market regulation. The G20 is assumed to invest these initiatives with greater political legitimacy but, on occasion, the leaders' summit process has actually detracted from this legitimacy. Many of these initiatives arguably would have occurred without the overlay of the G20 leaders' or ministerial processes. In Australia's case, hosting the G20 summit in 2014 detracted from domestic political leadership and the government's ability to advance a domestic economic reform agenda.

G7/G8 Precedents and the Origins of the G20

The G20 can trace its origins to the collapse of the Bretton Woods system in the early 1970s. With the demise of fixed exchange rates, new informal arrangements arose through which the major Western economies sought to address international economic issues. The finance ministers of the United States, Britain, France, and Germany (G4) met in the White House library, forming the so-called Library Group in March 1973 (Bradford and Linn 2011: 1). Japan joined the group in September 1973 to form the G5, followed by Italy (1975), Canada (1976), and the EC/EU presidency (1977), forming what ultimately became the Group of Seven (G7) finance ministers.

A parallel grouping of G7 heads of state began meeting annually from 1975. Russia joined the G7 at the invitation of President Clinton to form the G8 leaders' summit from 1998.

From the late 1960s, a literature grounded in formal theoretical models identified potential economic gains from international economic policy coordination (Cooper 1969). Yet this literature lacked historical and institutional context, and proved misleading about the effective scope and potential of such cooperation. The G7 finance ministers' and G7/8 leaders' meetings were subject to major questions about their agenda, representativeness, and effectiveness.

The G7 presided over two major episodes of economic policy coordination in the 1980s. The 1985 Plaza and 1987 Louvre Accords were designed to address episodes of U.S. dollar strength and weakness respectively. The effect of coordinated foreign exchange market intervention on exchange rates, and its economic significance, has long been disputed, starting with the G7's 1983 Report of the Working Group on Exchange Market Intervention

(the Jurgensen Report). The report concluded that “the role of intervention can only be limited” (Truman 2003: 247).

Mina Balamoune (2000) examined the announcement and compliance effects of the G7 summits held between Rambouillet in 1975 and Munich in 1992. She found no evidence for an announcement effect from these summits on a range of economic and financial market variables, implying a lack of credibility for the G7. She also found that compliance with summit announcements was low. Where there was compliance with summit goals, Balamoune found this did not necessarily improve economic performance.

An analysis by Marcel Fratzscher (2009) of 76 G7 communiqués since 1975 found they were able to move G3 exchange rates in the desired direction, especially when supported by foreign exchange market intervention. However, Fratzscher suggests this is due to the G7 successfully identifying, rather than correcting, episodes of currency misalignment. G7 communication did not cause the subsequent realignment of exchange rates.

Another analysis of G7 summit commitments between 1975 and 1989 found a compliance rate of only one-third in relation to economic policy issues more broadly (von Furstenberg and Daniels 1992). A related study found that many of the economic relationships asserted in summit declarations were contestable, demonstrating that the G7’s attempts at policy cooperation were not economically well founded (Daniels 1993).

By the late 1980s, there was growing skepticism about the prospects for effective international economic policy coordination. Stanley Fischer (1987: 4) was representative of this, maintaining that “continued systematic policy coordination on a grand scale among the major economies is unlikely . . . the best that each country can do for other countries is to keep its own economy in shape.” Fischer did hold out the prospect that improved understanding of policy, and growing interdependence, might see greater cooperation in future, “but only in the very long run.” More recently, Jeffrey Frankel (2016: 1) has sought to rehabilitate this theoretical literature “after a 30-year absence,” despite its lack of historical and institutional context.

Issues of international economic cooperation again came to fore with the 1997–98 emerging markets crisis, which went beyond the traditional geographic focus of the G7. In this context, it was thought that a broader grouping was needed to better represent the emerging-market economies that were at the center of the crisis.

The first meeting of G20 finance ministers and central bank governors took place in Berlin in December 1999, at the initiative of the United States, Germany, and Canada. The G20 added Argentina, Australia, Brazil, China, India, Indonesia, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey, and the presidency of the EU to the G7's membership. The managing director of the International Monetary Fund (IMF) and the president of the World Bank were also included. The G20 was given a rotating presidency based on five regional groupings but, like the G7/G8, lacked a permanent secretariat or standing organization. As was the case with its predecessors, this lack of a standing organization has been variously interpreted as both a strength and a weakness.

An important outgrowth of the G20 ministerial meetings was the Financial Stability Forum (FSF), which took on responsibility for coordinating issues in relation to international financial regulation and later became the Financial Stability Board (FSB). However, the meetings addressed a wide range of other issues between 1999 and the onset of the 2008–09 financial crisis, with the G20's agenda inevitably reflecting the issues of the moment.

The period between 1999 and 2008 saw continued debate about how best to expand the G8 and address its shortcomings. A small, often government-funded industry grew up in academia and think tanks to address the issue of global governance; the form and effectiveness of the G7/8 and the G20 ministerial meetings was its major focus. This industry is remarkable for its lack of realism: Alex Brill (2012: 20) writes that “too many experts are unclear on exactly how to improve existing multilateral institutions. Many scholars have critiqued these institutions and proposed strategies for reform, but in reviewing various recommendations, the utter vagueness of much of the advice is striking.”

One concrete idea that did emerge from this industry was that an elevation of the G20 ministerial meetings process was preferable to an expansion or devolution of the G8 leaders' grouping (Bradford and Linn 2011: 7). This was put into action as a result of the financial crisis of 2008–09. Yet, it is noteworthy that in the run-up to the crisis, there was considerable skepticism from close observers about the value of international economic cooperation. Joseph Daniels (2005: 84) was representative, arguing that “global economic stability depends, first and foremost, on good domestic economic policymaking. International policy cooperation, although ever more important

in light of global economic integration, is of second order importance for global stability.” Razeen Sally (2001: 55) put it more bluntly: “Most arguments for global governance are in fact bad economics and even worse political economy.” A key issue is whether the financial crisis of 2008–09 confirmed or demanded a reconsideration of these judgments about the prospects for international economic policy cooperation and coordination.

The G20 Leaders’ Meetings and the Financial Crisis

The onset of the financial crisis in 2008 gave new impetus to international economic cooperation on a broader basis than the G7/G8. The existing G20 ministerial meeting process provided a convenient forum for the organization of a leaders’ summit in Washington on November 15, 2008, which also added the Netherlands and Spain to the G20 meetings.

The elevation of the G20 ministerial process to the status of a leaders’ meeting was an attempt to address some of the long-standing issues around the effectiveness and legitimacy of the G7/8. However, the origins of the G20 in the emerging markets crisis of 1997–98 and the global financial crisis of 2008–09 meant that many of the long-standing issues around the representativeness and effectiveness of the G7/8 and its expansion were avoided. The G20 was seen as a workaround for many of these problems, without actually addressing them other than by expanding on the G7/G8’s membership. As Peter Drysdale and Kemal Dervis (2014: 4) note, “the fact that a finance ministers’ G20 already existed allowed the United States to circumvent debate on inclusion. It was easier to simply invite the government leaders of the existing group of twenty than to try to agree on who should be included. . . . There was no time for such a debate.”

The initial membership of the G20 was an arbitrary selection drawn at the initiative of the United States, Germany, and Canada in 1999, with subsequent additions to the membership made on an ad hoc basis. While the G20 is more representative in a numerical sense, this comes at a cost of commonality of interest and values. The G20 has no natural focus, geographic or otherwise, to bind its members together in the way of the Asia-Pacific Economic Cooperation forum (APEC), and does not have a set of common values, such as NATO does. It also has no membership criteria from which it might derive legitimacy, unlike, for example, the World Trade Organization

(WTO). Brill (2012: 22) notes that applying membership criteria consistent with the stated aims of the G20 would lead to the removal of Argentina, Indonesia, Mexico, and Russia, and the addition of Malaysia, Norway, Singapore, and Switzerland.

The G20 leaders' process thus inherited many of the problems and unresolved issues of its G7/G8 predecessor. The expansion in membership addressed the issue of inclusiveness but multiplied the problems of a lack of common interests and values. Steven Slaughter (2013: 1) writes that "the legitimacy of the G20 is fundamentally uncertain and problematic because the G20's membership and connection to existing forms of multilateralism remain contentious." The extent to which the G20 is an empty vessel is demonstrated by the readiness of policymakers and nonstate actors to project a growing list of issues onto the G20 agenda, leading to the production of insubstantial and often wordy communiqués. Though the G20 is seen as a vehicle for addressing the world's problems, the leaders' meetings bring only a modest amount of new institutional capacity, legitimacy, and political capital to the task. While this conclusion is ultimately a matter of judgment and interpretation, it is nonetheless borne out by a critical review of the G20's achievements to date. As I will argue below, many of the claimed successes of the G20 owe little to its process.

The G20 and Other Multilateral Institutions and Processes

The emergence of the G20 also reflects the limitations of the Bretton Woods institutions, the IMF and the World Bank. Originally designed to manage a world of fixed exchange rates and fund postwar reconstruction, these institutions have survived the demise of the Bretton Woods system. With floating exchange rates largely eliminating the problem of balance-of-payments crises among developed and many developing economies, and with private capital markets readily financing economic development in emerging economies, the IMF and World Bank have struggled to maintain their relevance. Indeed, before being recapitalized during the financial crisis, the IMF faced its own financial crisis because of a lack of demand for its lending on the part of developing economies. The International Financial Institutions Advisory Commission (2000) presided over by Allan Meltzer highlighted many of the problems

with the Bretton Woods institutions, though none of its recommendations were implemented.

The Bretton Woods institutions have sought to maintain their relevance by addressing the postcrisis international economic issues on the G20's agenda. They bring to this task some of the institutional capacity that the G20 lacks. Yet it is notable that reform of the Bretton Woods institutions is itself high on the G20 agenda. As Brill (2012: 24) observes, this creates the problem that the G20 relies for some of its work program on "institutions that are outdated or known to be riddled with problems and pathologies."

Other multilateral processes also offer lessons on what it is reasonable to expect of the G20. The multilateral trade liberalization process under the auspices of the WTO has addressed the legitimacy problem in international organization by demanding unanimous agreement among its membership, but at the cost of narrowing its negotiating agenda (for example, the exclusion of investment) and making little substantive progress since its fifth ministerial meeting in Cancún in 2003, the Bali Agreement on customs procedures in 2013 notwithstanding. The WTO process highlights some of the tradeoffs between inclusiveness, legitimacy, and effectiveness in international economic cooperation. The growing resort to bilateral and regional trade agreements, such as the Trans-Pacific Partnership, has also highlighted the limits of multilateralism in the area of trade liberalization.

It has been suggested that the G20 be given the task of rebooting the WTO's Doha round (Thirwell 2013), though it is notable that this suggestion was also made for the G7 (Daniels 2005: 93). The G7's role in the completion of the Uruguay round of the General Agreement on Tariffs and Trade would serve as a precedent for this, and it would represent an attempt to narrow the G20 agenda and give it more realistic and achievable goals. Yet at the same time, it is unrealistic to expect the G20 to succeed where the G7 and the WTO itself have failed. Indeed, the G20's attempt to put in place a standstill on trade protection at its November 2008 summit is one of its more notable failures. Since 2008, the G20's membership has been responsible for more than half of the thousands of new protectionist measures implemented worldwide, pointing to substantial underreporting of protectionist measures by the WTO in its role as the official monitor of the G20's adherence to its standstill on protection (Global Trade Alert 2014: 4).

This track record makes the G20 an unlikely vehicle for rebooting the multilateral trade liberalization process. As Drysdale and Dervis (2014: 14) note, “International trade has been a poor cousin of global macroeconomic and financial reform on the G20 agenda.” Some analysts have suggested a counterfactual in which the world would have descended into 1930s-style protectionism in the absence of the G20’s “standstill.” Barry Carin and David Short (2013: 9) suggest that “compared with the rampant ‘beggar thy neighborism’ of the 1930s, G20 nations have shown notable restraint.” This is a very low standard against which to assess the success of the G20’s “standstill.” A more straightforward conclusion is that the G20’s commitment was simply irrelevant to national trade policies.

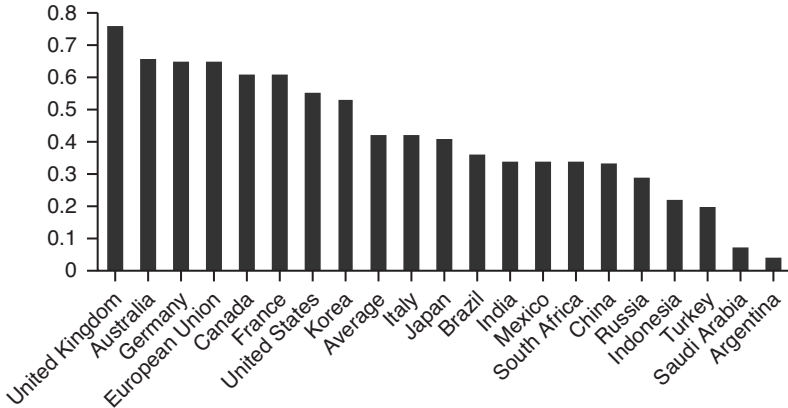
Experience with the G7/G8, Bretton Woods, and other multilateral institutions such as the WTO and the previous G20 ministerial process should have led to more modest expectations of the G20 leaders’ process. The G20 was assumed to have the capacity to succeed where other institutions and processes had often failed. A consistent theme in many discussions of the G20 is the idea that the institution has underperformed relative to its potential. Some of this commentary comes from the same academic think tank industry that was previously preoccupied with the shortcomings of the G20’s predecessors. This projection of previously disappointed aspirations for workable models of global governance onto the relatively new G20 leaders’ process suggests a willful blindness to the G20’s historical, political, and institutional context.

Financial markets have been far more skeptical of the importance of the G20 process. A European Central Bank (ECB) Working Paper by Marco Lo Duca and Livio Stracca (2014: 1) estimated the effects of G20 leaders’ meetings on financial markets, and found that “G20 summits have not had a strong, consistent and durable effect on any of the markets that we consider, suggesting that the information and decision content of G20 summits is of limited relevance for market participants.” Financial market prices give a relatively unbiased and independent assessment of the effectiveness of the G20 process and related policies.

Evaluating the G20’s Commitments

The G20 Information Centre at the University of Toronto monitors compliance with commitments made at the G20 Leaders’

FIGURE 1
G20 OVERALL COMPLIANCE AVERAGE, 2008–13
(SCALE -1 TO $+1$)



SOURCE: G20 Information Centre, University of Toronto.

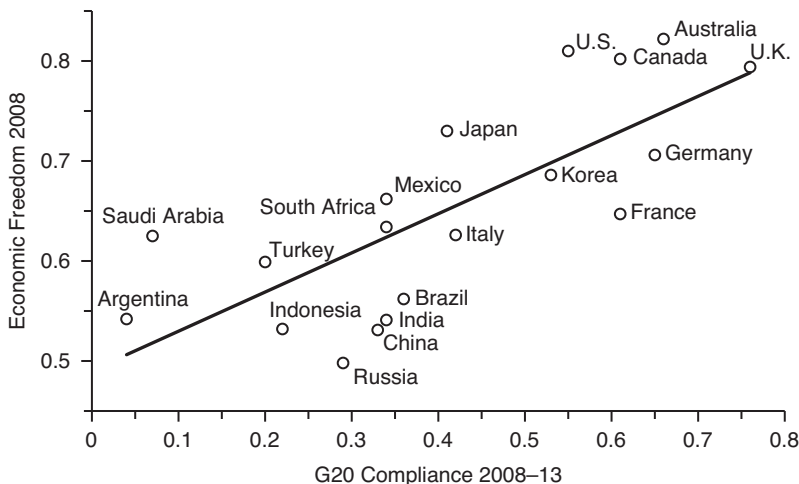
Summit meetings. Figure 1 shows the average scores for G20 countries and the EU between 2008 and 2013 on a scale of $+1$ (compliant) to -1 (noncompliant). A value of zero is assigned to indicate incomplete progress or compliance with a particular commitment.

The data point to a relatively high level of compliance by Anglo-American and most European countries, and Korea, with below-average compliance from the remaining members. Compliance is partly a matter of interpretation, and it is possible to disagree with some of the G20 Information Centre's coding. However, the center's analysis nonetheless represents a comprehensive and relatively independent attempt to monitor G20 compliance.

One approach to evaluating G20 compliance is to consider the relationship between compliance and measures of the quality of domestic institutions and policies, for which the Heritage Foundation's index of economic freedom can serve as a proxy. Plotting individual country average compliance scores against the Heritage Foundation's economic freedom scores for 2008 (divided by 100 for scaling purposes) shows a positive relationship between G20 compliance and the quality of domestic institutions (Figure 2).

Since the 2008 economic freedom scores predate the 2008–13 G20 summit commitments and compliance—actual data used to

FIGURE 2
G20 COMPLIANCE 2008–13 AND
ECONOMIC FREEDOM SCORE 2008



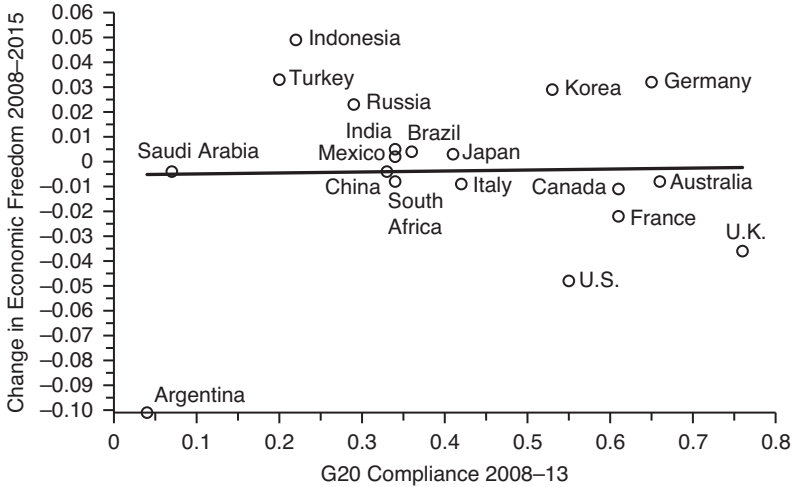
SOURCES: Heritage Foundation; G20 Information Centre.

compile the economic freedom scores typically lagging by around 12 months—we can rule out the possibility that summit compliance caused changes to domestic policies and institutions. On the other hand, the quality of domestic institutions and policies appears to have predictive power for subsequent compliance with G20 summit commitments, which in turn can be taken as a proxy for willingness to engage in international economic-policy cooperation or coordination. An important implication of this data is that the ability to successfully address the issues raised at G20 summits is a function of domestic institutional capacity and quality.

We can examine the possibility that G20 compliance caused improvement to the quality of domestic institutions and policies by comparing average G20 compliance with the change in the economic freedom index between 2008 and 2015 (Figure 3). This allows for a two-year lag in the effects of compliance.

Figure 3 is remarkable for showing no relationship between compliance with G20 summit commitments between 2008 and 2013 and the change in economic freedom over the same period plus a two-year lag. Most countries are clustered around the zero line for the

FIGURE 3
G20 COMPLIANCE 2008–13 AND
CHANGE IN ECONOMIC FREEDOM SCORE 2008–2015



SOURCES: Heritage Foundation; G20 Information Centre.

change in economic freedom while being highly dispersed along the compliance dimension, though it is notable that the major Anglo-American economies saw a reduction in economic freedom despite a high level of G20 compliance.

A similar relationship is found when comparing G20 compliance with the Legatum Institute’s Prosperity Index, which is designed to measure national well-being using a very broad range of measures, and Transparency International’s Corruption Perceptions Index (data available from the author on request).

These data demonstrate that international economic and political cooperation is a symptom, not a cause, of domestic policies and institutions. Domestic policies and institutional settings contribute to advancing the G20’s agenda, but these settings do not appear to depend on the G20 summit process in a measurable way.

Postcrisis Economic Policy Coordination

The strengths and limitations of the G20 process, discussed above, can be further considered by evaluating specific policy areas where

substantive actions have been agreed on and implemented, and have been widely claimed as successes by some observers. A key issue is how important the G20 process has been to actual policy outcomes versus plausible counterfactuals. The IMF has concluded that “evidence to date does not suggest that any of the large countries have made significant adjustments to their economic policies in response to peer pressure” under the G20-mandated mutual assessment of policies (Ostry and Ghosh 2013: 25). A former deputy managing director of the IMF has said that “it is hard to say with certainty that any G20 member has altered its policy plans in the interest of achieving greater policy coherence—and therefore effectiveness—with its G20 partners” (Taylor 2014). The two policy areas considered here are postcrisis macroeconomic policy coordination and financial market regulation.

The 2008 Washington G20 summit communiqué attributed the financial crisis in part to “inconsistent and insufficiently coordinated macroeconomic policies” (G20 2008). If we think of macroeconomic policy as monetary, fiscal, and exchange rate policy, it is far from obvious how greater coordination of these policies could have averted or mitigated the financial crisis, which had its origins in the politicization of housing finance in the United States (Morgenson and Rosner 2012, Acharya et al. 2011, Wallison 2015). The Bretton Woods system of fixed exchange rates provides a historical example of an internationally coordinated system of exchange rates and, occasionally, macroeconomic policies, which lacked the flexibility to accommodate the economic shocks that ultimately brought it undone in the early 1970s.

The evidence for effective postcrisis economic policy coordination is weak. I have already mentioned the data on compliance with the G20’s “standstill” on protection, which shows that despite their commitments, G20 members have contributed substantially to the introduction of new protectionist measures. We can also consider the communiqué from the London G20 Leaders’ Summit on April 2, 2009, which referenced an “unprecedented and concerted fiscal expansion, which will save or create millions of jobs” (G20 2009a). This “concerted fiscal expansion” is cited by some as a major achievement of the G20. For example, Colin Bradford and Johannes Linn (2011: 11) claim that “the major achievement of the London G20 Summit was the confirmation of the \$5 trillion ‘concerted fiscal expansion’ undertaken by G20 governments between

the Washington and London G20 Summits.” Carin and Short (2013: 2 and 11) credit the G20 with “marshalling nearly a trillion dollars to give the global economy some shock absorbers,” through which the G20 “earned an ample supply of legitimacy and effectiveness points.” They even suggest the G20’s response to the financial crisis was “triumphal.”

There is little evidence that these fiscal expansions were attributable to the G20 leaders’ process rather than just being the sum of individual country efforts that would have been undertaken in the absence of that process. The stimulus was not an outcome of the summit itself, having been preannounced by individual countries. For example, Australia’s two discretionary fiscal expansions were announced in October 2008 and February 2009. The counterfactual scenario in which no or less fiscal stimulus would have occurred in the absence of the summit is difficult to accept, given the willingness of G20 governments to adopt discretionary fiscal expansions.

It could be argued that the summit gave some additional political legitimacy to these efforts. Yet the London summit was the scene of a sharp disagreement between the United States and Europeans over what constituted fiscal stimulus. This led the London summit’s U.K. organizers to deemphasize fiscal stimulus as the summit’s main outcome at the time. Sharp disagreements about the appropriate stance of fiscal policy were also a feature of the 2010 G8 and G20 summits in Canada (Bradford and Linn 2011: 10). The 2010 Toronto summit commitments to reduce fiscal deficits by half were abandoned by the time of the St. Petersburg summit in 2013, when they were meant to have been achieved. If anything, these high-profile disagreements over fiscal policy detracted from the political legitimacy of fiscal policy both domestically and internationally, while the failure to meet the G20’s fiscal targets damaged credibility.

The effectiveness of fiscal stimulus at a national level is highly contentious. In the presence of an inflation-targeting central bank, monetary policy will effectively discount fiscal policy actions (DeLong and Olney 2005: 399–400). While some have argued that monetary policy is ineffective when the zero bound on nominal interest rates becomes a binding constraint on official interest rates, the widespread use of quantitative operating instruments for monetary policy in the wake of the crisis demonstrates that this is not the case. In this context, “estimates of [positive] fiscal multipliers become little more than forecasts of central bank incompetence” (Sumner 2013: 3).

It has also been argued that a concerted fiscal expansion can be effective because the world as a whole is a closed economy. However, this merely relocates the fiscal crowding-out effects from the national to the internationally relative change in budget balance as a share of the economy. Given a concerted global fiscal expansion, those economies with floating exchange rates that expand fiscal policy the most will see relatively greater upward pressure on their exchange rate and downward pressure on net exports, all else being equal. Economies with more modest fiscal expansions or larger fiscal contractions will benefit by creating more room for domestic monetary policy and putting downward pressure on their exchange rate. Fiscal policy may still be effective in the context of a fixed exchange rate regime, but this comes at the cost of an independent monetary policy and the benefits of a flexible exchange rate as an economic shock absorber.

The November 2010 Seoul summit saw a U.S. proposal to impose a ± 4 percent of GDP floor and ceiling on current account balances to address the alleged problem of global “imbalances.” The role of current account imbalances in pre- and postcrisis macroeconomic outcomes is controversial and reflects a debate that has persisted since Ben Bernanke (2005) raised the issue of a “glut” of global saving. The issue also reflects long-standing tensions between China and the United States over the former’s managed exchange rate regime, an echo of G7 tensions between Japan and the United States over trade balances and the U.S. dollar–Japanese yen exchange rate in the 1980s and 1990s.

The proposal to limit current account imbalances would have constrained the economic growth of countries that import a large amount of capital as a share of GDP, like Australia, and exacerbated the surplus of domestic saving in capital-exporting countries. The mercantilist assumptions underpinning the concern with such imbalances renders the issue tailor-made for raising tensions between governments, and made the G20 summit “look like an exercise in discord rather than international cooperation” (Bradford and Linn 2011: 14), further undermining the international legitimacy of macroeconomic policy coordination efforts.

Current account imbalances are not a macroeconomic problem under floating exchange rates, although they may be symptomatic of the problems associated with fixed exchange rate regimes. They are symptomatic of the global trade in capital that is an important

benefit of globalization (Cooper 2005). To that extent, they do not demand macroeconomic policy coordination, and so the G20's effort to promote "global rebalancing" was economically and politically counterproductive.

The 2014 Brisbane summit saw an attempt by the host country, Australia, to better focus the G20 agenda on concrete strategies for promoting economic growth. This was a close fit with the domestic policy agenda of the Australian government, but also reflected concern that the G20 had lost its way. G20 members were encouraged to develop plans to raise global GDP by 2 percent by 2018, relative to the IMF's October 2013 forecasts assuming the absence of these measures. These policies were then formally assessed by the IMF and OECD in a black-box process, in time for the Brisbane summit, and declared to be capable of achieving the growth objective—but only under the heroic assumption of full and simultaneous implementation from day one. Global economic growth has since fallen short of even the baseline scenario, as G20 governments implemented less than half of the measures required to meet the Brisbane growth target (Greber 2016a).

While raising global economic growth rates is a worthy objective, the Australian approach only served to highlight the extent to which global growth is a function of national economic policies rather than international policy coordination. Most governments are already committed to promoting economic growth, subject to domestic policy constraints that international policy cooperation and coordination do little to alleviate.

A review of Australia's proposed growth strategy invites skepticism about its origins in the G20 process (Australian Government 2014). The proposed strategy was an inventory of long-standing policy commitments otherwise unrelated to the G20 process or the Brisbane summit.

It has been argued by well-informed observers that the effort invested by the Australian treasurer Joe Hockey in pursuing outcomes for the G20's Brisbane summit compromised his ability to implement domestic policy proposals, undermining Australia's contribution to the G20 growth target (Uren 2015). Indeed, the distraction of the G20 summit was arguably a factor in the ousting of Australia's conservative prime minister, Tony Abbott, and of Hockey, his treasurer, in an internal party revolt in September 2015. In seeking to explain their political demise, journalist Phillip Coorey (2015)

observed that “Australia was chair of the G20 when the Coalition was elected and Hockey had no choice but to assume a prominent international role, which increased his burden.” Hockey’s successor as treasurer, Scott Morrison, said he would not attend meetings of the G20, IMF, or World Bank, deputizing to a junior minister “so he could stay home and sell tax reform” (Coorey 2015). More recently, “Australia hasn’t sent a single senior government minister to a top level meeting of finance chiefs in Paris . . . aimed at preparing for the next crisis” (Greber 2016b). Hosting the G20 summit not only harmed the domestic political fortunes of Australia’s political and economic leadership, but also compromised the domestic economic reform agenda that would have supported its G20 commitments.

The Australian approach to the Brisbane summit was hailed by some as a success in refocusing the G20 agenda. However, it did so in a way that undermined the case for international policy coordination by highlighting the importance of domestic policies and national rather than international political leadership.

International Financial Regulation and the Financial Stability Board

The G20 has been most active in presiding over a program of international financial regulatory reform, including in the areas of prudential and liquidity standards, and derivatives markets. A key institutional innovation of the G20 has been the creation of the FSB out of the former FSF, with a mandate to oversee these efforts. Like the G20 leaders’ process, the pre-financial-crisis history of the FSF provides important insights into the prospects for the FSB.

Until the financial crisis, the Basel Committee on Banking Supervision (BCBS) was largely made up of the G7 countries, and focused on the development of international prudential standards among developed economies. The 1988 Basel Accord (Basel I) was adopted in an environment where Japanese financial institutions were placing competitive pressure on U.S. and European institutions, creating demands for a leveling of the international playing field through common standards.

The emerging markets crises of the 1990s saw the G7 create the FSF in 1999, with a view to promoting international financial standards that would take in both developed and emerging

market economies. The FSF's membership included G7 central banks, finance ministries, and prudential regulators; international financial institutions (IFIs) such as the World Bank, IMF, and Bank for International Settlements (BIS); and international standard-setting bodies (SSBs) such as the BCBS and International Organization of Securities Commissions. Its first chair was the general manager of the BIS, and was supported by a small secretariat within that institution. The FSF's membership was subsequently expanded to include Australia, Hong Kong, the Netherlands, Singapore, and Switzerland.

The FSF's agenda largely reflected issues seen to be arising from the emerging markets crises of the 1990s. The FSF developed financial standards that were the subject of a Financial Sector Assessment Program (FSAP) run by the IMF and the World Bank. However, the FSAP was notable for the refusal of some countries to participate, not least the United States. The precrisis FSAP process is estimated to have cost around \$1 billion. Like Eric Helleiner (2010: 286), we may ask, "given the eventual system blow-up, was this time and money well spent?" Howard Davies and David Green (2008: 116) maintain that the FSF was notable for failing "to carve out a distinctive position, integrating the various perspectives of the diverse membership, as was originally hoped."

The first G20 Leaders' Summit in Washington in November 2008 mandated that the FSF and SSBs should expand their membership. At the Pittsburgh summit in September 2009, the G20 approved a charter for the FSB, and at the Cannes summit in November 2011 it called for a strengthening of the FSB's resources and governance through establishment of the FSB on a permanent organizational basis.

In its report to the 2012 Los Cabos summit, the FSB set out a new organizational, governance, and resourcing framework, and an amended charter, including its role in setting standards. The FSB was established as an association under Swiss law on January 28, 2013. Article 23 of the 2012 FSB Charter is notable for stating, "This Charter is not intended to create any legal rights or obligations" (Financial Stability Board 2012).

The institutionalization of the FSB can be seen as an attempt to invest the G20 process with greater organizational capacity in the area of financial regulation, particularly with respect to coordinating the activities of the SSBs. The broader membership of the G20 and

the FSB, relative to the G7-dominated precrisis processes for international regulatory coordination, is seen as giving these efforts greater legitimacy. However, the postcrisis broadening in the membership of the SSBs was perhaps more important in this regard, even if it was an outcome of the G20 process. Greater regulatory coordination could arguably have been achieved through the SSBs without the overlay of the G20 leaders' or even the G20 ministerial process.

The United States has been an important driver of the FSB process. One interpretation of these efforts is that the United States does not want to be competitively disadvantaged by tightening financial regulation at home. The FSB is thus seen as a vehicle to level the international regulatory playing field (Helleiner 2010: 285). International regulatory policy coordination is a mechanism for managing the competitive implications of domestic regulatory change.

The postcrisis regulatory agenda driven by the FSB has been blamed for a reduction in liquidity in some financial markets and an increase in asset price volatility (PricewaterhouseCoopers 2015). Reforms to derivatives markets have for the most part simply redistributed rather than reduced systemic risks (Pirrong 2010). There are also risks in imposing common regulatory standards, not least the creation of new systemic risks due to a lack of regulatory diversity and competition. Indeed, this should be viewed as an important lesson from the 2008–09 financial crisis. As Dani Rodrik (2009) has argued:

The world economy will be far more stable and prosperous with a thin veneer of international cooperation superimposed on strong national regulations than with attempts to construct a bold global regulatory and supervisory framework. The risk we run is that pursuing an ambitious goal will detract us from something that is more desirable and more easily attained. . . . Global financial regulation is neither feasible, nor prudent, nor desirable.

Conclusion

As an institution for global governance, the G20 inherits many of the problems of its predecessor, the G7/8. The original sin of the G20's creation was to avoid confronting the important issues surrounding G7/8 expansion and the need to reform existing Bretton Woods institutions, which were highlighted by the 2000 Meltzer Commission. An expanded membership was achieved at the expense

of developing a common set of interests and values (or membership criteria) that could have bound the members of the G20 together and served as a source of international political legitimacy.

The best defense of the G20 is that it is the only major forum for global governance; the G20's role in this regard is the outcome of deliberate policy choices and decisions about how to invest scarce political, diplomatic, and other capital. This investment has been undertaken without sufficient attention to the lessons presented by the history of multilateral institutions and processes, which were well known to international governance scholars, but ignored in the rush to embrace the G20 leaders' process.

There is an air of unreality in much of the academic commentary on the G20. The G20 process is often treated as an end in itself, where analysis would be better focused on the G20's substantive achievements and their relationship to that process. Even government-funded think tanks most supportive of the G20 exhibit a notably fading enthusiasm. One journalist compared a Lowy Institute for International Policy forum on the G20 to a "rarefied session of Alcoholics Anonymous," noting that the "hard bitten, one-time true believer . . . Barry Carin . . . simply describes the institution as a 'dead forum walking'" (Earl 2015). The disappointment is palpable, but should not have come as a surprise to well-informed observers.

References

- Acharya, V. V.; Richardson, M.; van Nieuwerburgh, S.; and White, L. J. (2011) *Guaranteed to Fail: Fannie Mae, Freddie Mac, and the Debacle of Mortgage Finance*. Princeton, N.J.: Princeton University Press.
- Australian Government (2014) "Comprehensive Growth Strategy: Australia." G20. Available at https://g20.org.tr/wp-content/uploads/2014/12/g20_comprehensive_growth_strategy_australia.pdf.
- Baliamoune, M. (2000) "Economics of Summitry: An Empirical Assessment of the Economic Effects of Summits." *Empirica* 27: 295–314.
- Bernanke, B. (2005) "The Global Saving Glut and the U.S. Current Account Deficit." Richmond, Va.: Virginia Association of Economics Sandridge Lecture (April 14).

- Bradford, C., and Linn, J. (2011) “A History of G20 Summits: The Evolving Dynamic of Global Leadership.” *Journal of Globalization and Development* 2 (2): 1–23.
- Brill, A. (2012) “A Modest Approach for Effective Multilateral Institutions.” *SAIS Review* 32 (2): 17–25.
- Carin, B., and Short, D. (2013) “The G-20 as a Lever for Progress.” Policy Analysis Brief. Muscatine, Iowa: The Stanley Foundation.
- Cooper, R. (1969) “Macroeconomic Policy Adjustment in Interdependent Economies.” *Quarterly Journal of Economics* 83: 1–24.
- _____ (2005) “Living with Global Imbalances: A Contrarian View.” Policy Briefs in International Economics PB05-3. Washington: Institute for International Economics.
- Coorey, P. (2015) “Treasurer Scott Morrison Wants to Unlock Retirees’ Home Wealth.” *Financial Review* (September 23).
- Daniels, J. (1993) *The Meaning and Reliability of Economic Summit Undertakings, 1975–1989*. New York: Garland Press.
- _____ (2005) “The Significance of the Economic Summits.” In W. Hoppenstedt, R. Pruessen, and O. Rathkolb (eds.), *Global Management*. Munster: Center for Global Economic Studies.
- Davies, H., and Green, D. (2008) *Global Financial Regulation: The Essential Guide*. Cambridge: Polity Press.
- DeLong, B., and Olney, M. (2005) *Macroeconomics*. New York: McGraw-Hill.
- Drysdale, P., and Dervis, K. (2014) *The G-20 Summit at Five: Time for Strategic Leadership*. Washington: Brookings Institution Press.
- Earl, G. (2015) “My G20-Soaked Season in the Sun.” *Financial Review* (June 5).
- Financial Stability Board (2012) “Charter of the Financial Stability Board.” Available at www.fsb.org/wp-content/uploads/r_120809.pdf.
- Fischer, S. (1987) “International Macroeconomic Policy Coordination.” NBER Working Paper No. 2244. Cambridge, Mass.: National Bureau of Economic Research.
- Frankel, J. A. (2016) “International Coordination.” NBER Working Paper No. 21878. Cambridge, Mass.: National Bureau of Economic Research.
- Fratzscher, M. (2009) “How Successful Is the G7 in Managing Exchange Rates?” *Journal of International Economics* 79: 78–88.
- G20 (2008) “Statement From G-20 Summit.” *New York Times* (November 16).

- _____ (2009a) “Global Plan for Recovery and Reform: Statement Issued by the G20 Leaders.” London: G20.
- _____ (2009b) “G20 Leaders Statement: The Pittsburgh Summit.” Pittsburgh: G20.
- Furstenberg, G. von, and Daniels, J. (1992) *Economic Summit Declarations, 1965–1989: Examining the Written Record of International Cooperation*. Princeton Studies in International Finance, No. 72. Princeton, N.J.: International Finance Section, Princeton University.
- Global Trade Alert (2014) “The Global Trade Disorder: The 16th GTA Report.” London: Centre for Economic Policy Research.
- Greber, J. (2016a) “OECD Blasts Reform Fatigue, Calls for More Rate Cuts, Budget Stimulus.” *Financial Review* (February 18).
- Greber, J. (2016b) “Australian Ministers Duck G20 Crisis-Prepping Talks.” *Financial Review* (March 31).
- Helleiner, E. (2010) “What Role for the New Financial Stability Board? The Politics of International Standards after the Crisis.” *Global Policy* 1 (3): 282–90.
- International Financial Institutions Advisory Commission (Meltzer Commission) (2000). Washington: Government Printing Office.
- Lo Duca, M., and Stracca, L. (2014) “The Effect of G20 Summits on Global Financial Markets.” Working Paper Series No. 1668. Frankfurt: European Central Bank.
- Morgenson, G., and Rosner, J. (2012) *Reckless Endangerment: How Outsized Ambition, Greed, and Corruption Created the Worst Financial Crisis of Our Time*. New York: St. Martin’s Griffin.
- Ostry, J., and Ghosh, A. (2013) “Obstacles to International Policy Coordination and How to Overcome Them.” IMF Staff Discussion Note. Washington: International Monetary Fund.
- Pirrong, C. (2010) “The Inefficiency of Clearing Mandates.” Washington: Cato Institute. Available at www.cato.org/publications/policy-analysis/inefficiency-clearing-mandates.
- PricewaterhouseCoopers (2015) “Global Financial Markets Liquidity Study” (August 12). Available at www.iif.com/publication/regulatory-report/pwc-global-financial-markets-liquidity-study.
- Rodrik, D. (2009) “A Plan B for Global Finance.” *The Economist* (March 12).
- Sally, R. (2001) “Looking Askance at Global Governance.” In J. Kirton, J. Daniels, and A. Freytag (eds.), *Guiding Global Order*:

- G8 Governance in the 21st Century*. Aldershot, England: Ashgate Publishing.
- Slaughter, S. (2013) “Debating the International Legitimacy of the G20: Global Policymaking and Contemporary International Society.” *Global Policy* 4 (1): 43–52.
- Sumner, S. (2013) “Why the Fiscal Multiplier is Roughly Zero.” Arlington, Va.: Mercatus Center, George Mason University. Available at mercatus.org/sites/default/files/Sumner_Fiscal_Multiplier_MOP_090313.pdf
- Taylor, L. (2014) “G20 Leaders to Meet in Australia under Pressure to Prove Group’s Relevance.” *The Guardian* (November 13).
- Thirwell, M. (2013) “Saving Multilateralism: The G20, the WTO, and Global Trade.” Sydney: Lowy Institute for International Policy.
- Truman, E. M. (2003) “The Limits of Exchange Market Intervention.” In J. Williamson and C. Fred Bergsten (eds.), *Dollar Overvaluation and the World Economy*. Special Report 16. Washington: Institute for International Economics (Peterson Institute).
- Uren, D. (2015) “Growth Strategy Seizes up as G20 and Budget Targets Hit the Wall.” *The Australian* (February 12).
- Wallison, P. J. (2015) *Hidden in Plain Sight: What Really Caused the World’s Worst Financial Crisis and Why It Could Happen Again*. New York: Encounter Books.

INTEREST GROUP ACTIVITY AND
GOVERNMENT GROWTH:
A CAUSALITY ANALYSIS
Russell S. Sobel and J. R. Clark

The special interest group model of government, employed throughout public choice theory, models the outcomes of government as a function of special interest group activity. Early work in this area by authors such as Stigler (1971) and Peltzman (1976) focused on the role of interest groups in securing regulation beneficial to the regulated industry. Subsequent formulations—including McCormick and Tollison (1981), Yandle (1983), Mueller and Murrell (1986), Shughart and Tollison (1986), Becker (1983), Sobel (2004), and Holcombe (1999)—use the interest group model to explain not only a wide variety of individual government programs and policies, but also the overall growth of government spending.¹ Even the passage of child labor laws has been attributed to interest groups such as owners of steam-driven mills, physicians, and teachers.²

In contrast, the large literature on rent seeking beginning with the seminal paper by Tullock (1967), and continuing with the contributions of Krueger (1974) and Posner (1975), explains how interest

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Russell S. Sobel is Professor of Economics and Entrepreneurship at The Citadel, and J. R. Clark is Professor of Economics at the University of Tennessee at Chattanooga where he holds the Probasco Chair of Free Enterprise.

¹See Ekelund and Tollison (2001) for a comprehensive review of this literature.

²See Marvel (1977); Anderson, Ekelund, and Tollison (1989); and Ekelund and Tollison (2001).

groups will expend resources to capture the economic rents created by government policies. In its most basic form, the argument goes that when a \$20 bill is up for grabs through a bidding process, the maximum amount of resources a group would devote to capturing that gain is \$20. Whether the rents *created by* government policy are over, under, or perfectly dissipated has been the subject of debate and exploration, but the general consensus is that the amount of rent seeking that is visibly measured is far less than would be expected given the size of government rents up for capturing (i.e., the “Tullock paradox”), although some of this differential may be explained by less visible and hard to measure in-kind rent-seeking activities.³ But merely the idea that rent-seeking activity falls below what would be expected *given* the transfers created by government implies a direction of causality in the opposite direction of the literature on the special interest group theory of government. Some extensions within this literature actually model government as trying to maximize the opportunities for politicians and regulators to “rent extract.”⁴ In this literature, governments pick policy targets and regimes that maximize the amount of rent seeking created by their actions.

Far from being a purely academic exercise, the ambiguity of the implied direction of causation between interest group activity and the size of government has both social and policy implications. On the policy side, reformers who want to slow the growth of government can be broken into two camps. The first is those who want to constrain interest group activity *as a route* to lowering government spending. Suggesting that campaign finance reform or PAC disclosure rules, for example, would be an effective way to help get government spending under control is an argument that takes for granted that it is the interest group activity that causes government spending. If the causality worked in the opposite direction, these types of reforms would be ineffective as the total rents created by policy would remain unchanged, and the means of competing for them would simply change to other avenues similar to how under rent controls side payments for items such as furniture could be used to compensate the landlord in alternate means. In other words,

³See Tullock (1980, 1989, 1997, 1998); Tollison (1982); Mateer and Lawson (1995); Mixon, Laband, and Ekelund (1994); Mixon (1995); and Sobel and Garrett (2002).

⁴See McChesney (1987), Shleifer and Vishny (1998), and Frye and Shleifer (1997).

if government spending causes interest group activity in the vein of the rent-seeking model, these reforms would simply be ineffective as they target the consequence not the cause. On the other side are those who suggest that we can curb interest group activity and rent seeking by limiting the power and spending of government through items such as constitutional restrictions, a line-item veto, or a balanced budget amendment, for example. By constraining government spending, this argument holds, there would be less interest group activity. But this conclusion relies on the direction of causality: if interest groups cause government, and not vice versa, the only means to constrain government is to first constrain interest group activity.

Nowhere is this causal distinction more blurred than in the current debates about the significant 25 percent increase in lobbying and interest group activity in the 2007–2010 period and how it relates to the federal government’s greatly expanded budget including the \$700 billion Troubled Asset Relief Program (TARP) program in October 2008, and the \$797 billion “fiscal stimulus” legislated in the 2009 American Recovery and Reinvestment Act (ARRA). Lobbying by the finance, insurance, and real estate sectors alone has been over \$450 million per year since 2008. The industry now has approximately 2,500 individual registered federal lobbyists and increased donations directly to federal political campaigns from \$287 million during the 2006 election cycle to \$503 million during the 2008 election cycle. Other sectors, such as energy, have followed similar paths of this period, with a 66 percent increase in federal lobbying expenditures, over 2,200 registered federal lobbyists, and increases in campaign contributions from \$51 million during the 2006 election cycle to \$81 million during the 2008 election cycle (Center for Responsive Politics 2013). Office space in Washington, D.C., has now become the highest priced in the country and many businesses have opened or moved their offices to the Washington, D.C., area, and the popular logic clearly relies on an argument that these responses have been *caused* by the expansion in government activity.⁵ Thus, while some argue that programs such as TARP have *caused* the increase in bank lobbying, others argue that it was the increase in bank lobbying that *caused* the passage of TARP (Allison 2013).

⁵See Cho, Mufson, and Tse (2009); Clabaugh (2010); and Lewis (2010).

Interestingly, while the so-called “Occupy Movement” and conservative/libertarian leaning scholars both argue against what they see as a large recent increase in bailouts and crony capitalism, the root cause each group identifies can be separated by the direction of causality. Followers of the “Occupy Movement” blame big, well-funded corporations and the political activity they fund for an out of control government, while the other side blames the out of control government for the rise in crony activity among firms.

The recent strand of literature on productive and unproductive entrepreneurship first elaborated by Baumol (1990, 1993, 2002), and expanded by Boettke (2001), Boettke and Coyne (2003), Coyne and Leeson (2004), and Sobel (2008), not only incorporates the government causes interest group logic, but provides it a theoretical underpinning. In this literature, the allocation of a society’s entrepreneurial talent between productive, market-based entrepreneurship and unproductive political and legal entrepreneurship (e.g., lobbying) is driven by the relative profitability of the two activities, which is a function of the quality of a country’s institutions. In countries with institutions providing secure property rights, a fair and balanced judicial system, contract enforcement, and effective limits on government’s ability to transfer wealth through taxation and regulation, the returns to unproductive entrepreneurship are low, while the returns to productive market entrepreneurship are high, thus causing fewer resources to be devoted toward interest group activity. In areas without strong institutions, entrepreneurial individuals are instead more likely to engage in attempts to manipulate the political or legal process to capture rents as the returns to unproductive activity are relatively higher. Again, in this literature it is the actions and undertakings of government that come first, and the amount of political interest group activity and lobbying is simply a consequence caused by the policies of government.

Of course, it is clearly possible that *both* are true—that is, bi-directional causality. Exogenous changes in government spending may subsequently cause changes in interest group activity, while exogenous changes in interest group activity may subsequently cause changes in government spending. This simultaneous equations-type logic implies a theoretical relationship similar to the relationship between club membership and output modeled in Buchanan’s (1965) “Theory of Clubs” in which there is an optimal membership for every given club output, and an optimal club

output for every given club membership size, and only one point at which both are simultaneously satisfied. Here, however, the argument would be that for every given level of government spending there is some optimal level of interest group activity that dissipates these rents, while for every given level of interest group activity there is a level of government spending produced by that special interest group activity, and an equilibrium would be reached at the point where both relationships are simultaneously satisfied. Under bidirectional causality, the problem of growing government and interest group activity can be effectively controlled by changes to either side, implying that constraints on government and restrictions on interest group activity both potentially can be effective tools.

In this article, we present a theoretical and empirical treatment of this issue of the direction of causality. We begin with the presentation of models that capture each side of the argument individually, and also a bidirectional simultaneous equation model. We then continue with an empirical examination of data on government spending and interest group activity to see if the nature of the causality can be identified empirically employing Granger Causality tests. We use data on total federal expenditures and two different measures of interest group activity, expenditures on lobbying, and the payroll of political organizations in Washington, D.C., and confirm the presence of a bidirectional causal relationship.

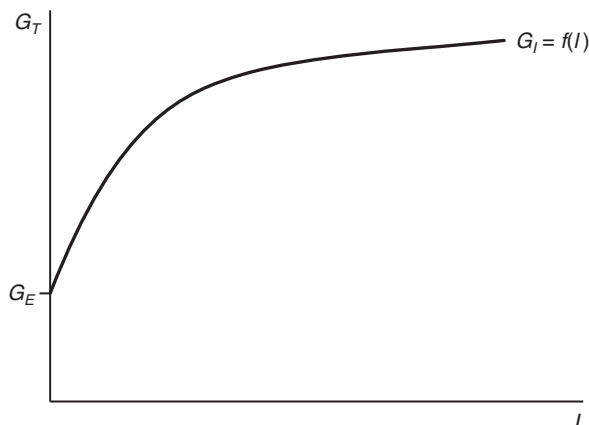
The Competing Models: A Theoretical Framework

In this section, we outline models for each of the three possibilities (two of one-direction causality, but with causality flowing in opposite directions, and a bidirectional simultaneous equation model).

The interest group causes government framework may best be thought of within a production function framework. That is, interest group activity is exogenous, and it produces government spending. In the remainder of this article, we use the notation I to refer to the value of resources devoted to interest group activity, and G_I to refer to the level of government action/spending caused by interest group activity. The production function approach may be thought of therefore as:

$$(1) G_I = f(I), \text{ where } G'_I > 0, \text{ and } G''_I < 0.$$

FIGURE 1
INTEREST GROUP ACTIVITY PRODUCES/CAUSES
GOVERNMENT SPENDING



As with any production process, output increases with additional inputs, but at a decreasing rate, so therefore $G'_I > 0$, and $G''_I < 0$. In many respects, this can be viewed equivalently to Becker's (1983) formulation of the process of the production of political pressure. In Equation (1) above, G'_I is the marginal productivity of interest group activity in generating additional government spending or activity. G'_I itself is obviously a function of many well-known parameters including the cost of organizing, cost of controlling free-riders, the size of the group, and various aspects and legal limits on lobbying and campaign finance.⁶

Allowing for the possibility of other government spending that is not caused by interest group pressure, for example, "exogenous" noninterest group spending, denoted G_E , where $G_E \geq 0$, allows for a total level of government activity/spending of $G_T = G_I + G_E$. Graphically, this relationship is now depicted in Figure 1.

We now turn to the second case, in which government spending causes interest group activity per a rent-seeking-type model. In this framework, the level of interest group activity (I) is a function of the total level of government spending (G_T) such that:

$$(2) \quad I = f(G_T), \text{ where } I'_G > 0, \text{ and } I''_G < 0?$$

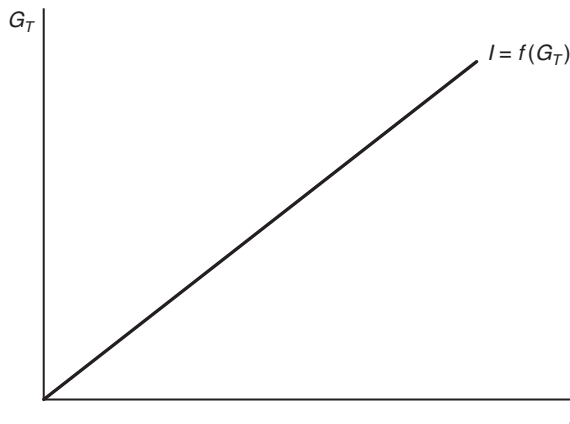
⁶See, for example, Olson (1965) and Becker (1983).

Interest group activity should be increasing in government spending, therefore $I'_G > 0$. The exact structure of this function, and the second derivative in particular, are not as obvious and deserve explicit discussion. Starting from a naïve view, let's assume that government spending is always fully and perfectly dissipated, such that if the government spends \$2 billion, then \$2 billion in interest group activity will be caused to dissipate or compete for the rent at stake. In this simple framework, $I = G_T$ and the two have identical values, implying a relationship depicted by a 45 degree line in a graph (and $I'_G = 1$). While there are many alternative theoretical game-theory models of the rent-seeking process, the most frequently used assumption is from the one-shot simultaneous move pure strategy Nash equilibrium in which each of the two players expends one-fourth of the total rent at stake, thus creating a total rent-seeking expenditure of all parties of one-half the total rent at stake.⁷ If one wished to use that assumption, the relationship depicted graphically would again be a straight line, but with a slope of one-half. In addition, using a formulation more closely related to Baumol (1990, 1993, 2002), one would want to specify that there should be some normal rate of return (zero economic profit) built into this relationship, in a present value form. This by itself would make the slope less than one as the investment must generate a normal rate of return at a minimum. Staying generalized, we assume nothing of the slope, but offer the possibility that the slope will depend on the rate of rent dissipation. Essentially, the line will be a line representing the level of interest group activity required such that the rent-seeking industry is in zero economic profit (risk-adjusted) equilibrium.

Complicating the relationship is that, viewed on a large scale, the rent-seeking industry may be either a decreasing cost, constant cost, or increasing cost industry. In the constant cost case, the line representing the relationship will be linear, $I''_G = 0$. However, if the industry is either an increasing or decreasing cost industry, then the relationship will be nonlinear. In the case of increasing cost, $I''_G > 0$ while in the case of a decreasing cost industry, $I''_G < 0$. While we allow for and consider all possibilities within our model, for current purposes, no loss of generality will result from continuing with the simple linear exposition that would be associated with a constant cost industry, and a constant ambiguous, but positive, slope in the

⁷See Godwin, López, and Seldon (2006) for a review.

FIGURE 2
GOVERNMENT SPENDING PRODUCES/CAUSES INTEREST
GROUP ACTIVITY



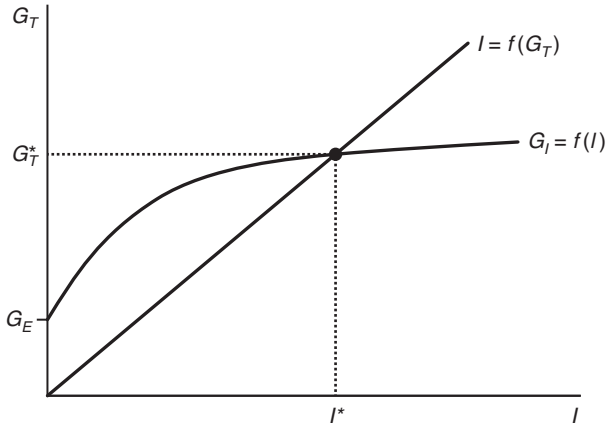
range 0 to 1. Such a relationship is illustrated in Figure 2. Note that we have set the intercept to zero in Figure 2, but that one could also include an intercept into this relationship, without altering the general results we derive.

The final possibility is the case of bidirectional causality. Here both relationships must be satisfied simultaneously. That is, the value of resources devoted to interest group activity that is required for the zero economic profit, rent-dissipation equilibrium *given* the level of government spending must also satisfy the condition that that level of government spending is the amount produced by that *given* level of interest group activity. This is equivalent to the framework of the model in Buchanan's (1965) "Theory of Clubs" in which there is an optimal club membership for every given level of club output, and also an optimal club output for every given club membership size, and only one point at which both are simultaneously satisfied. Graphically, this is shown in Figure 3 when the relationships from the two previous figures are combined.

To illustrate the nature of this bidirectional process, let's consider two situations in which the equilibrium relationship shown in Figure 3 is not satisfied. These are illustrated in Figure 4.

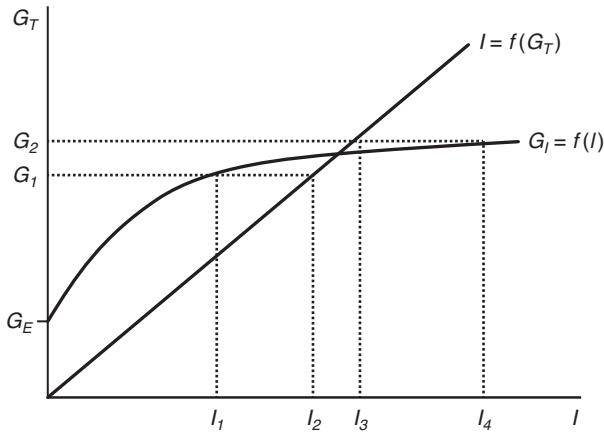
First, consider a level of interest group activity equal to I_1 . This level of interest group activity would produce a level of government

FIGURE 3
 BIDIRECTIONAL CAUSALITY EQUILIBRIUM



spending equal to G_I which can be found by moving up vertically to the line representing $G_I = f(I)$ and continuing horizontally to the left vertical axis at G_I . This is not an equilibrium, however, because at the level of spending G_I there is disequilibrium in the rent-seeking market in that there are excess economic profits. According to the rent-seeking relationship $I = f(G_T)$, the level of rent seeking required to appropriately dissipate (zero economic profit) that level of spending

FIGURE 4
 DISEQUILIBRIUM SITUATIONS



would be I_2 which can be found by moving horizontally from G_1 to the line representing the relationship $I = f(G_T)$. Therefore, due to the excess profits, additional interest group activity would enter the industry. Resources would move away from productive private market entrepreneurship into unproductive political entrepreneurship in the public sector. As this happens, it also produces additional government spending, and the movement continues until the equilibrium at the intersection of the two lines depicted in Figure 3 is achieved.

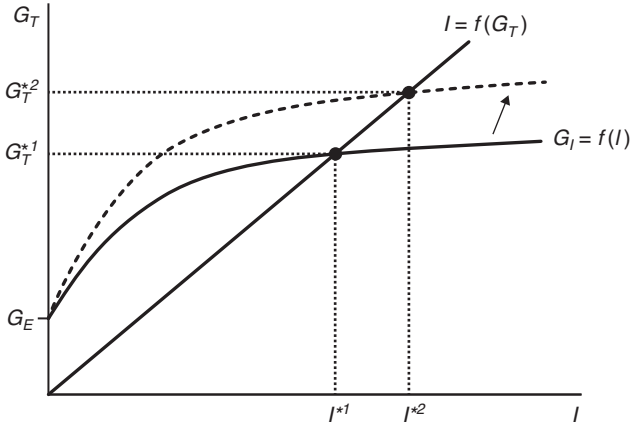
Similarly, consider a level of interest group activity equal to I_4 . This level of interest group activity would produce a level of government spending equal to G_2 . This is not an equilibrium because at the level of spending G_2 there is again disequilibrium in the rent-seeking market in that there are below-normal economic profits (i.e., economic losses). According to the rent-seeking relationship $I = f(G_T)$, the level of rent seeking required to appropriately dissipate (zero economic profit) that level of spending would be I_3 . Therefore, due to the losses, interest group activity would shrink, resources would move away from unproductive entrepreneurship back into productive entrepreneurship in the private sector. According to the production relationship, as the level of interest group activity falls, so does the level of government spending it produces, and the movement continues until the equilibrium at the intersection of the two lines depicted in Figure 3 is again achieved.

Using the Model to Understand Recent Expansions in Government Spending and Interest Group Activity

This model may now be used to illustrate and better understand the recent expansions in both government spending and interest group activity that occurred after the recent financial crisis, and resulting expansion in government spending. Rather than shifts of the curves, here because the y-axis intercepts are fixed by theory, changes are illustrated by alterations of the slopes of the two lines. We consider two cases. First, we consider the case of something happening to increase the marginal productivity of interest group activity at producing government spending (G'_I). As G'_I gets larger, the curve illustrating the $G_I = f(I)$ equation rotates upward as is shown in Figure 5.

In Figure 5, the marginal productivity of interest group activity at producing government spending has increased, rotating the line

FIGURE 5
 INCREASING MARGINAL PRODUCTIVITY OF
 INTEREST GROUP ACTIVITY



upward to the new one illustrated by the dashed line. This would result in both an expansion in government spending (from G_T^{*1} to G_T^{*2}) and an expansion in interest group activity (from I^{*1} to I^{*2}).

Are there reasons to believe that events have unfolded in recent years that have increased the productivity of interest group activity at producing spending? The most obvious chain of logic suggesting this has occurred is provided by the “Baptist and Bootleggers” model of Yandle (1983). That model argues that the simple economic interest of an interest group does not become salient or politically possible unless there is a “moral cover” to the story for providing the interest group benefits. The moral cover in recent years has been the political rhetoric of “too big to fail” and “Keynesian stimulus.” Giving billions in direct subsidies to individual businesses, from banks to energy and car companies, perhaps was only possible with the cover that these expenditures needed to be done to avoid economic collapse and to promote recovery.

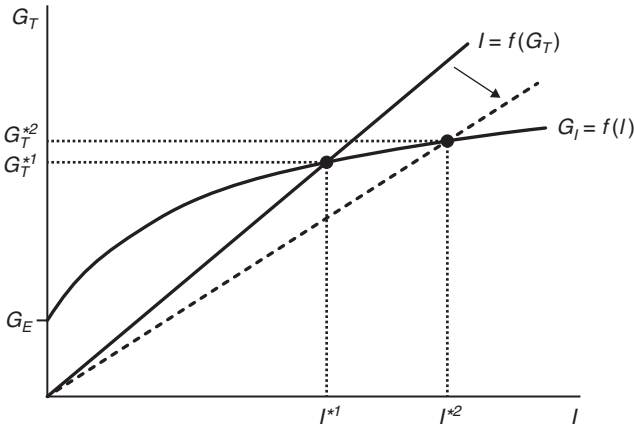
Interestingly, this line of logic helps to solve the mystery posed by Young (2013) in his policy analysis entitled “Why in the World Are We All Keynesians Again? The Flimsy Case for Stimulus Spending.” As he argues, the textbook macroeconomic literature on the eve of the financial crisis had pretty much settled on the idea that monetary policy was the more potent tool for macropolicy, and

that fiscal stimulus was far less potent, if effective at all. As an example, he points to a quote from Alan Blinder, the former vice chairman of the Federal Reserve's Board of Governors, who in 2004 concludes "virtually every contemporary discussion of stabilization policy by economists—whether it is abstract or concrete, theoretical or practical—is about monetary policy, not fiscal policy." Young (2013) even notes that some of the individuals involved in crafting and promoting the ARRA stimulus had done previous published research that would seem to argue for alternative policies that should have been followed instead. Nonetheless, his argument is simply that the shaky evidence for fiscal policy, in the past history of the United States, as well as the lack of potent current effects, and also in countries like Japan, clearly leaves one to wonder why Keynesian deficit spending has become so in fashion in recent years, despite the evidence of its ineffectiveness. The answer may very well be that these arguments provided the "Baptist/moral cover" for the special interest bootleggers to get government transfers purely in their economic self-interest. In other words, the fiscal crisis created a situation in which the lore of Keynesian stimulus became politically salient enough to allow passage of special interest spending that otherwise would not have passed without this moral cover story. In a nutshell, this Keynesian moral cover simply increased the marginal productivity of interest group activity at producing spending. If so, this would result in the change illustrated in Figure 5.

A related argument follows the logic of Clark and Lee (2003, 2005a, 2005b) who explain why and how special interest groups get better at lobbying through time (e.g., develop more human capital in lobbying as opposed to productive activity), contributing to government growth using a prisoners' dilemma model. As these interest groups form on specific issues (e.g., the banking interest groups energized in the recent financial crisis), through time their productivity grows with experience, and they may undertake actions beneficial for themselves (especially the leaders of these interest groups), even if the policies they pursue may not be in the best interest of the group they represent.

An alternative change that could produce the recent expansion in government spending and interest group activity would be a reduction in the slope of the other line, the one representing rent-seeking equilibrium, $I = f(G_T)$ as is illustrated in Figure 6.

FIGURE 6
DECREASING OPPORTUNITY COST OF
INTEREST GROUP ACTIVITY



If the slope were to fall, the line would rotate downward to the dashed line in Figure 6, producing both an expansion in government spending (from G_T^{*1} to G_T^{*2}) and an expansion in interest group activity (from I^{*1} to I^{*2}).

Are there reasons to believe that events have unfolded in recent years that have altered the zero economic profit (dissipation) equilibrium conditions in the market for interest group activity? Baumol's model of productive and unproductive entrepreneurship suggests that the amount of rent seeking is not simply a function of the profitability of rent seeking, but of the relative return from rent seeking *versus* the return from productive market activities. The financial crisis did significantly reduce profit margins, cause a reduction in private employment, displace resources, and shrink private markets. To the extent that the return in the private sector represents the opportunity cost of devoting resources to rent seeking, this lower private return should have been expected to result in a shift of resources into political entrepreneurship and rent seeking until the profitability of that activity was reduced to a level equal to the now lower profitability of private activity. That is, a recession-induced reduction in the profitability of private market activity should also result in an equilibrium reduction in the returns to rent seeking as well under the Baumol model. This reduced opportunity cost of resources devoted

to interest group activity would result in the exact change illustrated in Figure 6, with the slope of the $I = f(G_T)$ line, G' , falling.

The best way of understanding this link between a recession-induced reduction in private sector profitability and interest group activity is through a graph similar to how the mechanics of the incidence of the corporate income tax are often modeled. There are two sectors, here the private sector and the government sector (lobbying). There is a fixed stock of capital that will be allocated between the two sectors based on the rates of return. Assuming both have diminishing returns to investment, the allocation of capital between the two sectors can be illustrated graphically as in Figure 7.

In Figure 7, the length of the horizontal axis equals the total stock of capital. The “demand” for capital in the private sector (line D_P) on the left side of the graph is a line whose (diminishing) height shows that the return to capital in the sector falls with increased investment. A similar line from the right axis shows the diminishing returns to capital involved in lobbying / interest group activity in the

FIGURE 7
BAUMOL-TYPE ALLOCATION OF
CAPITAL (HUMAN & PHYSICAL)
BETWEEN PRODUCTIVE PRIVATE ACTIVITY AND
UNPRODUCTIVE INTEREST GROUP ACTIVITY

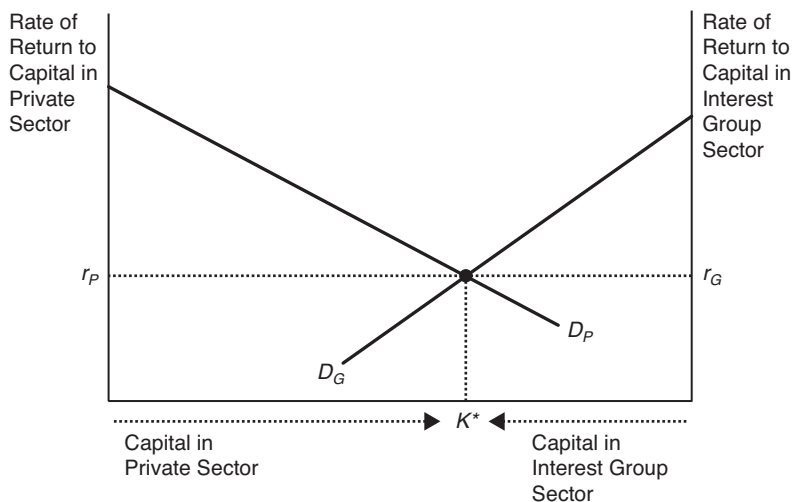
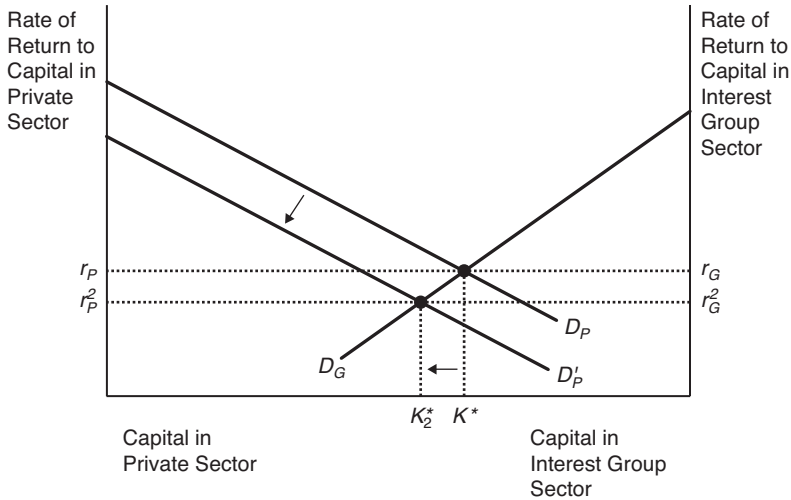


FIGURE 8
 RECESSION-INDUCED SHIFT OF
 CAPITAL (HUMAN & PHYSICAL) INTO
 UNPRODUCTIVE INTEREST GROUP



government sector (D_G). The allocation of capital between the two sectors (both human and physical) will be at an equilibrium where the two sectors, at the margin, are equally profitable for investing capital, at point K^* where $r_P = r_G$ in Figure 7 where the two curves intersect.⁸

Figure 8 shows how a recession-induced reduction in the return to capital employed in the private sector would impact this equilibrium. The curve showing the private returns would fall to D'_P . This will cause capital at the margin to flow into the higher returns in interest group activity in the government sector, pushing down returns in that sector as well, until a new equilibrium is restored at point K_2^* with a lower profitability in both sectors ($r_P^2 = r_G^2$), and a larger proportion of capital now in the interest group sector. This is

⁸While Baumol's original theory focuses on the allocation of human capital (e.g., entrepreneurial effort) between the sectors, Hall, Sobel, and Crowley (2010) extend this model to include both human and physical capital and provide evidence that the logic applies to both equally. We therefore treat the capital stock generally in the figures, and conjecture that both human and physical capital are shifted between the sectors in the general manor shown.

the fundamental change that causes and is reflected in the $I = f(G_T)$ line rotating downward, due to the lower slope, in Figure 6.

A final, and likely, possibility is that both of the changes in Figures 5 and 6, the higher productivity of interest group activity and the shift of resources into interest group activity caused by the recession, contributed to the recent expansion in both government spending and interest group activity. That is, perhaps both shifts happened, and reinforced each other to produce the expansions in government spending and interest group activity that have unfolded since the financial crisis.

As an important aside in terms of using the model, we also note that the shift represented in Figure 5 is how one would illustrate, more generally, learning effects in the model when interest group activity (or the government) is relatively new in a geographic area. As is suggested by authors such as Olson (1982), Johnson and Libecap (1994), and Browder (2015), the productivity of interest groups may rise with their tenure and experience with a newly formed government through time. In addition, closer in line with Browder is the idea that government's ability to create and extract rents may also grow through time as politicians learn how to better participate in this process of generating benefits to interest groups. These learning type effects, which are in keeping with work in both experimental economics and evolutionary biology, would result in an increase in the productivity of interest group inputs in producing each level of government spending. These type effects would be represented in a similar manner to the shifts shown in Figure 5.

Testing the Alternatives

The previous sections outlined a model that relied on the direction of causality between government spending and the value of resources devoted to interest group activity. While the rent-seeking model stresses how expansions in government spending result in more interest group activity to dissipate the rents, the special interest model stresses how increases in interest group activity produce expansions in government. There are three alternatives:

1. The value of resources devoted to interest group activity causes government spending.
2. Government spending causes the value of resources devoted to interest group activity.
3. Both (1) and (2) are true, and there is bidirectional causality.

In this section, we attempt to discern these alternatives empirically employing Granger-Sims causality tests. These tests, despite their drawbacks, essentially see, on average, which series moves “first” and which moves “second” in a time series framework.

Our Granger-Sims causality tests are conducted by estimating the following system of equations as a structural Vector Autoregression (VAR):

$$(3) \quad G_t = \beta_1 + \sum_{i=1}^r (\beta_{1i} \cdot G_{t-i}) + \sum_{i=1}^s (\alpha_{1i} \cdot I_{t-i}) + \varepsilon_{1t}$$

$$(4) \quad I_t = \alpha_1 + \sum_{i=1}^s (\alpha_{2i} \cdot I_{t-i}) + \sum_{i=1}^r (\beta_{2i} \cdot G_{t-i}) + \varepsilon_{2t}$$

where G is government spending and I is a measure of the value of resources devoted to interest group activity. We set up the null hypotheses that $\alpha_{1i} = 0$ and $\beta_{2i} = 0$, for all $i = 1$ to r, s . Intuitively, if the set of lagged values of I are jointly significant in the equation for G , then we can reject the null that I does not cause G , in favor of the alternative hypothesis that indeed I Granger-causes G (and vice versa). The optimal lags (r and s) are determined by using the Bayesian (Schwarz) information criterion (BIC) on the vector autoregressive equations, and the test of the joint significance of the lagged variables is performed using an F-test to determine if causality exists, and in which direction. All three alternatives are possible findings (only G Granger-causes I , only I Granger-causes G , or there is bidirectional causality and both are true).

Because our data are time series, we need to first ensure that our series are stationary. To test for unit roots, we employ two tests, the first of which is the Augmented Dickey-Fuller (ADF) test with the following regression specification:

$$(5) \quad \Delta y_t = \theta_0 + \gamma \cdot y_{t-1} + \sum_{i=1}^P \beta_i \cdot \Delta y_{t-i} + \varepsilon_t$$

where y is the variable of interest and P is the number of lags determined using the Bayesian (Schwarz) information criterion (BIC). The series is stationary if and only if $\alpha < 1$, which by Equation 2 is equivalent to a test for $\gamma < 0$, where $\gamma = (\alpha - 1)$. Because the standard t-statistic for γ is a test of $H_0: \gamma = 0$, a negative and significant t-statistic for γ implies that $\alpha < 1$, and the series is stationary. If the t-statistic for γ is not significant, the series is nonstationary. If the series is nonstationary, it must be first-differenced (annual change) and the ADF test performed again to ensure the resulting series is stationary. If it is not, the process continues until the order of differencing required to make the series stationary is found.

The second test we employ to check for stationarity is the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test based on the following test statistic:

$$(6) \quad \eta = \frac{\sum_{t=1}^T S_t^2}{T^2 \bar{\sigma}^2}$$

where $S_t = \sum_{s=1}^t e_s$ and $\bar{\sigma}^2$ is an estimate of the long-run variance of $e_t = (y_t - \bar{y})$. This Lagrange-multiplier test is intended to complement the Dickey–Fuller test. In the KPSS test, the null hypothesis is opposite to that in the ADF test; therefore, a statistically significant coefficient indicates the series is nonstationary and requires differencing (this process continues until the test statistic is no longer statistically significant.). The reason is, under the null, the long-run variance is a well-defined finite number, and the test statistic has a well-defined asymptotic distribution (but not under the alternative).

We conduct our empirical tests using measures of U.S. federal government spending and interest group activity targeted at federal legislation or in the Washington, D.C., area. Because government spending data is normally reported on a fiscal year basis, which would detract from the timing of the lag structure and possibly influence our results, we begin by estimating the models on calendar year federal spending data from the U.S. Department of Commerce, Bureau of Economic Analysis (BEA). The differences in what is and is not included in the BEA measure versus the federal budget are minor.⁹ In addition, we also estimate the models using the budget-based fiscal year data for comparison. We estimate our models using total federal expenditures (in billions), and all nominal values throughout our empirical analysis are converted to constant (real) 2012 dollars using the Consumer Price Index (CPI).

For measures of the value of resources devoted to interest group activity/rent seeking, we employ the only two dollar-based measures available, lobbying spending and the payroll of political/lobbying organizations located in the Washington, D.C.,

⁹For a full listing of the differences, see U.S. Department of Commerce, Bureau of Economic Analysis, Table 3.18B: “Relation of Federal Government Current Receipts and Expenditures in the National Income and Product Accounts to the Budget, Fiscal Years and Quarters.” Available at www.bea.gov/scb/pdf/2012/10%20October/1012_new_nipa_tables.pdf.

area. Annual data on lobbying spending (in millions) is obtained from The Center for Responsive Politics (www.opensecrets.org), and annual data on the payroll of lobbying/political organizations is obtained from The United States Census Bureau, County Business Patterns database (www.census.gov/econ/cbp), both series beginning in 1998. We recognize that this limits our number of observations to less than would be desired for both series (15), but given the sparse nature of data on interest group activity, we move forward as the model can be estimated efficiently with the small sample available. We include industry code 813xxx which was identified by Sobel and Garrett (2002) as a key industry subset expanded in state capital cities which is a reliable indicator of lobbying activity. Again, all nominal values are corrected for inflation to 2012 real dollars. Payroll (in millions) is a good measure of the value of labor resources employed in lobbying/rent seeking, and because labor is the variable input in the short run, it should rapidly reflect changes in the level of interest group activity making it very suitable for a time-based causality test.

First, we must ensure our series are stationary, and Table 1 presents the results of our unit-root/stationarity tests. All three of our main variables are found to be nonstationary in their levels form in both tests (ADF and KPSS), so all three are then converted to annual change versions (first differenced), and the resulting series are all stationary in both tests (ADF and KPSS). We therefore employ these first-differenced versions of our variables in our Granger-Sims causality tests.

With our transformed series, we perform the vector-autoregression-based systems estimation to test the direction of causality, and our results are presented in Table 2. The upper two rows of results show the results using total federal expenditures using the calendar year data, while the lower two rows of results show the results using the fiscal year data. The tests on the calendar year data indeed indicate bidirectional Granger causality for total federal spending and both measures of the value of resources devoted to interest group activity. This suggests that the theoretical model presented earlier, in which both have to be in simultaneous equilibrium, is the correct model.

While we believe the BEA data on federal spending by calendar year matches up with our measures of interest group activity more properly for a lag-based empirical test, we also perform our analysis

TABLE 1
UNIT ROOT TESTS

Variable	ADF	KPSS	Result
Variables in Levels			
Real Federal Expenditures	1.00	0.96***	Nonstationary
Real Lobbying Spending	-0.93	0.55**	Nonstationary
Real Payroll of Political Organizations in Washington, D.C.	-0.31	0.54**	Nonstationary
Variables in First-Difference (Change) Form			
Change in Real Federal Expenditures	-3.71***	0.18	Stationary
Change in Real Lobbying Spending	-3.45***	0.17	Stationary
Change in Real Payroll of Political Organizations in Washington, D.C.	-2.79**	0.10	Stationary

NOTES: ADF is the augmented Dickey-Fuller χ^2 test; KPSS is the Kwiatkowski, Phillips, Schmidt, and Schin (1992) test. The null hypothesis for the ADF test is nonstationarity (unit root), while the null hypothesis for the KPSS test is stationarity (no unit root). All tests include a constant, and lag length determined by BIC. Statistical significance as follows: * = 10 percent, ** = 5 percent, and *** = 1 percent.

on fiscal year spending data to check for robustness. The results using the fiscal year data (which, for the federal government, runs from October 1 of the previous year to September 30 of the year indicated by the fiscal year) is presented in the final two rows of Table 2. Using lobbying spending the model again supports bidirectional causality, but using the payroll of political organizations, one of the test statistics falls just short of the 10 percent normal threshold. Taken at face value, the final row of results indicate only one-way Granger

TABLE 2
GRANGER CAUSALITY TESTS

Variable	H ₀ : Government Spending Does Not Cause Interest Group Activity (<i>F</i> -statistic)	H ₀ : Interest Group Activity Does Not Cause Government Spending (<i>F</i> -statistic)	BIC	Lags	Result
Measure: Total Federal Expenditures (Calendar Year)					
Real Lobbying Spending	12.57***	3.43*	24.31	2	Bidirectional Granger Causality
Real Payroll of Political Organizations in Washington, D.C.	4.75*	4.92*	22.54	1	Bidirectional Granger Causality
Measure: Total Federal Expenditures (Fiscal Year)					
Real Lobbying Spending	6.49**	6.21**	24.76	2	Bidirectional Granger Causality
Real Payroll of Political Organizations in Washington, D.C.	7.41**	3.29	23.49	1	Government Spending Granger-causes Interests Groups (one way) [The significance level of the other was 10.7% for bidirectional causality.]

NOTES: All variables in first-difference (change) form for stationarity per unit root tests. The null hypothesis for the tests is NON-causality; therefore, significant test statistics imply there is causality. All tests include a constant, and lag length determined by BIC. Statistical significance as follows: * = 10%, ** = 5%, and *** = 1%.

causality running from government spending to interest group activity (supportive of the rent-seeking model but not the interest group model of government), however, the significance level of the one that is not significant at traditional levels is 10.7 percent, just slightly higher than the 10 percent normal threshold, suggesting the relationship is extremely close to bidirectional causality.

Thus, in three of the four models our data confirm the idea of bidirectional causality between government spending and interest group activity. Given the limited historical data available, however, which makes our sample size smaller than desired, the model does converge and estimate efficiently. We do hope future research will one day take advantage of the longer time series of data then available to confirm our results.

Conclusion

The interest group theory of government holds that the activities of well-organized interest groups produce government spending and policies. That is, government action is a result, or product, of interest group activity. The separate, but related, literature on rent seeking, to the contrary, stipulates that when government “rents” are available, interest group activity rises to dissipate, or compete over, these benefits created by government. In this later view, interest group activity is a causal result of government action.

Recent years have seen a massive expansion in both federal government spending and also of interest group activity. Since the financial crisis the U.S. federal government, through bailout and stimulus programs such as TARP and ARRA, has made available trillions in new spending benefitting well-defined interest groups and constituencies. Office space in Washington, D.C., is now the most expensive in the nation, and the measured lobbying activities at the federal level have risen by 25 percent, with some industries such as finance, insurance, and real estate now spending over \$450 million per year on lobbying (those industries are now represented by approximately 2,500 registered lobbyists). While some accounts of these events blame the increased lobbying activity on the expansions in government spending, an equally large number blame the increased government spending on the rising pressure of interest groups for new spending programs to aid in the economic recovery efforts.

Whether interest groups cause government spending, or whether government spending causes interest group activities, or both are true, is the central question in this paper. We present both a theoretical and empirical examination. Our theoretical models begin by graphically illustrating each of the three alternatives. A unified framework, within which causality runs in both directions, is best understood in a manner similar to Buchanan's "Theory of Clubs" model. We then use the model to illustrate recent events through the recession-induced reduction in private-sector profitability lowering the opportunity cost of lobbying, coupled with an increasing productivity of interest groups in producing government spending with a new Keynesian-themed "Baptist" cover for their bootlegging-based demands for government handouts.

We then perform empirical tests of the direction of the causality that confirm bidirectional causality. Specifically, we use Granger Causality tests to determine the causal relationships. We employ data on total federal expenditures and two different measures of interest group activity, expenditures on lobbying, and the payroll of political organizations in Washington, D.C., and confirm the presence of a bidirectional causal relationship. Thus, exogenous changes in government spending will produce changes in interest group activity, and vice versa. Government outcomes, and the level of interest group activity are simultaneously determined in a framework where both sides of the market must be in the equilibriums postulated by their respective theories (the interest group theory and the rent seeking theory). These results have important implications for those who would wish to curb the growth of government and reduce interest group activity and lobbying. Because of the bidirectional causality, the level of both activities may be curtailed by policy changes on either side of the equation. That is, reforms such as a balanced budget act that curb the growth in government spending will also reduce interest group activity, and policies that restrict interest group activity and lobbying (lobbying disclosure rules, donation limits, etc.) will also curb both interest group activity and the growth of government spending. Given the limited historical data available for our tests, however, we hope future research may employ longer samples to reestimate and confirm our findings.

References

- Allison, J. A. (2013) *The Financial Crisis and the Free Market Cure*. New York: McGraw Hill.
- Anderson, G. M.; Ekelund, R. B. Jr.; and Tollison, R. D. (1989) “Nassau Senior as Economic Consultant: The Factory Acts Reconsidered.” *Economica* 56 (221): 71–81.
- Baumol, W. J. (1990) “Entrepreneurship: Productive, Unproductive and Destructive.” *Journal of Political Economy* 98 (5): 893–921.
- _____ (1993) *Entrepreneurship, Management, and the Structure of Payoffs*. Cambridge, Mass.: MIT Press.
- _____ (2002) *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*. Princeton, N.J.: Princeton University Press.
- Becker, G. (1983) “A Theory of Competition among Pressure Groups for Political Influence.” *Quarterly Journal of Economics* 98 (3): 371–400.
- Blinder, A. S. (2004) “The Case against Discretionary Fiscal Policy.” CEPS Working Paper No. 100 (June).
- Boettke, P. J. (2001) *Calculation and Coordination: Essays on Socialism and Transitional Political Economy*. New York: Routledge.
- Boettke, P. J., and Coyne, C. J. (2003) “Entrepreneurship and Development: Cause or Consequence?” *Advances in Austrian Economics* 6: 67–87.
- Browder, B. (2015) *Red Notice: A True Story of High Finance, Murder, and One Man’s Fight for Justice*. New York: Simon & Schuster.
- Buchanan, J. M. (1965) “An Economic Theory of Clubs.” *Economica* 32 (125): 1–14.
- Center for Responsive Politics (2013) *OpenSecrets.org*. (May 14). Available at www.opensecrets.org.
- Cho, D.; Mufson, S.; and Tse, T. M. (2009) “In Shift, Wall Street Goes to Washington.” *Washington Post* (December 9).
- Clabaugh, J. (2010) “D.C. Office Demand Highest in 10 Years.” *Washington Business Journal* (October 28).
- Clark, J. R., and Lee, D. R. (2003) “The Increasing Difficulty of Reversing Government Growth: A Prisoners’ Dilemma That Gets Worse with Time.” *Journal of Public Finance and Public Choice* 21 (2–3): 151–65.

- _____ (2005a) "Leadership, Prisoners' Dilemmas, and Politics." *Cato Journal* 25 (2): 379–97.
- _____ (2005b) "Prisoners' Dilemmas, Leadership, and the Destructive Growth in Government." *Troy University Business and Economic Review* 29 (1): 16–22.
- Coyne, C., and Leeson, P. T. (2004) "The Plight of Underdeveloped Countries." *Cato Journal* 24 (3): 235–49.
- Ekelund, R. B., and Tollison, R. D. (2001) *The Elgar Companion to Public Choice*. Cheltenham, U.K.: Edward Elgar.
- Frye, T., and Shleifer, A. (1997) "The Invisible Hand and the Grabbing Hand." *American Economic Review* 87 (2): 354–58.
- Godwin, K. R.; López, E. J.; and Seldon, B. J. (2006) "Incorporating Policymaker Costs and Political Competition into Rent-Seeking Games." *Southern Economic Journal* 73 (1): 37–54.
- Hall, J. C.; Sobel, R. C.; and Crowley, G. R. (2010) "Institutions, Capital, and Growth." *Southern Economic Journal* 77 (2): 385–405.
- Holcombe, R. G. (1999) "Veterans Interests and the Transition to Government Growth: 1870–1915." *Public Choice* 99 (3–4): 311–26.
- Johnson, R. N., and Libecap, G. D. (1994) *The Federal Civil Service System and the Problem of Bureaucracy*. Chicago: University of Chicago Press.
- Kreuger, A. O. (1974) "The Political Economy of a Rent-Seeking Society." *American Economic Review* 64 (3): 291–303.
- Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; and Shin, Y. (1992) "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root." *Journal of Econometrics* 54 (1–3): 159–78.
- Lewis, C. S. N. (2010) "Rents Signal Rise of D.C., Fall of N.Y." *Wall Street Journal* (January 10).
- Marvel, H. P. (1977) "Factory Regulation: A Reinterpretation of Early English Experience." *Journal of Law and Economics* 20 (2): 379–402.
- Mateer, G. D., and Lawson, R. A. (1995) "The Thrill of Victory, the Agony of Defeat." *Public Choice* 83 (3): 305–12.
- McChesney, F. S. (1987) "Rent Extraction and Rent Creation in the Economic Theory of Regulation." *Journal of Legal Studies* 16 (1): 101–18.

- McCormick, R. E., and Tollison, R. D. (1981) *Politicians, Legislation, and the Economy: An Inquiry into the Interest Group Theory of Government*. Boston: Kluwer.
- Mixon, F. G. Jr. (1995) "To the Capitol, Driver: Limousine Services as a Rent Seeking Device in State Capital Cities." *Rivista Internazionale di Scienze Economiche e Commerciali* 42: 663–70.
- Mixon, F. G. Jr.; Laband, D. N.; and Ekelund, R. B. Jr. (1994) "Rent Seeking and Hidden In-Kind Resource Distortion: Some Empirical Evidence." *Public Choice* 78: 171–85.
- Mueller, D. C., and Murrell, P. (1986) "Interest Groups and the Size of Government." *Public Choice* 48 (2): 125–46.
- Olson, M. Jr. (1965) *The Logic of Collective Action*. Cambridge, Mass.: Harvard University Press.
- _____ (1982) *The Rise and Decline of Nations*. New Haven, Conn.: Yale University Press.
- Peltzman, S. (1976) "Toward a More General Theory of Regulation." *Journal of Law and Economics* 19: 211–40.
- Posner, R. A. (1975) "The Social Costs of Monopoly and Regulation." *Journal of Political Economy* 83 (4): 807–27.
- Shleifer, A., and Vishny, R. W. (1998) *The Grabbing Hand: Government Pathologies and Their Cures*. Cambridge, Mass.: Harvard University Press.
- Shughart, W. F., and Tollison, R. D. (1986) "On the Growth of Government and the Political Economy of Legislation." *Research in Law and Economics* 9 (1): 111–27.
- Sobel, R. S. (2004) "Can Public Choice Theory Explain the U.S. Budget Surpluses of the 1990s?" *Journal of Public Finance and Public Choice* 23 (3): 169–82.
- _____ (2008) "Testing Baumol: Institutional Quality and the Productivity of Entrepreneurship." *Journal of Business Venturing* 23 (6): 641–55.
- Sobel, R. S., and Garrett, T. A. (2002) "On the Measurement of Rent Seeking and Its Social Opportunity Cost." *Public Choice* 112 (1–2): 115–36.
- Stigler, G. J. (1971) "The Theory of Economic Regulation." *Bell Journal of Economics and Management Science* 3 (1): 3–18.
- Tollison, R. D. (1982) "Rent Seeking: A Survey." *Kyklos* 35: 575–602.
- Tullock, G. (1967) "The Welfare Cost of Tariffs, Monopolies, and Theft." *Western Economic Journal* 5 (3): 224–32.

- _____ (1980) *Toward a Theory of the Rent Seeking Society*. College Station: Texas A&M Press.
- _____ (1989) *The Economics of Special Privilege and Rent Seeking*. Boston: Kluwer.
- _____ (1997) “Where Is the Rectangle?” *Public Choice* 91 (2): 149–59.
- _____ (1998) “Which Rectangle?” *Public Choice* 96 (3): 405–10.
- Yandle, B. (1983) “Bootleggers and Baptists: The Education of a Regulatory Economist.” *Regulation* 7 (3): 12–16.
- Young, A. T. (2013) “Why in the World Are We All Keynesians Again? The Flimsy Case for Stimulus Spending.” *Cato Policy Analysis*, No. 721 (February 14). Washington: Cato Institute.

THE PECULIAR BUSINESS OF POLITICS

Richard E. Wagner

This article stems from Jim Dorn's invitation to encapsulate my recent book, *Politics as a Peculiar Business: Insights from a Theory of Entangled Political Economy* (Wagner 2016). Where standard political economy treats states as singular entities that intervene in economies, I treat states as networks of *peculiar* enterprises that operate inside a society's market arrangements. This peculiar quality ramifies throughout a society. To be successful, political enterprises must raise sufficient revenue to return profits to investors. Those profits are disguised through indirect transactions and ideological formulations, but are profits all the same. There is thus a simple explanation for why political enterprises grow relative to commercial enterprises: they grow because they offer higher returns to *relevant investors* than what those investors could obtain through commercial enterprises. The analytical challenge a theory of entangled political economy must face is to explain the profit-seeking reality of the sub-system of political enterprises that operates inside a society. This article sketches the main conceptual issues with which a theory of entangled political economy must contend, and closes by considering possible implications for efforts to limit the reach of the political within society.

To start, equilibrium theory and its associated method of comparative statics is incapable of conveying the analytical vision of

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Richard E. Wagner is the Hobart R. Harris Professor of Economics at George Mason University, and Distinguished Senior Fellow in the F. A. Hayek Program for Advanced Study in Philosophy, Politics, and Economics.

entangled political economy. This vision is one of continual change injected into societies by creative agents with interests that sometimes support and at other times oppose those of other agents. Cooperation and conflict are both ineradicable features of societies (Hirshleifer 2001). This theory must be constructed in the active voice, in contrast to the passive voice of equilibrium theory. This construction requires a theory of creative systems, in contrast to the standard vision of robotic or mechanical systems (Bertalanffy 1968). The magic number for conveying the central features of market interaction under private ordering is two, for two is sufficient to convey the generation of mutual gains from trade. By contrast, the magic number for illustrating public ordering is three, for three is necessary to convey how power can be used to reward friends at the expense of enemies (Schmitt [1932] 1996). The peculiar commercial quality of political interaction generates a continuing parade of societal tectonics (Young 1991), in contrast to the placidity of equilibrium theory. These tectonics can transform a constitution of liberty into a constitution of servility as private ordering recedes relative to public ordering. This article sets forth the principal analytical constructions by which the entangled vision is executed, and it closes by considering both the scope for and difficulties of restricting the reach of the political within an unavoidably entangled system of political economy.

Two Visions of Political Economy: Additive and Entangled

The compound noun “political economy” denotes a relationship between the two simpler nouns polity and economy. Just how that relationship is construed entails much of significance, as I explain in *Politics as a Peculiar Business*. Standard formulations of political economy, as illustrated crisply by Persson and Tabellini (2000), construe polity and economy as independent entities, with polity acting on economy to change economic outcomes in some fashion. This is the standard vision of public policy wherein a policy mechanic works on an economic engine to improve its performance, or perhaps to degrade it depending on whom you ask.

This standard framework can be reasonably described as *additive* political economy to indicate that political economy denotes an adding together of polity and economy wherein each entity maintains its

original qualities, much as a collision between two billiard balls leaves their original qualities unchanged. While *Politics as a Peculiar Business* explains that this vision of political economy is incoherent, that vision nonetheless provides ideological cover for nearly the entire gamut of state activity. There is no political activity that won't find supporters who claim to seek to fix some societal defect. After all, what political program has ever been advocated by someone who claims that there is nothing wrong with the economic engine or process? Additive political economy provides ideological cover for the interventionist state by playing upon the universal recognition that machines sometimes malfunction and require mechanics to repair them. In contrast, entangled political economy locates such so-called policy mechanics as profit-seeking enterprises that operate within the economic process, except that those enterprises have peculiar features relative to ordinary enterprises, as Eusepi and Wagner (2011) elaborate.

Orthodox welfare economics and its claim on behalf of the Pareto efficiency of competitive equilibrium reflects the additive vision of political economy. The first-order conditions for Pareto efficiency pertain to transactions among commercial actors. If those conditions are satisfied, the economic engine is working efficiently. Such conceptual constructions as public goods and external effects invoke claims that the engine is not working efficiently and requires intervention by policy mechanics. To be sure, there has been extensive debate over whether common claims of inefficient performance are truly what proponents of state intervention claim they are, as Coase (1974) and Krause (2015) illustrate for lighthouses and Cheung (1973) and Johnson (1973) illustrate for bees.

The language of welfare economics enables someone to support or oppose nearly any imaginable policy proposal. Markets work well or they don't, depending on the proclivities of the observer and not on any so-called facts of the matter. This situation arises because the conditions for Pareto efficiency refer to states of mind. Those states cannot be observed. Only action can be observed, and action speaks ambiguously. Does a ship crashing on rocks indicate market failure? Or does it show that scarcity will be present even in a Pareto-efficient world? An economist can argue either position. One could argue that the ship's sinking reflects the unfortunate fact of scarcity, because if it were Pareto efficient to build another lighthouse it would have been built already (Stigler and Becker 1977). One could

argue alternatively that transaction costs are involved in trying to aggregate small valuations across large numbers of people, and that a public agency is able to lower that cost (Kahn 1966). Yet again, someone could argue that political agencies have neither the incentive to perform such aggregation nor the knowledge that can only be generated through market transactions (Hayek 1937, 1945). In short, the categories of welfare economics do not provide a recipe for the resolution of disputes but rather provide a grammar for generating interminable disputes (Wagner 2015).

This interminable quality arises because of the presumption buried within the framework of welfare economics that those who articulate those arguments are not ordinary economic actors with ordinary commercial interests, but rather are detached from and independent of commercial activity. Yet real mechanics operate within the market, and an economist can use the theory of markets to say useful and intelligible things about observed patterns of mechanical activity within a society. How is it any different for so-called policy mechanics? The scheme of entangled political economy claims that there is no formal difference, for policy advocacy reflects the same economizing logic as ordinary commerce (McCormick and Tollison 1981, Tollison and Wagner 1991). To be sure, there are significant substantive differences due to the peculiar quality of political activity relative to commercial activity. Political activity *is* commercial activity, only it is not organized through the private law framework of property and contract (Streit 1992). Indeed, these days there is perhaps little commercial activity that stands outside the direct influence of public ordering.

It should be noted that entangled political economy is not some new theoretical development. Entanglement was a common scheme of thought during the classical period of political economy, as illustrated by Robbins (1952) and Samuels (1966), as well as by Frank Knight's (1960) examination of the problem of bringing intelligence to bear on democratic action. With respect to practice, moreover, Jonathan Hughes (1977) explained that entanglement among political and commercial enterprises was widespread in Colonial America, though not nearly to the extent that it is today. It was only with the coming of the neoclassical period in economics starting late in the 19th century, with its replacement of political economy with economics (Milonakis and Fine 2009), that the construction of additive political economy starts to take shape with its isolation and separation

of what were thought to be purely economic phenomena from political phenomena. Entangled political economy represents a reformation of classical modes of thinking, as Maria Paganelli (2014) illustrates with respect to Adam Smith, making use of new modes of thought in the process.

Most significantly, the entangled vision of political economy treats economics as a genuine social science and not a science of rational action writ large (Wagner 2010). To execute this alternative scheme of thought requires replacement of the common presumption that economic observations pertain to states of societal equilibrium. If observations are thought to reflect states of societal equilibrium, sound economic reasoning requires that society be reduced to a representative individual to economize on analytical effort (Kirman 1992). Once this reduction is accomplished, there is no option but to treat economics as a science of rational action writ large because society has effectively become a single individual for analytical purposes. By contrast, if economics is to be treated as a genuine social science it must embrace a scheme of thought that entails creative and open-ended interaction among individuals being continually injected into society, as Giandomenica Becchio (2014) illustrates for Carl Menger. Hence, entangled political economy is articulated within a nonequilibrium analytical framework at the social, though not individual, level of analysis. There is no presumption that mutual consistency among individual plans is present within a society. Indeed, the typical presumption is that such consistency is never attained.

Executing the Entangled Vision: Challenges and Opportunities

Additive political economy reflects a conceptual antinomy wherein a market economy is described as an order consisting of myriad organizations while a polity is treated as a single organization. In an earlier age, when governmental budgets were in single digits as a share of GDP, and where public ordering had but modest presence within a society, it would surely have been a reasonable approximation to use the theory of markets to characterize the preponderance of economic activity, while leaving politically organized activity standing apart from the theory of markets. To be sure, entanglement was present even in those earlier times, as Hughes

(1977) explains. Yet that presence was trifling in comparison with the presence of public ordering now.

In modern times, such approximation-based reasoning can lead thought significantly astray. It is common to think of the U.S. federal government as an organization directed by a president. The federal budget, for instance, is commonly described as “the president’s budget.” To speak this way is to engage in formal and not in substantive speech. Aaron Wildavsky (1975: 211) estimates that the American president spends around 20 hours a year working on what is designated as “his” budget. In no reasonable way can such a budget be described as a product of a president’s deliberate choice. To the contrary, that budget is the product of a process through which enterprises compete for political support. This competitive process is, at present, only incompletely understood (Wagner 2007: 125–54; Wagner 2012a). This situation reflects what Gordon Tullock (1965) described as “bureaucratic free enterprise,” which denotes a system of interaction among competing entities that operate outside a framework of private property, and does not represent the activity of a coherent organization. There is no way that a single mind can truly choose the budget of the federal government by devoting 20 hours to the task. To develop a reasonable understanding of the world we experience, there is no option but to incorporate politically organized activity into a theory of order, which is what entangled political economy seeks to accomplish.

Throughout the social sciences, though not only the social sciences, theorists must theorize about objects that they can apprehend theoretically but not observationally. This situation means that inputs into our theories are also constructed through theory. The predominant line of economic theory for the past century or so has held that our observations pertain to states of systemic equilibrium. Yet there is no reason to do this other than to ease the theorist’s life. Proceeding in this fashion also means that change can come only as an exogenous shock to an equilibrated system. It further means that an economic system can be reasonably reduced to a representative agent, because no useful theoretical work is accomplished by working with a larger number of agents. Theorists understandably emphasize the ability of a scheme of thought to sharpen the image of the phenomena they are examining. It should also be noted that a scheme of thought also precludes other phenomena from being seen. For instance, a theory based on universal price taking will not be able

to serve as a theory of price formation. Should a theorist desire to construct a framework where change comes about through interaction among conflicting plans within a society, it will be necessary to develop some nonequilibrium framework.

Politics as a Peculiar Business rejects these conventional theoretical presumptions and seeks to create some elements of a nonequilibrium approach to social theory. It accepts the presumption that individuals are economizing agents, and so seek to replace conditions they value less highly with conditions they value more highly; however, it treats those agents as creative and not mechanical. A person who faces unchanged prices and incomes can thus choose differently as time passes due simply to a desire to avoid boredom, to say nothing of recognizing that action occurs in combinatorial space where exploration for fresh ideas always beckons. While societies are generally orderly systems of interaction, they also entail turbulence. That turbulence, however, stems from the creative actions of people within the society which continually upset other plans within that society. Turbulence and not placidity is the normal state of a system of entangled political economy. A system based on private ordering, however, is generally better able to calm turbulence than a system where public ordering occupies the social foreground (Wagner 2012b).

In any case, societies are conceptualized as complex systems of creative action and interaction wherein novelty is continually inserted. Within this analytical framework, a conceptual problem arises in bringing commercial and political enterprises onto the same field of vision. Commercial enterprises operate with several or alienable property rights. The theory of market order explains how prices emerge out of economic interaction within the institutional framework of private property and freedom of contract. Those prices aid entrepreneurial judgments in organizing economic activities within society. Among other things, market prices enable the development of profit-and-loss accounting and help to establish market values for enterprises (Boettke 2001). In contrast, political enterprises operate with common property where prices do not emerge and market values cannot be established. Yet political enterprises must engage in economic calculation just as surely as do commercial enterprises. The chief executive of a politically organized park service, for instance, must decide how much land to devote to campgrounds relative to playgrounds, what kinds of equipment to install on playgrounds, and

many other facets of the enterprise's operation. At the same time, that park service must remain competitive within the budgetary process of the governmental unit of which it is a part. Commercial and political enterprises are subsystems within a common social system, but interaction among those enterprises generates societal tectonics and not placid equilibrium due to institutional divergences in their respective modes of entrepreneurial action.

Social Systems Theorizing: Mechanical vs. Creative

It is common to speak of economies and polities as "systems." But what kinds of systems are these analytical objects? Any system can be described as a network that contains a set of nodes along with edges that connect those nodes. For purposes of economic analysis, the nodes can be described variously as persons, enterprises, or agencies. In any case, the edges would denote patterns of relationship among the nodes. A network-based depiction is a useful way to characterize human population systems where each node operates with particular knowledge within a social division of labor (Hayek 1937, 1945). The performance properties of any system depend on both the capabilities possessed by the nodes and the patterns of connection among the nodes. Following Bertalanffy (1968) systems can be characterized as mechanical or creative.

Conventional economic theory characterizes an economic system as mechanical, which follows from the presumption that economic observations pertain to states of systemic equilibrium for a given set of preferences and techniques. Within this conceptual framework, it is a simple step to imagine a policy mechanic who inserts new data into the system at particular nodes, which results in a new equilibrium for the system. For a creative system, however, this conventional framework is incoherent because the relevant data don't exist prior to action but rather emerge through action. As Buchanan (1964a) explains, a market is not competitive through prior assumption but becomes competitive as action takes place through time. A policy mechanic can still act on particular nodes within the system, but creative agents can also by-pass those efforts in open-ended, creative fashion. Even if they face unchanged prices and incomes, creative agents can choose to alter their patterns of action, much as Ross Emmett (2006) explains in his analysis of Frank Knight's likely reaction to Stigler and Becker's (1977) argument to the contrary.

It is easy enough to characterize the difference between robotic and creative systems by using images that *Politics as a Peculiar Business* sets forth. A parade illustrates a robotic or equilibrium system. A parade is an organization. It might contain 5,000 participants stretched over two miles. All the same, those participants can be reasonably reduced to a center of gravity and represented as a point on a map, and with the progress of the parade tracked on the map as it moves along its route. By consulting the parade marshal's order of march, someone standing on a reviewing stand can tell which particular unit is passing by simply by consulting a watch. The only obstacle to doing this resides in the possible occurrence of some exogenous event such as the engine of a float breaking down, with the parade stopped until a tow-vehicle arrives. A parade is a robotic system, and is suitable for analysis by a theory of equilibrium because this theory captures the central features of a well-working robotic system where the constituent elements perform as they are programmed unless they break down.

A crowd of pedestrians leaving a stadium after an event illustrates a creative or nonequilibrium system. The crowd of pedestrians is orderly just as is a parade. The orderliness of the crowd, however, has different sources than the orderliness of the parade. In no reasonable way can a crowd be reduced to a parade. The members of the crowd are heading in different directions, whereas the members of the parade are heading in the same direction. The members of the crowd are also moving at different speeds, in contrast to the common speed of movement among the members of a parade. Within a creative system, the constituent elements perform as they choose. So-called policy mechanics cannot change a creative system as they choose because the elements in that system can in effect talk back through their creative actions. This situation means that systemic properties are not the province of so-called policymakers or policy mechanics. To the contrary, systemic properties are products of interaction among the constituent elements, as those interactions are guided and shaped by the actions of the participants.

The pedestrian crowd is still an orderly system, only that orderliness has different qualities and sources than does the parade. The parade is actually a form of public ordering in that the parade marshal establishes a set of rules to which the members of the parade agree to conform. Parades do, of course, differ in their observed qualities, and those differences reside in such matters as differences

in the musical and marching abilities of the members of the parade. For instance, some members of the parade might be able to keep in step better than other members, or play their music more deftly. For the pedestrian crowd, however, there is no common timing to the movement of the members, nor is there any common music that they play. Each member chooses his or her own pace and direction of movement. Orderliness within the pedestrian crowd resides in such matters as people having a general dislike for crashing into other people, in conjunction with having internalized some principles of courteous conduct in social settings. The pedestrian crowd is a privately ordered system that illustrates the merits of Edward Stringham's (2015) claim on behalf of private ordering.

Experiences with the prohibition of alcohol early in the 20th century and with current efforts to prohibit recreational drugs illustrate in sharp relief the difference between mechanical and creative systems. If a social system were a mechanical system, prohibition would forthrightly stop the activity in question, provided only that policy increased prices sufficiently. In response to policy-induced changes in prices, the individual agents would eliminate that activity from their repertoires and would revise their pattern of activity in light of the new set of prices. Within a creative system, however, policy-induced changes in prices also induce creative searches for new paths of commerce that are accompanied by new sets of prices. Within a creative system, there is no reason to expect individual agents to cease particular activities just because other agents, political or otherwise, seek to make them do so. In other words, prohibition, then or now, does not eliminate the prohibited activity from agents' repertoires. All prohibition accomplishes, and all it can accomplish, is to empower one set of agents to try to prevent activities that another set of agents desires to undertake. How successful that effort will be will depend in large measure on the creative responses of those other agents.

Market Transactions and Parasitical Political Calculation

Commercial and political enterprises both operate within the same societal territory and confront the same kinds problems of engaging in successful economizing action. Every actor is engaged in seeking to attain and maintain desired states of existence in the presence of other agents seeking to do the same. All enterprises are thus

engaged in seeking to make value-increasing choices as the executives of the enterprise appraise their situation. In this respect, Buchanan's (1969) analysis of the reciprocal relationship between cost and choice tells us that cost always pertains to the options for choice that an acting individual faces. If you want to explain the choice of an enterprise, whether commercial or political, the relevant situs of cost pertains to the options a choosing agent faces and not some abstract construction of enterprise-level accounting. While agency costs are always present in corporate situations, the transferability of ownership and the resulting establishment of corporate valuation presents tools to limit agency costs in commercial enterprises relative to political enterprises, as Meckling and Jensen (1976) and Fama (1980) explain.

Both commercial and political enterprises may be reasonably regarded as profit-seeking enterprises. After all, to seek profit means nothing more than to avoid loss. Loss, moreover, means shifting resources from higher-valued to lower-valued uses. Profit seeking, in other words, is simply different wording applied to mutual gains from trade. Commercial enterprises can remove profits directly from the enterprise. Political enterprises can't. This doesn't mean that profits can't be removed from political enterprises, nor does it mean that profits don't result from the action of political enterprises. It means only that profit from political enterprises follows some indirect channel. In this respect, Pauly and Redish (1973) explain how nonprofit hospitals can serve as conduits for the transfer of profits to physicians. As a general matter and in the spirit of Pauly and Redish, politically generated profits are removed through transactions with associated commercial enterprises, resembling money laundering in the process.

Political enterprises lack some significant tools for economizing action because they operate under nonalienable ownership. As a result, they carry no market value. These enterprises still seek to be successful within their arenas of action. Those arenas, however, are governed by budgeting and not commercial transaction. This situation leads to a sequential pattern for interaction among commercial and political enterprises: commercial enterprises locate profit opportunities, after which political enterprises enter that commercial territory and do their business as peculiar forms of commercial enterprise. Within this environment, political enterprises must act parasitically on commercial enterprises to secure

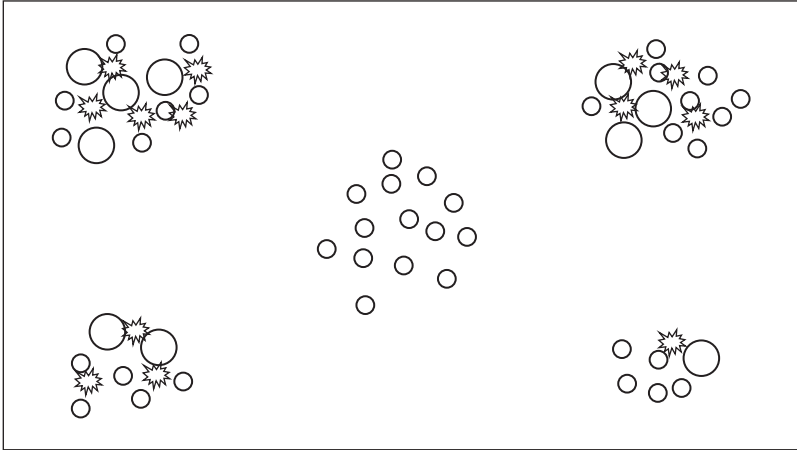
navigational guidance and financial support. By navigational guidance, I mean that the location of commercial enterprises provides information for the location and substance of political programs. By financial support, I mean that political enterprises operate within the fiscal commons (Wagner 2007) and not within the market, and therefore derive their revenues through making parasitical attachments to market transactions. Hence, political enterprises are doubly indebted to commercial enterprises: commercial enterprises provide navigational guidance and revenues.

By parasitical political pricing I mean something similar to what Maffeo Pantaleoni (1911) set forth in treating a system of political prices as being a parasitical attachment to a system of market prices. Pantaleoni asked us to imagine a world inhabited by two bazaars, one a market bazaar and the other a political bazaar. In the market bazaar, vendors sell their wares at market prices which Pantaleoni, in keeping with the economic theory of his time, treated as being equal to marginal costs of production. In the political bazaar, vendors sell their wares at what Pantaleoni described as political prices, which depends on the particular tax system in place. Pantaleoni worked with a flat tax on all income, as did Buchanan (1964b) in creating a model in which all members of a society would agree on the same rate of political output. The parasitical quality of political pricing comes into play because political enterprises do not generate their revenues directly from the sale of services, but rather derive them from making parasitical attachment to commercial transactions.

Figure 1 illustrates the location of commercial and political enterprises in a manner consistent with Pantaleoni's theory of parasitical political pricing. It illustrates the simple point that commercial enterprises first locate profitable opportunities, and then political enterprises locate nearby to complete the parasitical relationship. The figure describes a network of enterprises that inhabit different locations on some abstract notion of commodity space.

Figure 1 shows two types of nodes. The circles are commercial enterprises, with some being large and others small. The irregular spheres denote political enterprises. The enterprises are located in five distinct parts of the abstract commercial space to represent different types of commercial activity. Suppose momentarily that the political enterprises are removed from view. We can liken the resulting distribution of firms over commercial space to the distribution of

FIGURE 1
 ENTERPRISE LOCATION AND PARASITICAL
 POLITICAL CALCULATION



NOTE: circles = commercial enterprises; irregular spheres = political enterprises.

prospectors among potential sites. Just as prospectors will clump in some areas and avoid other areas, so will commercial enterprises.

Figure 1, with the irregular spheres suppressed, is a portrait of the location of enterprises in commercial space. The center shows 14 small firms. The largest concentration of large firms is in the northwestern part of Figure 1, where four of the 11 firms are large. At the northeastern part of Figure 1, three of the 12 firms are large. At the southwest, two of the seven firms are large, while in the southeast there is one large firm among the six firms. Figure 1 represents the location of commercial enterprises in commercial space, and that pattern would be explained through the economizing efforts of those enterprises. Figure 1 with political enterprises suppressed from view characterizes the use of knowledge within a framework of wholly private ordering. The locational pattern of enterprises in Figure 1 provides information about the pattern of human concerns. In this respect, firms located in the northwest and the northeast would seem to be particularly strongly implicated in dealing with matters of great human concern and interest. These are the activities that would attract attention in national newscasts, whereas going on in the southeastern part of commercial space

would likely attract little attention. Enterprises located in the center of Figure 1 would attract even less attention reflecting that those enterprises deal with matters of highly local concern, such as cutting grass and painting houses.

How might political enterprises be incorporated into this map of enterprises, keeping in mind their necessarily parasitical character in light of their lack of tools for economic calculation? Commercial firms identify locations that are of most significant human interest and concern, after which political enterprises enter the territory looking for margins along which they can develop clientele and extract support. Figure 1 distributes 13 political enterprises within the commercial territory portrayed there. Those enterprises are placed systematically according to market principles of economic calculation. The center of Figure 1 has no political enterprises, which indicates that there is no political profit to be gained by locating there. In contrast, five political enterprises are located in the northwest, to indicate that much political profit is thought to be attainable there. The northeast also contains much potential for political profit, as indicated by the location there of four political enterprises. The southwest is commercially less exciting than the northeast, as illustrated by only three political enterprises being located there. The southeast is even sleeper commercially speaking, and so contains only one political enterprise.

Most of the political enterprises locate adjacent to large firms, and sometimes straddle large and small firms. When public ordering appears, dyadic exchange gives way to triadic exchange (Podemska-Mikluch and Wagner 2013). With dyadic exchange, no matter how numerous the participants, commercial relationships are agreeable to the participants. With triadic exchange, however, there is agreement among a dominant subset of participants to prey upon other members of society, using ideological formulations to soften the appearance of preying. With private ordering, a failed enterprise will liquidate its assets or otherwise reorganize, depending on what the participants are able to work out among themselves. With the triadic exchange that public ordering enables, the executives of a failing enterprise might be able to secure the cooperation of a political enterprise in order to establish a class of forced investors to keep the enterprise going without reorganization.

Private vs. Public Ordering and the Generation of Societal Tectonics

What are the economic properties of a social system where commercial and political enterprises occupy similar territory but are constituted through the different operating principles that the distinction between dyadic and triadic exchange conveys? Jane Jacobs (1992) advanced the reasonable claim that well-working societies required some reasonable mix of commercial and guardian activities, with the two operating in different domains of action. She also recognized that excessive commingling between commercial and guardian activities could have socially troubling consequences—what she described as “monstrous moral hybrids.” The resulting admixture of private and public ordering can create societal tectonics in economic regions where the two forms of ordering commingle, generating economic rubble in the process. In a similar vein, William Niskanen (1978) explains that classical Greek theorists, unlike their successors two millennia later, recognized that democracy was a difficult form of government to maintain because of its strong tendency to erode the sense of personal responsibility that is bound up with the principle of private property and residual claimancy.

The history of urban transportation since the time of the jitneys in the early 20th century (Eckert and Hilton 1972) provides a good illustration of a situation that is still in play today with controversy over Uber and Lyft. Suppose initially that transportation within a city is organized wholly through private ordering. At some subsequent moment a city council sponsors a municipal bus enterprise. This could have been organized as a private enterprise with the members of the city council being the shareholders. But instead it is organized as a municipal enterprise. While the enterprise doubtlessly will charge fares to riders, equally surely the municipal enterprise will not be able to survive in the face of truly open competition with private enterprises because it lacks the managerial tools that private property enables commercial enterprises to develop. The municipal enterprise, however, will be exempt from taxes that the private enterprises must pay. While the municipal enterprise might be required to charge fares to cover its direct operating expenses, it might also receive its buses through the city budget, enabling the municipal enterprise to appear more competitive than otherwise. Should the

municipal enterprise still face competitive difficulty, the city could regulate private competitors as in abolishing competitive forms of transportation. Short of doing this, it could impose costs on commercial enterprises through regulation, again enabling the municipal enterprise to appear more competitive. In any case, the insertion of public ordering into a system grounded on private ordering will generate the social equivalents of earthquakes where the supporters of political enterprises clash with the supporters of commercial enterprises. Earthquakes create rubble, of course, and how that rubble is removed has significant social implications. One possible form of removal would be to replace public with private ordering, as in privatizing the municipal bus enterprise. Another form of removal would promote some fusion between private and public ordering by inserting politically expressed interests into the operation of commercial enterprises.

Alternatively, within modern credit markets private and public ordering are deeply entangled, as Smith, Yandle, and Wagner (2011) explore with reference to the depression-era National Recovery Act and the more recent Troubled Asset Relief Program. Along many margins of activity, public ordering now governs the conduct of commercial lending enterprises. Entangled political economy captures the idea that political enterprises are involved in commercial activity, just as commercial enterprises are engaged in political activity. Rather than engaging in debates over whether the events of 2008 were caused by excessive or insufficient regulation, entangled political economy would explain such events as reflecting *systemic properties* of a system in which businesses engage in politics and politicians engage in business. When political agencies become involved in directing the composition of asset portfolios and when financial institutions engage in negotiation over meeting regulatory requirements, what Jacobs (1992) describes as monstrous moral hybrids come into the commercial foreground to do their work.

Within a framework of private ordering, debtor-creditor relationships are governed by principles of contractual obligation (Fried 1981). Within this framework, borrowers and lenders alike regard their relationship as mutually advantageous. Not all debtor-creditor relationships work out to mutual advantage, often because of unanticipated changes that affect one of the two parties. A borrower might have thought a new commercial venture would be more profitable than it turned out to be. In consequence, the debtor might have

declared bankruptcy after having tried unsuccessfully to discharge the debt by liquidating some other assets. Given the creditor's place in the bankruptcy pecking order, the creditor's loss on this transaction might have upset other commercial relationships that called for renegotiation. Regardless of the simplicity or complexity of those renegotiations, the legal principles that accompany private ordering provide clarity as to just who it is that owes what to whom in sorting out the pattern of obligations that accompany the collapse of earlier expectations.

With private ordering, promises are exchanged among specific individuals, and controversy over fulfillment of those promises resides with those individuals. This is so with a modest amendment for promises involving corporate bodies. A corporate officer who occupies one side of a promise might leave the corporation before the promise is discharged, leaving the promise as a general obligation of the corporation. To the extent corporate governance operates to limit agency costs, promises made by different corporate officials have much the same effect as promises made by a single official, as De Angelo (1981) and Makowski (1983) explain.

The situation changes sharply when public ordering enters into debtor-creditor relationships (Wagner 2012a). Whether we are talking about some \$20 trillion of official public debt or some \$100 trillion of unfunded liabilities, there is no clarity about who owes what to whom. These obligations point to past promises of an indefinite form that cannot be kept going forward. Public debt doesn't entail specific promises and obligations. Public debt operates in the realm of status and not that of contract.

Present magnitudes point toward a form of *systemic lying* of large magnitude. When viewed from a contractual perspective, public debt denotes a gap between the promises that have been made to people in their capacities as taxpayers and the promises made to people in their capacities as beneficiaries. I describe this as "systemic lying" to indicate that it is a systemic quality of a system of entangled political economy wherein principles and concepts derived from private ordering are extended to public ordering. It is, after all, not necessary for governments to use public debt. A government that found revenues falling short of expenditures could finance the gap by taxing. If it did this, some taxpayers doubtlessly would borrow to discharge their liability rather than paying cash. Hence, the increased spending would induce increased borrowing within the society, but

this indebtedness would be organized through private ordering and not through public debt and the public ordering this entails.

Public ordering uses the language of contract without the practice of contracting. In consequence, public debt doesn't reflect promises in the same sense that promises are made under private ordering. With public ordering, debtor-creditor relationships are based on status. While bondholders agree to buy bonds and so agree to become creditors, there is no agreement behind the acquisition of the status of being a debtor. One could, of course, claim that taxpayers accept the status of being indebted by virtue of their being a citizen of the indebted government, but this is just a particular ideology talking, which claims that any democratic process reflects consensus among everyone caught within the web of governmental action.

The different operating properties of private and public ordering are captured crisply in Vilfredo Pareto's ([1923] 1935) distinction between logical and nonlogical action. This distinction does significant work in explaining the generation of societal earthquakes of varying magnitudes, as Patrick and Wagner (2015) explore. By logical action, Pareto meant actions that correspond to the template of scientific experiments, which he described as the logico-experimental method. A firm that introduces a new line of products under the presumption that this will increase the value of the firm can test this conjecture against actual market valuations. Political environments accommodate no such testing, and yet those environments are fiercely competitive all the same. Competition within public ordering will manifest itself in ideological constructions; competitors will seek to advance ideological formulations that resonate more strongly with the relevant electorate than other formulations. Without private property and residual claimacy, choice evaporates and is replaced by expressions of sentiment.

Conclusion

My primary interest in writing *Politics as a Peculiar Business* was to contribute to the construction of an analytical framework in which economics would be construed as a genuine social science, the starting point of which would be the formal ubiquity of economizing action. In the process, I hoped to contribute to the development of a science of association along similar lines to Michael Oakeshott (1975) and Vincent Ostrom (1997).

In pursuing this effort, I have located political activity within the sphere of enterprise association, recognizing that political enterprises have peculiar qualities that manifest themselves throughout a society. Those peculiar qualities, moreover, bring to life Benjamin Franklin's reported retort to a lady who asked him, at the close of the American Constitutional Convention in 1787, what kind of government the Convention had established. Franklin is alleged to have replied, "a republic if you can keep it." It's clear that Franklin thought in terms of open or evolutionary processes and not in terms or closed states of equilibrium. It's equally clear that what was established as a constitution of liberty has moved significantly toward becoming a constitution of servility, of the democratic sort that Tocqueville described in his chapter on democratic despotism in *Democracy in America*.

Whether the reach of government can be stopped and even reversed is an open question that awaits future observation. There is very little that governments do that cannot be accomplished through nongovernmental action, though with different distributional consequences and with different people occupying positions of honor. Within a system based on private ordering, entrepreneurs and other creators are the principle holders of posts of honor within a society. These people hold center stage in human drama, with political officials serving mainly as stagehands. With public ordering the positions are reversed, with political officials coming to occupy significant posts of honor and gain control of center stage. But the human drama is continuing and patterns of human association are continually changing. My presumption in writing *Politics as a Peculiar Business* is that locating economics as a genuine social science, and politics as a peculiar form of commercial enterprise, will contribute to the emergence of a better appreciation of the constitution of liberty we once had, relative to the constitution of servility that has been growing over the past century or so.

References

- Becchio, G. (2014) "Carl Menger on States as Orders, Not Organizations." *Advances in Austrian Economics* 18: 55–66.
- Bertalanffy, L. (1968) *General Systems Theory*. New York: George Braziller.
- Boettke, P. J. (2001) *Calculation and Coordination*. London: Routledge.

- Buchanan, J. M. (1964a) "What Should Economists Do?" *Southern Economic Journal* 30: 213–22.
- _____ (1964b) "Fiscal Institutions and Efficiency in Collective Outlay." *American Economic Review*, Papers and Proceedings 54 (May): 227–35.
- _____ (1969) *Cost and Choice*. Chicago: Markham.
- Cheung, S. N. S. (1973) "The Fable of the Bees: An Economic Investigation." *Journal of Law and Economics* 16: 11–33.
- Coase, R. H. (1974) "The Lighthouse in Economics." *Journal of Law and Economics* 17: 357–76.
- De Angelo, H. (1981) "Competition and Unanimity." *American Economic Review* 71: 18–27.
- Eckert, R. D., and Hilton, G. W. (1972) "The Jitneys." *Journal of Law and Economics* 15: 293–325.
- Emmett, R. B. (2006) "Die Gustibus Est Disputandum: Frank H. Knight's Response to George Stigler and Gary Becker's 'Die Gustibus Non Est Disputandum.'" *Journal of Economic Methodology* 13: 97–111.
- Eusepi, G., and Wagner, R. E. (2011) "States as Ecologies of Political Enterprises." *Review of Political Economy* 23: 573–85.
- Fama, E. (1980) "Agency Problems and the Theory of the Firm." *Journal of Political Economy* 88: 327–49.
- Fried, C. (1981) *Contract as Promise: A Theory of Contractual Obligation*. Cambridge, Mass.: Harvard University Press.
- Hayek, F.A. (1937) "Economics and Knowledge." *Economica* 4: 33–54.
- _____ (1945) "The Use of Knowledge in Society." *American Economic Review* 35: 519–30.
- Hirshleifer, J. (2001) *The Dark Side of the Force: Economic Foundations of Conflict Theory*. Cambridge, U.K.: Cambridge University Press.
- Hughes, J. R. T. (1977) *The Governmental Habit: Economic Controls from Colonial Times to the Present*. New York: Basic Books.
- Jacobs, J. (1992) *Systems of Survival*. New York: Random House.
- Johnson, D. B. (1973) "Meade, Bees, and Externalities." *Journal of Law and Economics* 16: 35–66.
- Kahn, A. E. (1966) "The Tyranny of Small Decisions: Market Failures, Imperfections, and the Limits of Economics." *Kyklos* 19: 23–47.

- Kirman, A. P. (1992) "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives* 6: 117–36.
- Knight, F. H. (1960) *Intelligence and Democratic Action*. Cambridge, Mass.: Harvard University Press.
- Krause, M. (2015) "Buoys and Beacons in Economics." *Journal of Private Enterprise* 30: 45–59.
- Makowski, L. (1983) "Competition and Unanimity Revisited." *American Economic Review* 73: 329–39.
- McCormick, R. E., and Tollison, R. D. (1981) *Politicians, Legislation, and the Economy*. Boston: Kluwer.
- Meckling, W. H., and Jensen, M. C. (1976) "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure." *Journal of Financial Economics* 3: 305–60.
- Milonakis, D., and Fine, B. (2009) *From Political Economy to Economics*. London: Routledge.
- Niskanen, W.A. (1978) "The Prospect for Liberal Democracy." In J. M. Buchanan and R. E. Wagner (eds.), *Fiscal Responsibility in Constitutional Democracy*, 157–74. Leiden, Netherlands: Martinus Nijhoff.
- Oakeshott, M. (1975) *On Human Conduct*. Oxford, U.K.: Clarendon Press.
- Ostrom, V. (1997) *The Meaning of Democracy and the Vulnerability of Societies: A Response to Tocqueville's Challenge*. Ann Arbor: University of Michigan Press.
- Paganelli, M. P. (2014) "Adam Smith and Entangled Political Economy." *Advances in Austrian Economics* 16: 37–54.
- Pantaleoni, M. (1911) "Considerazioni sulle proprietà di un sistema di prezzi politici." *Giornale degli Economisti* 42: 9–29, 114–33.
- Pareto, V. ([1923] 1935) *The Mind and Society*. New York: Harcourt Brace.
- Patrick, M., and Wagner, R. E. (2015) "From Mixed Economy to Entangled Political Economy: A Paretian Social-theoretic Orientation." *Public Choice* 164: 103–16.
- Pauly, M. V., and Redish, M. R. (1973) "The Not-for-Profit Hospital as a Physicians' Cooperative." *American Economic Review* 63: 87–99.
- Persson, T., and Tabellini, G. (2000) *Political Economics: Explaining Public Policy*. Cambridge, Mass.: MIT Press.
- Podemska-Mikluch, M., and Wagner, R. E. (2013) "Dyads, Triads, and the Theory of Exchange: Between Liberty and Coercion." *Review of Austrian Economics* 26: 171–82.

- Robbins, L. (1952) *The Theory of Economic Policy in English Classical Political Economy*. London: Macmillan.
- Samuels, W. J. (1966) *The Classical Theory of Economic Policy*. Cleveland: World.
- Schmitt, C. ([1932]1996) *The Concept of the Political*. Chicago: University of Chicago Press.
- Smith, A.; Yandle, B.; and Wagner, R. E. (2011) “A Theory of Entangled Political Economy, with Application to TARP and NRA.” *Public Choice* 148: 45–66.
- Stigler, G. J., and Becker, G. S. (1977) “De Gustibus Non Est Disputandum.” *American Economic Review* 67: 76–90.
- Streit, M. E. (1992) “Economic Order, Private Law, and Public Policy.” *Journal of Institutional and Theoretical Economics* 148: 675–704.
- Stringham, E. (2015) *Private Governance: Creating Order in Economic and Social Life*. New York: Oxford University Press.
- Tollison, R. D., and Wagner, R. E. (1991) “Romance, Realism, and Economic Reform.” *Kyklos* 44: 57–70.
- Tullock, G. (1965) *The Politics of Bureaucracy*. Washington: Public Affairs Press.
- Wagner, R. E. (2007) *Fiscal Sociology and the Theory of Public Finance*. Cheltenham, UK: Edward Elgar.
- _____ (2010) *Mind, Society, and Human Action: Time and Knowledge in a Theory of Social Economy*. London: Routledge.
- _____ (2012a) *Deficits, Debt, and Democracy: Wrestling with Tragedy on the Fiscal Commons*. Cheltenham, U.K.: Edward Elgar.
- _____ (2012b) “Viennese Kaleidics: Why It’s Liberty More than Policy that Calms Turbulence.” *Review of Austrian Economics* 25: 283–97.
- _____ (2015) “Welfare Economics and Second-Best Theory: Filling Imaginary Economic Boxes.” *Cato Journal* 35 (1): 133–46.
- _____ (2016) *Politics as a Peculiar Business: Insights from a Theory of Entangled Political Economy*. Cheltenham, UK: Edward Elgar.
- Wildavsky, A. (1975) *Budgeting: A Comparative Theory of Budgetary Process*. Boston: Little, Brown.
- Young, R. A. (1991) “Tectonic Policies and Political Competition.” In A. Breton et al. (eds.), *The Competitive State*, 129–45. Dordrecht, Netherlands: Kluwer Academic Publishers.

A CRITIQUE OF PROPOSALS TO RAISE THE FED'S INFLATION TARGET

William T. Gavin

During the last seven years, the unemployment rate fell from 10 percent to less than 5 percent, but policymakers say that there is still an underutilization of labor resources. Why? Because inflation is below the Federal Reserve's 2 percent target and GDP is below official estimates of potential GDP. Normally, in this situation the Federal Open Market Committee (FOMC) would lower short-term interest rates in order to stimulate aggregate spending. However, short-term interest rates are near zero—as low as they can go when people have the alternative of holding cash. Those who want to use lower interest rates to stimulate the economy also want a higher inflation target so that when the economy is at full employment nominal interest rates will be higher, and when something bad happens, policymakers will have more flexibility to lower interest rates before hitting the zero lower bound.

In a *Financial Times* interview on April 20, 2015, Federal Reserve Bank of Boston President Eric Rosengren called on his fellow policymakers at the Fed and around the world to consider raising their inflation targets: “As we learn more about the real interest rate potentially being lower, we may at least want to have a broader debate about whether we have set the inflation targets too low” (Fleming 2015).

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

William T. Gavin is a former Vice President and Economist in the Research Department at the Federal Reserve Bank of St. Louis. The author thanks Kevin Kliesen, Chris Neely, and Dan Thornton for helpful comments on an earlier draft of this article.

The rationale for a higher inflation target does not depend on there being a long-run tradeoff between inflation and unemployment. It is about countercyclical policy, the desire to have plenty of flexibility for the Fed to lower interest rates when there is a string of bad news. Others, who do believe that there is a long-run tradeoff between inflation and unemployment, also call for *permanently* higher inflation. They think that inflation aids labor market adjustments when the demand for labor falls both for an individual firm and for the economy overall.

This article explains why a higher inflation rate is not a good idea. As a cyclical policy, it would do more harm than good and, as a permanent policy, would not take us to a better economy. I begin by reviewing the calls for more inflation, explaining the rationale that is put forward for each case. Next, I lay out the reasons why raising the inflation target would be a bad idea. In particular, raising the inflation target damages the value of inflation targeting as a nominal anchor. Moreover, I summarize what we have learned about the costs of inflation, both anticipated and unanticipated. Finally, I explain why the perceived benefits suggested by advocates of higher inflation are ephemeral and not likely to be achieved in practice.

Calls for Higher Inflation

Some economists have recommended that the Federal Reserve raise its long-run inflation target from 2 percent to 4 percent—not because they think this would be useful in the near term, but rather because they think a 4 percent inflation economy would perform better than a 2 percent inflation economy. To the best of my knowledge, this argument was first made by Summers (1991) at a conference on the optimal inflation target sponsored by the Federal Reserve Bank of Cleveland and the *Journal of Money, Credit, and Banking* in October 1990. He specifically commented on proposed legislation, House Joint Resolution 409, which would have mandated a zero inflation target for the Federal Reserve. More recently, Ball (2014) and Blanchard, Dell’Ariccia, and Mauro (2010) have argued that 2 percent steady state inflation is too low and causes market interest rates to hit zero too often, frustrating Fed attempts to promote full employment.

The reasoning is simple and rather mechanical, involving two ideas. The first is simply that the market interest rate is the sum of

the real return and the expected inflation rate that should be equal to the inflation target. A higher inflation target during normal times means higher inflation expectations and a higher market interest rate, giving the Fed more room to lower the policy rate when a recession begins. The second idea, open to debate, is that lower interest rates will lead to more aggregate demand and more output. Ball calculates that “if interest rates had been two points lower during 2009, output in 2010 would have been 2 percent higher.” Moreover, he argues that “the output gain for 2013 would be 5.9 percent, and the cumulative gain over 2010–13 would be 16.4 percent of annual output” (Ball 2014; 4–5). Similar calculations can be found in Blanchard, Dell’Ariccio, and Mouro (2010).

Well before the 2008–09 financial crisis, Akerlof, Dickens, and Perry (1996 and 2000) argued in favor of higher inflation targets. They want a higher inflation target in order to “grease” the labor market. They assume that workers are ignorant about the effects of inflation on the market for their labor services. These workers would rather have the purchasing power of their paycheck cut by inflation than take a direct cut in their nominal take-home pay. They argue that workers fail to understand that inflation increases wages elsewhere in the economy, so that other wages will be rising while theirs are held constant. Their relative wage will fall just as it does in the case where they get an explicit wage cut. They calculate that “the difference in the sustainable rate of unemployment between operating with a steady 3 percent inflation rate and a steady zero percent inflation rate is estimated as 1 to 2 percentage points” (Akerlof, Dickens, and Perry 1996: 51). Essentially, these authors are arguing that higher inflation will reduce conflict in labor markets and lead to higher aggregate output. Ironically, they are arguing that, by confusing individuals about the relative price of labor, the higher inflation will improve economic efficiency.

There were also calls for a *temporary* increase in the inflation target made early in the financial crisis. The sharp decline in housing and other asset prices in 2008 and 2009 left many highly leveraged homeowners and investors underwater with debt levels that were thought to be a drag on economic recovery. Kenneth Rogoff recommended that the Fed pursue 6 percent inflation for a couple years to help such debtors (see Evans-Pritchard 2009). In 2011, during an NPR interview, he recommended that the Fed print money until the inflation rate reached 5 percent (National Public Radio 2011).

Since most debt is fixed in nominal terms, the policy would intentionally shift wealth from creditors to debtors. Rogoff (2014) clearly intended this to be a temporary policy as he argues against Ball's permanent 4 percent inflation target.

The Inflation Target as a Nominal Anchor

The adoption of inflation targets to stabilize the purchasing power of paper money evolved gradually, after almost two decades of failing attempts to implement money supply targets. The need for a nominal anchor became apparent in the late 1960s and early 1970s as the modified dollar/gold standard adopted at Bretton Woods began to come apart. The U.S. dollar lost its (admittedly weak) anchor to gold and the result was high and uncertain inflation.¹ Initially, the government used wage and price controls to try to control inflation. But the economic distortions were obvious. There were many shortages and non-price rationing schemes such as lines at gasoline stations. The price controls were abandoned and economists debated about how to repair the damage and implement a new anchor for the dollar. Monetarists, led by Milton Friedman, advocated a fixed growth rate for the money supply. In 1976, Congress passed a resolution requiring the Fed to announce targets for the money supply. Congressman Ron Paul (R-TX) and Lewis Lehrman (1982) called for a return to the gold standard. Others recommended adoption of targets for nominal GDP growth. By targeting nominal GDP at the estimated growth rate of real GDP, the cost of living would rise or fall with fluctuations in productivity.² There was little support for inflation targeting because, as Milton Friedman said in his 1967 presidential address to the American Economic Association, the control mechanism linking monetary policy actions to the price level was thought to be too uncertain. He argued there were long and variable lags between monetary policy actions and their effect on inflation. Thus, attempts to target inflation would destabilize the economy and the price level. He admitted that "Perhaps, as our understanding of monetary phenomena

¹By the late 1970s, high inflation was considered the number one policy problem in the United States. See Gallup Poll results reported in Hibbs (1982).

²Looking back to a pre-Keynesian era, Selgin (1990) revisits the classical rationale for a "productivity standard" that looks much like a form of nominal GDP targeting. He also provides a survey of the debate about nominal GDP targeting.

advances, the situation will change. But at the present stage of our understanding, the long way around [money supply targeting] seems the surer way to our objective [price stability]" (Friedman 1968: 15).

Following Friedman's advice, on October 6, 1979, Fed Chairman Paul Volcker announced that the Fed would stop the daily targeting of interest rates and begin targeting bank reserves directly in an attempt to achieve targets for the money supply. Although money supply growth actually rose and became more volatile, attempting to achieve money supply targets drove money market interest rates to almost 20 percent. Interest rates remained high as unemployment rose to 10.4 percent and inflation in the Consumer Price Index (CPI) fell from double-digit levels to 2.5 percent (year over year) in January 1983. Other nations followed the United States in raising interest rates and lowering inflation, some more successfully than others. However, New Zealand and Canada continued to have high inflation, which led their politicians and central banks to adopt inflation targets despite the advice of Milton Friedman and with little support from macroeconomists. There was a big surprise when inflation came down quickly and stabilized around the targets in both countries.

As Friedman was making the case for money supply targeting based on long and variable lags, our knowledge was advancing in the form of the "rational expectations revolution." Friedman advocated for money supply targets because he imagined the central bank could control money in both the short and long run. He was worried about long and variable lags in the control mechanism. In contrast, the theory of rational expectations predicts that people will forecast inflation using all the information they have about monetary policy.³ The Fed uses the basic premise of rational expectations in its public policy statements that are explicitly aimed at influencing people's expectations about future policy.

Because of inflation targeting's almost immediate success, by the mid-1990s economists scrambled to show how and why inflation targeting worked. It worked indirectly by coordinating the public's inflation expectations around a common number. By deciding on and

³A useful case study can be found in Sargent and Zeira (2011). They document the case of an Israeli government promise, in October 1983, to bail out bank stockholders, with the payout to be made four or five years into the future. Inflation began to rise rapidly on the announcement because people believed the promise was credible.

announcing an inflation target, the central bank gave the public the information needed to make good decisions about wages, prices, and long-term financial plans. Contrary to Friedman's concern, inflation targeting did not destabilize the price level or the economy. Rather, it led people to make contracts, follow pricing policies, and make financial plans that incorporated the central bank's objective. Inflation targeting worked with standard interest rate operating procedures because the interest rate includes a premium for inflation expectations. It turns out that stabilizing the policy interest rate around an appropriate level stabilizes the inflation rate. Influencing expectations of future policy has given central banks indirect, but substantial, control over the trend rate of inflation. It is not, however, direct control that can be used to manipulate the inflation rate within a business cycle frequency as advocated by Rogoff.

As more central banks adopted inflation targeting during the 1990s, they converged around a 2 percent target. Why 2 percent? Some believed Summers's (1991) argument about business cycle stabilization and the zero lower bound. Others thought the inflation target could be insurance against the sort of deflation that was associated with the Great Depression. In the United States, Fed Chairman Alan Greenspan argued that the appropriate target would be zero inflation, "properly measured" (FOMC 1996: 51). He cited the Boskin Commission's (1996) finding that the CPI was biased upward, maybe as much as 1 to 2 percent. Because of the perceived bias in the CPI, Greenspan thought that 2 percent inflation target would be closer to 0 percent in a true cost-of-living index. Finally, and perhaps most importantly, by choosing 2 percent the Fed was following the lead of central banks that had already adopted explicit inflation targets.

In practice, central banks around the world have implemented inflation targeting by choosing the same target year in and year out. This process has led the public to believe that inflation-targeting central banks have a long-term inflation objective. So, inflation targeting has been successful because the central bank decides on and announces an inflation objective, not because of a change in its day-to-day behavior in money markets, or in the way it reacts to news about unemployment or real GDP at policy meetings.

The Fed did not announce an explicit numerical objective for inflation until January 25, 2012. Although the Fed did not publicly announce any inflation target before 2012, there is evidence from

16 years earlier that the FOMC had reached a consensus that 2 percent was the appropriate objective (Heller 2015, FOMC 1996). Throughout much of this period, speeches of policymakers and Fed watchers noted that the Fed had a “comfort zone” for inflation that was close to 2 percent. From 1996 through 2014, the average inflation rate in the CPI was just a bit over 2 percent, and the long-run forecasts of business economists have been centered on 2 percent. It appears that the 2 percent inflation objective is working to anchor the U.S. dollar.

Raising the inflation target now would be a setback for the progress that has been made over the past three decades—allowing the anchor to drift would disrupt expectations and risk a return to the high and uncertain inflation of the 1970s. By deciding on and announcing a numerical objective for inflation, the central bank is providing an anchor for the dollar—an anchor based on a price path for an evolving basket of consumer goods. The anchor encourages people to form expectations around a common trend. The central bank does not control the price level directly, but indirectly by creating information that makes it more likely that people will price things in a way that incorporates the central bank’s inflation objective.

The Costs of Inflation

The costs of inflation are associated with both anticipated and unanticipated inflation.⁴

Anticipated Inflation

Since currency pays no interest, anticipated inflation acts as a tax on cash balances and causes people to spend resources economizing on cash balances. However, this resource cost is considered to be small in most studies.⁵ The most important, yet difficult to quantify, cost of inflation is the damage it does to the monetary standard that we use to measure economic value. Anything that adds uncertainty to

⁴See Kessel and Alchian (1962) for an early and prescient exposition of these ideas. Marty and Thornton (1995) summarize the objections to the moderate inflation rate recommended by Summers (1991).

⁵Fischer (1981) and Dowd (1994) summarize the costs of inflation, including quantitative estimates of the cost of inflation associated with the inflation tax on cash balances.

the unit of account makes the price system less effective and causes the economy to operate less efficiently. Consider an analogy with the distance standard used to measure a unit of length. Suppose that the unit of length were to grow 2 percent a year so that all construction tools, blueprints, and the size of building materials had to be continuously adjusted for the changing unit of length. The example is ridiculous because the costs would be outrageous. But inflation creates the need for accountants to develop systems that can cope with a continuously adjusting unit of account. Avoiding such costs is the reason why the government adopts standards, including a standard for money. Ongoing inflation, even if fully anticipated, degrades the operation of the price-clearing mechanism and makes comparisons among economic values across time and markets less reliable.

The most obvious and quantifiable costs of anticipated inflation come from the interaction of inflation with the tax code. A well-known distortion is the tax deductibility of the interest component in the home mortgage payment. Higher inflation raises mortgage rates and, thus, raises the value of this deduction. This tax subsidy reduces the real cost of home ownership and likely contributes to overinvestment in housing.

Early research by Martin Feldstein analyzes the interaction of inflation with the tax code in the era before the Economic Recovery and Tax Act of 1981 (ERTA)—before the partial indexation of the tax code. Even then, the highest costs of the interaction of the tax code with inflation occurred with taxes on interest and capital income. Since the code continues to tax nominal interest and capital gains, much of the analysis still applies today. Feldstein (1976) explains how inflation interacts with the tax code to raise the cost of capital while lowering the return to household savings; Feldstein (1980) shows how expected inflation interacts with the tax code to reduce the share price per dollar of pretax earnings; and Feldstein (1982) demonstrates that inflation–tax code interaction distorts the measurement of profits, interest payments, and capital gains. Although this early work did not consider the effects of the Reagan tax reform, more recent work does. Feldstein (1997) and Abel (1997) provide further support for the idea that relatively moderate changes in the inflation target can have relatively large effects on welfare through the interaction of inflation with the tax code.

Altig and Carlstrom (1991) show that the imperfect indexation of personal income tax brackets in ERTA 1981 still left measurable

welfare losses associated with high inflation. In a comprehensive analysis of the U.S. tax code, Bullard and Russell (2004) estimate that the welfare costs of moving from a 2 percent to a 4 percent inflation trend would be expected to cause welfare losses equal to about 2 percent of one year's output. They attribute "the lion's share of the welfare cost of higher inflation . . . to its tendency to produce a downward shift in the entire structure of real interest rates, both before and after taxes" (Bullard and Russell 2004: 62).

Bullard and Russell (2004) do not include the capital gains tax. Gavin, Pakko, and Kydland (2007) analyze a tax on realized capital gains that interacts with high inflation. They show that it can account for a substantial share of the cyclical output and employment losses that occurred in the 1973–75 recession. They calculate that an increase in the inflation target from 2 to 4 percent would be expected to lead in the short run to as much as a 2 or 3 percent decline in output while reducing welfare in the long run by a smaller amount (equal to the value of 0.3 percent of a year's output).

Unanticipated Inflation

The discussion to this point has been about the effects of an inflation that is fully anticipated. However, any change in the inflation target necessarily includes an element that is not anticipated. Many plans and contracts that were made in the expectation of a 2 percent inflation economy will turn out badly for some if the Fed were to adopt a 4 percent inflation target. This will be especially true for institutions like life insurance companies, pension funds, and foreign central banks where the portfolio share of long-term U.S. Treasury securities is large. Of course, the large stock of outstanding long-term debt that was issued at low interest rates creates an incentive for the government to adopt a higher inflation objective to inflate away the real burden of the debt on taxpayers. Raising the inflation target would reinforce fears that the government would use higher inflation to effectively default on its debt.

Initially, raising the inflation objective from 2 to 4 percent would arbitrarily redistribute wealth away from savers and toward debtors. But these are distributional effects—some people would be made better off at the expense of others. Higher inflation is associated with more uncertainty about inflation and that hurts everyone. Unanticipated inflation distorts relative price signals, creating confusion about whether any given price change is due to real factors

affecting supply and demand or whether the price change is just part of the absolute change in the average price level. People must tax their memories and powers of calculation to compare the value of goods at different periods of time. Difficulties with accurate price comparisons lead to economic inefficiencies that reduce economic output or, at least, the social value of output.

When inflation becomes less predictable, people shy away from the use of fixed long-term supply contracts, whether they are for labor or other factors. People devote resources to activities that protect them from a future of uncertain inflation. Financial advisors specialize in giving advice about how to protect investments from inflation risk. CEOs are more likely to be chosen from a pool of those with expertise in finance and accounting rather than from a pool of those whose expertise is in the production and distribution of a firm's output and in the particular market for its product. During the 1970s, we saw the creation of a CPI futures market to sell and price inflation risk. When inflation came down this market closed, and the resources that were used to support it were reallocated to more productive activity. During the 1970s, additional accountants and economists were hired to solve problems of measurement and forecasting. The expenses that are incurred to avoid the costs of inflation uncertainty are included in the GDP accounts and cause reported GDP to be higher, even though social welfare may be lower. It is analogous to the surge in aggregate demand that follows a severe hurricane or earthquake. Billions of dollars may be spent to rebuild homes and businesses, but in the end, the total level of capital (and social welfare) is no higher than it was before the natural disaster.

The Flawed Policy Framework

The rationale behind the calls for a higher inflation rate are based on an incorrect reading of why the policy interest rate went to zero in December 2008 and flawed economic ideas about how lower interest rates affect the economy.

Interest Rates Went to Zero Because the Fed Flooded the Market with Bank Reserves

The rate has stayed at zero because the Fed has continued to keep the market flooded with bank reserves. A higher inflation target

would not have made any difference. The interest rate did not hit zero because the Fed was trying to stabilize the business cycle. It went to zero in the fourth quarter of 2008, because the Fed dumped \$600 billion in excess reserves into the banking system as it rescued large banks. The \$600 billion was injected mainly by purchasing short-term securities or lending funds with a maturity under 180 days. If the Fed had allowed this paper to run off as it matured, the balance sheet would have shrunk back to normal and the policy interest rate would have been back closer to normal as early as the middle of 2009 when the recession ended. But instead, in a series of announcements beginning on November 25, 2008, the FOMC promised to purchase over \$1 trillion in agency debt and mortgage-backed securities (MBS) as well as long-term Treasury debt in order to keep the interest rate at zero well past the middle of 2009.⁶ Even if the inflation target had been 4 percent, the Fed would have kept interest rates at zero in this circumstance.

The Fed Has Become Overly Aggressive in Lowering Rates and Overly Cautious in Raising Them

The past two decades have seen a dramatic increase in the responsiveness of the Fed to economic weakness. Williams (2009) looks back at earlier recessions and suggests that the Fed could and should have done more because he observes that the rate did not go to zero in any of these earlier recessions. Greenspan (2004) presents an overview of a risk management approach to monetary policy. The idea is that policymakers anticipate events (probabilistically) that will lead to bad outcomes for inflation or output. Policymakers should also estimate the harm done by the possible bad outcomes. They weigh the probability of the event occurring by the estimated damage done if it occurs. A low probability event (such as a depression) can have such large costs that the Fed will act quickly to avoid such a catastrophe. The intended consequence of this policy is that the Fed will lower interest rates more quickly following bad news.

⁶Agency debt refers to the debt of government-sponsored enterprises (GSEs) that support homeownership by issuing debt to buy mortgages, securitizing mortgages and selling them to get funds to buy more mortgages, and holding some mortgages. The GSEs include Fannie Mae, Freddie Mac, Ginnie Mae, and the Federal Home Loan Banks. For a detailed description of these asset purchases by the Fed, see Fawley and Neely (2013: 60).

The unintended consequence is that policy now is more likely to be at the zero lower bound for any given inflation target.

Lower Interest Rates Do Not Cause More Real Growth

The Fed's policy briefings show that lower interest rates will raise inflation and economic growth and lower the unemployment rate. But these predictions are not founded on reliable economic theory. They ignore the effect that lower interest rates have on people's willingness to work, save, and invest for the future. When setting interest rates monetary policymakers typically ignore how low interest rates drive investors to seek higher returns in riskier investments. The financial crisis occurred partly because the Fed ignored how its low interest policy encouraged "innovation" in finance that subsequently led to excessive leverage—namely, the use of new financial derivatives to make ever more risky investments. Advocates of low interest rate policies who are calling for a 4 percent inflation target understand this, but they believe that monetary policy should focus on the task of stabilizing inflation and the real economy while systemic risk can be managed by regulators and government supervisors, despite a long history of such supervisory failure, especially in recent years.

Lower Interest Rates Do Not Cause Higher Inflation

In the Fed's policy framework, lower interest rates lead to higher demand for goods which leads to higher inflation. The correlation between inflation and output changes over time with the nature of shocks hitting the economy and the inflation regime maintained by the central bank. The myth is that this correlation can be treated as a stable economic structure and used by policymakers as a logical framework with which to stabilize the business cycle and drive the economy to full employment. A look at the path of interest rates and inflation over the past 30 years shows that a downward trend in the Fed's policy interest rate has been associated with an ever lower inflation trend.

The data are largely silent on whether a stable relationship between inflation and output exists. If you start by assuming the relationship is stable and truly structural, there is not enough information in the data to reject this assumption. If you start by assuming that the relationship is unstable and not structural, there is also not enough information in the data to reject this assumption. Policymakers and

policy advisors want a theory in which the Fed can stabilize output, so they start with the assumption that this relationship is stable and can be exploited to control the real economy. If this assumption is wrong, then the rationale for raising the long-run inflation target to 4 percent evaporates.

In the conventional policy view, monetary policy can raise employment and output in the short run, but not in the long run. Akerlof, Dickens, and Perry (1996, 2000) disagree. They think that higher inflation makes labor markets work better because society can avoid some of the conflict associated with cutting wages when a firm faces negative demand shocks. Even if this were true, using inflation to cut real wages is a bad idea. Negative shocks affect some firms more than others. These shocks are often indicators of fundamental forces affecting costs and consumer demand. When a product falls out of fashion due to changing tastes or the development of new products, the old product is going to gradually disappear. The better information firms and workers get from price signals, the more quickly they can adjust to the structural changes that are needed in a dynamic market economy. Recent examples of products that appear to be in decline include desktop computers, pocket cameras, wrist watches, bookstores, newspapers, and LP records. Firms in these industries either reinvent themselves to produce new products or go out of business. Anyone working in these industries should be aware that their livelihood may be at risk. Disrupting price signals also hurts society overall because when economic adjustments are delayed resources are wasted and the economy operates below its potential. Generally, the recent experience of developed economies does not support the idea that the trend in unemployment will be higher when the inflation trend is lower. Indeed, in the United States both inflation and the unemployment rate have been unexpectedly low during the past few years, lower than forecasts by the Fed or the private sector.

Conclusion

The recommendation to increase the Fed's inflation target from 2 to 4 percent should be ignored for at least three reasons. First, raising the inflation target can damage the Fed's credibility and the nominal anchor for our fiat money standard. Second, there is published evidence that a 2 percentage point increase in the inflation target would cause important welfare losses. Third, the idea that lower

interest rates lead to higher inflation and more real growth is not supported by sound theory or U.S. data following the Volcker monetary policy reforms in 1979. That idea was discarded during the high inflation and high unemployment of the 1970s, but reemerged after monetary policy became credible in the 1980s. Yet, its reemergence has led to an environment during the 2000s in which interest rates have been kept too low for too long, spawning continued speculative behavior in housing and derivative markets.

References

- Abel, A. B. (1997) "Comment" on M. Feldstein, "The Costs and Benefits of Going from Low Inflation to Price Stability." In C. D. Romer and D. H. Romer (eds.) *Reducing Inflation: Motivation and Strategy*, 156–66. Chicago: University of Chicago Press.
- Akerlof, G. A.; Dickens, W. T.; and Perry, G. L. (1996) "The Macroeconomics of Low Inflation." *Brookings Papers on Economic Activity* 1: 1–76.
- (2000) "Near-Rational Wage and Price Setting and the Optimal Rates of Inflation and Unemployment." *Brookings Papers on Economic Activity* 1: 1–44.
- Altig, D., and Carlstrom, C. (1991) "Inflation, Personal Taxes, and Real Output: A Dynamic Analysis." *Journal of Money, Credit and Banking* 23 (3): 547–71.
- Ball, L. (1999) "Efficient Rules for Monetary Policy." *International Finance* 2 (1): 63–83.
- (2014) "The Case for a Long-Run Inflation Target of Four Percent." International Monetary Fund, Working Paper No. 14–92 (June).
- Blanchard, O.; Dell'Ariccia, G.; and Mauro, P. (2010) "Rethinking Macroeconomic Policy." International Monetary Fund SPN 10/03 (February 12).
- Boskin Commission (1996) "Toward a More Accurate Measure of the Cost of Living." Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index (December 4). Available at www.ssa.gov/history/reports/boskinrpt.html.
- Bullard, J. B., and Russell, S. (2004) "How Costly Is Sustained Low Inflation for the U.S. Economy?" Federal Reserve Bank of St. Louis *Review* 86 (3): 35–67.

- Dowd, K. (1994) “The Costs of Inflation and Disinflation.” *Cato Journal* 14 (2): 305–31.
- Evans-Pritchard, A. (2009) “Ken Rogoff Says Fed Needs to Set Inflation Target of 6 Percent to Help Ease Crisis.” *The Telegraph* (February 20).
- Fawley, B. W., and Neely, C. J. (2013) “Four Stories of Quantitative Easing.” Federal Reserve Bank of St. Louis *Review* 95 (1): 51–88.
- Feldstein, M. (1976) “Inflation, Income Taxes, and the Rate of Interest: A Theoretical Analysis.” *American Economic Review* 66 (5): 809–20.
- _____ (1980) “Inflation, Tax Rules, and the Stock Market.” *Journal of Monetary Economics* 6: 309–31.
- _____ (1982) “Inflation, Capital Taxation, and Monetary Policy.” In R. E. Hall (ed.), *Inflation: Causes and Effects*, 153–68. Chicago: University of Chicago Press.
- _____ (1997) “The Costs and Benefits of Going from Low Inflation to Price Stability.” In C. D. Romer and D. H. Romer (eds.), *Reducing Inflation: Motivation and Strategy*, 123–56. Chicago: University of Chicago Press.
- Fischer, S. (1981) “Towards an Understanding of the Costs of Inflation: II.” *Carnegie-Rochester Conference Series on Public Policy* 15: 5–42.
- Fleming, S. (2015) “Inflation Goal May Be Too Low, Says Rosengren.” *Financial Times* (April 20).
- Federal Open Market Committee (FOMC) (1996) Transcripts from the July 2–3, 1996, FOMC meeting. Available at <https://fraser.stlouisfed.org/docs/historical/FOMC/meetingdocuments/FOMC19960703meeting.pdf>.
- Friedman, M. (1968) “The Role of Monetary Policy.” *American Economic Review* 58 (1): 1–17.
- Gavin, W. T.; M. R. Pakko; and F. E. Kydland (2007) “Monetary Policy, Taxes, and the Business Cycle.” *Journal of Monetary Economics* 54 (6): 1587–1611.
- Greenspan, A. (2004) “Risk and Uncertainty in Monetary Policy.” *American Economic Review* 94 (2): 33–40.
- Heller, H. R. (2015) “The Fed versus Price Stability.” Project Syndicate (March 19). Available at www.project-syndicate.org/commentary/fed-inflation-target-by-robert-heller-2015-03#rmLjEvStyKTCyckQ.99.

- Hibbs, D. A. Jr. (1982) "Public Concern about Inflation and Unemployment in the United States: Trends, Correlates, and Political Implications." In R. E. Hall (ed.), *Inflation: Causes and Effects*, 211–32. Chicago: University of Chicago Press 1982.
- Kessel, R. A., and Alchian, A. A. (1962) "Effects of Inflation." *Journal of Political Economy* 70 (6): 521–37.
- Marty, A. L., and Thornton, D. L. (1995) "Is There a Case for 'Moderate' Inflation?" Federal Reserve Bank of St. Louis *Review* 77 (4): 27–37.
- National Public Radio (2011) "Does the Economy Need a Little Inflation?" Interview with Kenneth Rogoff. Available at www.npr.org/2011/10/07/141006642/does-the-economy-need-a-little-inflation.
- Paul, R., and Lehrman, L. (1982) "The Case for Gold: A Minority Report of the Gold Commission." Washington: Cato Institute. (Originally published in the *Congressional Record*.)
- Rogoff, K. (2014) "The 4% Non-Solution." *Project Syndicate* (June 5). Available at www.project-syndicate.org/commentary/kenneth-rogoff-examines-two-ways-to-beat-the-zero-bound-on-nominal-interest-rates.
- Sargent, T. J., and Zeira, J. (2011) "Israel 1983: A Bout of Unpleasant Monetarist Arithmetic." *Review of Economic Dynamics* 14 (3): 419–31.
- Selgin, G. A. (1990) "Monetary Equilibrium and the Productivity Norm of Price-Level Policy." *Cato Journal* 10 (1): 265–87.
- Summers, L. (1991) "Price Stability: How Should Long-Term Monetary Policy Be Determined?" *Journal of Money, Credit, and Banking* 23 (3): 625–31.
- Williams, J. C. (2009) "Heeding Daedalus: Optimal Inflation and the Zero Lower Bound." *Brookings Papers on Economic Activity* 2: 1–37.

BEHAVIORAL ECONOMICS AND FED POLICYMAKING

Mark A. Calabria

For every bias identified for individuals, there is an accompanying bias in the public sphere.

—Cass Sunstein (2014:102)

Behavioral economics has continued to gain momentum in challenging the standard rational actor model in economics. With a few exceptions, the emphasis has been on the cognitive failure of individuals outside of government. Niclas Berggren (2013: 200) estimates that 95.5 percent of behavioral economics articles in the leading economics journals do not contain an analysis of the cognitive ability of policymakers. In this article, I offer a preliminary analysis of potential cognitive failures in the Federal Reserve’s conduct of monetary policy. Proposals to “debias” monetary policymaking are offered, along with a discussion of how the Fed’s existing institutional structure ameliorates or exasperates potential biases.

Behavioral Economics in Monetary Policy

Initial attempts to incorporate behavioral economics into monetary policy focused on how policymakers might better account for the behavioral biases of market participants, especially in the area of

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Mark Calabria is Director of Financial Regulation Studies at the Cato Institute.

price and wage setting (Rotemberg 2008: 34–36). Behavioral economics has been used to provide a new rationale for reliance on a Phillips curve framework, as well as to place greater weight on unemployment. As Janet Yellen (2007: 15) has stated, “Behavioral macroeconomic models also provide theoretical underpinnings for the view held by most policymakers that, in the short run, monetary policy can and should strive to stabilize the real economy.” More recent scholars (e.g., Orphanides 2015) have attempted to explain the Federal Reserve’s own actions via the lens of behavioral economics. Those efforts are extended in this article.

Are policymakers at the Fed likely to experience cognitive failures? DellaVigna (2009: 364) has argued that politicians “are experienced agents facing high-stake incentives and significant competition” suggesting that at least some elements of the political realm are less susceptible to cognitive failure than households. Setting aside whether DellaVigna’s characterization of politicians is accurate, the question of competition and experience has been demonstrated to reduce cognitive failure (List 2003, 2004). However, while there may be some modest choice among currencies, the Fed has few, if any, real equals. The level of competition facing the Fed does not even approach that likely necessary to serve as an effective feedback mechanism.

If competition is unlikely to exist in sufficient degree to eliminate cognitive failures at the Fed, will experience suffice? Members of the Federal Open Market Committee (FOMC) are often professional economists and usually quite advanced in their careers. They serve long terms (in principle, at least). On the other hand, economic crises are, thankfully, not a frequent occurrence. This means that policymakers rarely have the opportunity to learn from them. Federal Reserve Chair G. William Miller stated as much during an FOMC meeting in December 1978 (quoted in Abolafia 2012: 99):

Boston Fed President Frank Morris: Mr. Chairman, I don’t think we understand what is really going on in the economy.

Chairman Miller: I’ll go along with that.

President Morris: I think it’s because we haven’t had enough experience judging the reaction of both the consumer and the investor to an economy with a high rate of inflation.

Chairman Miller: That’s right, we haven’t had any experience.

While Morris and Miller found themselves in an environment of unusually high inflation, other periods have found policymakers wondering how to maneuver in an environment of unusually low inflation (Akerlof, Dickens, and Perry 1996: 50–52). I shall return to this point in more detail below, but both statements from policymakers and the broad economic trends suggest experience alone will be insufficient to eliminate any behavioral biases among policymakers.

Potential Biases in Monetary Policy

Cognitive science has identified a number of potential decision-making biases. Some are more robust than others. Below, I discuss the following biases: availability bias, representativeness bias, status quo bias, loss aversion, and overconfidence. I also offer examples of how each of these may be present in monetary policymaking. The discussion is in no way exhaustive, but rather focuses on the most likely biases, as well as those for which we have some evidence. There remains considerable room for additional research into these and other cognitive biases and their related impact on monetary policy decisions.

Availability Bias

When asked to consider the frequency of possible outcomes, individuals often quickly go to what they can remember about similar situations. Ask any New Yorker about terrorist attacks and 9/11 is unlikely to be far from mind. Ask an economist about inflation and you may well get a lecture on the Great Inflation of the 1970s. Ask about a stock market bust, and you might get the answer “1929.” Ask about when allowing a large hedge fund to fail resulted in a collective financial yawn and you might get a puzzled look (think “Amaranth Advisors LLC”). Where an answer comes easily (“1929”) there is a tendency to overestimate the frequency of such events, whereas when an answer requires some research or consideration (“Amaranth”) the frequencies tend to be more accurately estimated or even underestimated. This is the heart of availability bias (Tversky and Kahneman 1973: 163–65.).

Monetary authorities must often make quick decisions. Should institution X be saved before the markets open in Asia on Monday morning? Will the smallest sign of deflation (or inflation) spiral out-of-control? Will the large scale purchase of assets stimulate lending

and the economy? While there will be, of course, some analysis conducted on these questions, it would not be too ungenerous to describe some actions by the Fed as “spur of the moment.” Witness the ever-changing rationales given for denying assistance to Lehman Brothers.

Given the handful of almost mythical events in the history of monetary policy and financial crises, it is perhaps not surprising that these same events dominate discussions of policy. Federal Reserve Chair Ben Bernanke has occasionally been called the “perfect man” for the 2008 crisis, because of his long study of the Great Depression. But this assumes we were facing another Great Depression. As it’s impossible to know whether that was actually the case in 2008, it is not a stretch to conclude that, given the actions ultimately taken, the implied probabilities of another Great Depression were magnitudes higher than the actual probability. There may be no bigger availability heuristic in macroeconomics than the Great Depression. In all likelihood, especially given their perceived failure to act appropriately in the 1930s, the Fed is operating under a number of availability biases.

Representativeness Bias

Financial and monetary policy decisions can be vulnerable to generalizations. Once we conclude that entity X has characteristic Y, and that entity X is a member of set Z, we may assume that all members of Z have characteristic Y. A common example of this is the stereotype that people who wear glasses are intelligent.

One possible example from Fed behavior is the belief that if bank A is experiencing an outflow of deposits then all other banks must also be experiencing an outflow of deposits (and therefore the broad provision of liquidity is necessary). Yet this assumes a degree of homogeneity across banks that may not be warranted. Without some investigation (preferably with a large sample), one cannot know whether liquidity or solvency pressures at a single institution (or a small handful of them) represents a larger systemic problem. In some cases, such as the Savings and Loan Crisis, individual problem institutions were largely representative; in the recent crisis, on the other hand, there is good reason to believe that troubled institutions were the exception rather than the rule. If the Fed fails to accurately gauge representativeness, it may unintentionally reduce

the transfer of activity from poorly managed institutions to better-run institutions.

Representativeness bias can also impact the conduct of monetary policy. Is an increase in energy or home prices indicative of coming broader price increases? Does a slowdown in manufacturing employment in Cleveland portend a larger trend? These are the sorts of questions that face the Fed on a constant basis. Of course there are attempts at minimizing representativeness bias, such as focusing on consumer prices excluding housing and energy, as well as looking at other price indexes besides the consumer price index. Nevertheless, the finding that the voting patterns of Federal Reserve governors are influenced by which district they came from offers evidence consistent with representativeness bias (Meade and Sheets 2005: 676). Such potential bias can be offset by having a diversity of geographic backgrounds represented on the Board of Governors. However, geographic diversity alone may not be sufficient to offset representativeness bias if Fed governors and regional presidents travel in similar social circles. Complaints about a perceived closeness to banks are in part a concern about representativeness bias.

Status Quo Bias

Even when circumstances dramatically change, we sometimes maintain the same course of action, discounting or even ignoring new information. Kahneman and Tversky (1982) argue that individuals feel greater regret for bad outcomes that result from new actions being taken than they do for bad outcomes that result from inaction. Samuelson and Zeckhauser (1988: 35–36) argue that status quo bias is consistent with loss aversion, as well as sunk cost thinking. Relatedly, procrastination can lead policymakers to delay taking important actions. Orphanides (2015: 183–86) suggests that procrastination “may explain the Federal Reserve’s current “fear of liftoff”—the reluctance to normalize its policy stance despite considerable evidence it should do so.

Loss Aversion

One of the strongest, if not the strongest, finding in experimental behavioral economics is that individuals appear to weight potential losses much higher than potential gains. One result of loss aversion is

that when gains and losses are symmetric or nearly so, risk aversion may set in. Loss aversion is the general explanation for the “endowment effect”—that is, the hypothesis that individuals value an object more simply because they already own it.

Loss aversion can be found in multiple contexts in monetary policy. The “hard fought” battle against the Great Inflation, for instance, might cause a bias against policies that risk greater inflation. On the other hand, moving to tighter policy late in a recovery could be viewed as jeopardizing gains in employment or asset values. Policy goals, especially when given numeric expression, may serve as reference points from which gains and losses are measured. For instance, there is nothing that is particularly special about an inflation rate of 2 percent, yet policymakers’ often expressed goal of hitting a 2 percent inflation target frames deviations from that rate, either below or above, as a loss. The same could hold with an unemployment rate target, although deviations below a specific unemployment target are rarely framed as losses. Witness, for example, the abandonment of the so-called “Evans rule” once unemployment broke its target of 6.5 percent (see Boesler 2014). To some degree, debates over whether the Fed should “consult” a policy rule and report deviations from it rests on an assumption that loss aversion will bias the policymakers toward following the rule even when deviations would be optimal.

Overconfidence

Another common finding in behavioral studies is that individuals regularly offer estimates of their own ability, competence, or judgment that far exceed an objective assessment. One example of overconfidence is the bias of illusory superiority, sometimes called the “Lake Wobegon effect” (in which all children are above average).

Overconfidence is not limited to the layperson, but has also been repeatedly found among experts. Most relevant for monetary policy is the finding that even economists are subject to overconfidence.¹ Policymakers at the Fed may fall victim to overconfidence in managing the macroeconomy in terms of timing, magnitude, and

¹See, for example, Erik Angner’s (2006) study of economic policymaking in the transition from Soviet planning toward privatization.

even the qualitative impact of interventions. Overconfidence can result in Fed actions that are at times “too little” but at other times “too much.” Too little intervention can result when policymakers believe their actions will have larger effects than objective analysis would indicate. Overconfidence can, for instance, cause problems when relying on interest rates to gauge the stance of monetary policy: low rates might mean that policy is easy, but they could also signal a weak economy. In addition, policymakers may also overestimate their ability to unwind an aggressive policy stance.

The presence of overconfidence may not offer conclusive guidance on which particular policies to follow, but it should suggest a greater degree of modesty in forecasting the impacts of those policies, as well as our ability to reverse them.

Can Expertise Overcome Bias?

Despite the possibility of widespread cognitive failures, modern American society largely manages to function. Arguably, it does so relatively well. One avenue for overcoming bias is to rely on those with expertise and experience in a particular field. Although experts suffer from biases like the rest of us, they also have greater opportunities for learning and adjusting to their biases. Some evidence suggests that in the presence of appropriate incentives, such learning does occur (Camerer and Hogarth 1999: 12–13). What is the likelihood that expertise can overcome bias in the realm of Federal Reserve monetary policymaking?

As Daniel Kahneman (2011) has observed, experts suffer from all sorts of biases that result in bad decisions and outcomes. Building upon the work of Paul Meehl (1954), Kahneman argues that experts are inferior to simple algorithms (like a Taylor Rule) because experts “try to be clever, think outside the box, and consider complex combinations of features in making their predictions” (Kahneman 2011: 224). In the studies reviewed and sometimes conducted by Kahneman, experts are always looking for that one additional data point that suggests a different course of action. We see that today, with the Fed claiming its decisions will be “data dependent,” but not telling us what data they will be dependent upon, or how different data will be weighted. Kahneman also notes that experts are inconsistent, giving different answers to the same (or similar) question. This characteristic may

be especially damaging in relaying to market participants the direction of monetary policy. Kahneman (2011: 225) summarized his research with a “surprising” conclusion: “To maximize predictive accuracy, final decisions should be left to formulas, especially in low-validity environments.”

Kahneman and psychologist Gary Klein have investigated which conditions are conducive to relying on the discretion of experts and which are not. It is not surprising that the Fed often characterizes itself as a “firefighter.” Scholars have indeed found that seasoned firefighters have good intuition about things like when the floor of a burning building is about to collapse (Klein 2003). Kahneman’s research, however, finds that these expert skills are built up over time. Novice firefighters do not display the same skills as veterans. This could be one justification for the long terms (14 years) allowed for Fed governors. But most Fed governors do not serve anywhere near that long. The average number of years of experience for FOMC members is just over six years (Woolley and Gardner 2009: 10). As financial crises and turning points in the economy happen less often than that—close to every 13 years in the United States for crises, according to Reinhart and Rogoff (2011)—the fact is that few Fed governors will operate in more than one or two crises. Nor are they even likely to operate in more than one or two inflections in the macroeconomy.

Monetary policy is also inherently subject to unpredictability. As Milton Friedman (1961: 447) observed, monetary policy operates with “long and variable lags.” This is one reason why the Fed often ends up promising specific outcomes that subsequently fail to materialize. The very complexity and unpredictability of monetary policy suggests that the Fed would be more accountable if it were rule-bound.

To summarize these findings, experts can be relied upon when (1) they operate in a regular, predictable environment, and (2) there is an opportunity for learning via repeated practice. Neither of these conditions characterize monetary policy. Behavioral economics has sometimes been presented as an avenue to justify government intervention to correct the failing of ordinary people. Yet the same literature reminds us that policymakers, even experts, also suffer from a variety of biases. Just as default rules may be useful in minimizing consumer errors, they are also likely to be useful in minimizing monetary errors.

Mechanisms to Remove Bias

Although the quantitative impact of cognitive biases on monetary policy is difficult to measure with any precision, the historical record, including statements from FOMC members, suggests that cognitive biases may have a substantial influence on the conduct of monetary policy. Given this possibility, coupled with the importance of monetary policy to macroeconomic stability, I suggest below a number of avenues for reducing the impact of cognitive biases.

Rules-Bound Monetary Policy

The “rules versus discretion” debate in monetary economics has traditionally focused upon the time inconsistency of policymaking. The problem is that in the short run “surprise” inflations by a central bank may produce increases in employment and output. Over time, however, market participants come to anticipate this inflation with employment and output reverting to baseline. The result is higher inflation but no long-run improvement in either employment or output. Rules-based policy that limited the choices of policymakers could resolve this time inconsistency.

Given the stagflation of the 1970s, it is not surprising that era also witnessed a rebirth in the economic debates over rules versus discretion in monetary policy, most associated with the work of Calvo (1978), Kydland and Prescott (1979), Barro and Gordon (1983a, 1983b), McCallum (1984), and Taylor (1985). This body of work largely assumes that policymakers are rational and that errors are the result of misaligned incentives.²

Orphanides (2015: 184–88) has suggested that cognitive biases among policymakers provide an argument for monetary rules. Kahneman (2011), as discussed above, specifies the conditions under which rules are preferred to expert discretion. These conditions would appear to characterize monetary policymaking. Accordingly, requiring an explicit monetary rule, or that deviations from rules be explained, offers considerable potential to minimize the impact of cognitive biases among FOMC members. Further research is warranted on whether different rules are more or less susceptible to inducing their own biases and reducing preexisting biases.

²For an overview of this literature, see Walsh (2010: chap. 7), as well as Alesina et al. (2011). Hetzel (1985) and Tavlas (2014) provide an overview of the earlier debates.

Decisionmaking within Committees

The primary focus of behavioral economics has been on individual decisionmaking, yet monetary policy is often, although far from exclusively, conducted by committee. Blinder and Morgan (2005: 800–1) suggest that, in the context of monetary policy, group decisions are superior to individual decisions. Bainbridge (2002) provides an extensive overview of the behavioral arguments in favor of board decisionmaking relative to individual decisionmaking. For a skeptical view see Sunstein and Hastie (2008).

The difference in these findings may be explained by the structure and composition of the board in question. If a board is constituted of like-minded individuals with similar backgrounds and experiences, then groupthink and confirmation bias become significant risks (see Schulz-Hardt et al. 2000: 666–67). The modern dominance of central banks by economists, for instance, has likely narrowed the range of deliberation and may have contributed to the observed decline in time spent on deliberation (Woolley and Gardner 2009: 15–18). That FOMC members increasingly come from the same geographic areas could also be contributing to a reduction in deliberation and an increase in groupthink. Sunstein (2002: 5) has suggested that “social pressures are likely to lead groups of like-minded people to extreme positions.”

Following Sunstein (2002: 5), if we are to expect committees to be effective at reducing cognitive biases in monetary policy, there must be sufficient encouragement for dissent. The Federal Reserve, in contrast, is well known for its history of discouraging formal dissent. Former Minneapolis Federal Reserve Bank President Narayana Kocherlakota (2016) recently argued that “consensus creates a strong status quo bias that reduces the sensitivity of monetary policy to incoming data.” At a minimum, the FOMC should transition to a norm that encourages greater dissent.

Increased diversity—in terms of views represented on the FOMC—may require legislation. Specifically, the requirements of Section 10 of the Federal Reserve Act should be updated to increase geographic, as well as occupational, diversity (Calabria 2016).

Overall, the behavioral literature provides some support for concluding that committees can potentially reduce cognitive biases. However, the same literature offers sufficient reason to

question whether that finding is robust to committee structure and composition.

Adversarial Review of Monetary Policy

A number of scholars have suggested that adversarial review, either judicial or congressional, can reduce behavioral biases in agency decisionmaking (see Seidenfeld 2002; Babcock, Lowenstein, and Issacharoff 1997). If policymakers know *ex ante* that they will later have to defend their choices, increased deliberation may occur. Third party review could also help in the sense that biases may be more transparent to others than to the biased decisionmakers. If the third party is a nonexpert, policymakers may be forced to more thoroughly examine their own choices in order to explain them satisfactorily.

The perceived reduction in deliberation among monetary policymakers may be related to the increased dominance of the Federal Reserve by economists. Economists deliberating among themselves will generally take for granted a number of assumptions which noneconomists will not. If these assumptions are in error and critical to the deliberations, the lack of scrutiny could greatly undermine the quality of monetary policymaking.

Most agency decisions are and should be subject to judicial review. Monetary policy is notably one that is not. Nor, indeed, is it likely to be. Whether it *should* be is an interesting question, but one which falls outside the scope of this article. The lack of judicial review does suggest, however, that monetary policy is subject to fewer checks than other agency actions. There is congressional review, which has generally been seen as rather ineffective. A potential avenue for reform is the proposal to subject monetary policy decisions to review by the Government Accountability Office (GAO). Having to regularly explain their actions to the GAO may well reduce cognitive bias among members of the FOMC, and so increase the quality of deliberations. A GAO audit could also increase the quality of deliberations at the biannual congressional monetary oversight hearings.

Conclusion

Members of the FOMC are human. Along with that humanity come a number of cognitive biases and limitations that can affect the conduct of monetary policy. Whether such biases have indeed affected policy remains an open empirical question, but anecdotal

evidence, as well as actual comments from FOMC members, suggests that cognitive biases do impact FOMC decisions. I have attempted here a preliminary sketch of some of these potential biases, along with proposed institutional changes that could reduce their impact. The policy that offers the greatest potential for reducing biases is a move toward greater reliance on rules-based decision-making. It should also be recognized, based on the extensive work of Gerd Gigerenzer and others (e.g., Gigerenzer and Gaissmaier 2011), that heuristic decisionmaking can deliver better results than extensive deliberation. However, for the reasons presented above, its usefulness may be less applicable to monetary policy.

References

- Abolafia, M. Y. (2012) "Central Banking and the Triumph of Technical Rationality." In K. K. Cetina and A. Preda (eds.), *The Oxford Handbook of the Sociology of Finance*, 94–112. Oxford: Oxford University Press.
- Akerlof, G.; Dickens, W.; and Perry, G. (1996) "The Macroeconomics of Low Inflation." *Brookings Papers on Economic Activity* (1): 1–72.
- Alesina, A.; Stella, A.; Friedman, B.; and Woodford, M. (2011) "The Politics of Monetary Policy." *Handbook of Monetary Economics*, 1000–54. Amsterdam: North Holland.
- Angner, E. (2006) "Economists as Experts: Overconfidence in Theory and Practice." *Journal of Economic Methodology* 13 (1): 1–24.
- Babcock, L.; Loewenstein, G.; and Issacharoff, S. (1997) "Creating Convergence: Debiasing Biased Litigants." *Law and Social Inquiry* 22 (4): 913–25.
- Bainbridge, S. (2002) "Why a Board? Group Decisionmaking in Corporate Governance." *Vanderbilt Law Review* 55 (1): 1–55.
- Barro, R., and Gordon, D. (1983a) "Rules, Discretion, and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics* 12 (1): 101–21
- (1983b) "A Positive Theory of Monetary Policy in a Natural-Rate Model." *Journal of Political Economy* 91 (4): 589–610.
- Berggren, N. (2013) "Time for Behavioral Political Economy? An Analysis of Articles in Behavioral Economics." *Review of Austrian Economics* 25 (3): 199–221.

- Blinder, A., and Morgan, J. (2005) "Are Two Heads Better Than One? Monetary Policy by Committee." *Journal of Money, Credit, and Banking* 37 (5): 789–811.
- Boesler, M. (2014) "Fed Ditches Evans Rule." *Business Insider* (March 19). Available at www.businessinsider.com/march-fomc-statement-2014-3.
- Calabria, M. A. (2016) "Yes, the Federal Reserve Has a Diversity Problem." *Alt-M* (June 23).
- Calvo, G. (1978) "On the Time Consistency of Optimal Policy in a Monetary Economy." *Econometrica* 46 (6): 1411–28.
- Camerer, C., and Hogarth, R. (1999) "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19 (1): 7–42.
- Choi, S. J., and Pritchard, A. C. (2003) "Behavioral Economics and the SEC." Michigan Law and Economics Research Paper No. 03–002. Ann Arbor, Mich.: University of Michigan.
- DellaVigna, S. (2009) "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315–72.
- Friedman, M. (1961) "The Lag in Effect of Monetary Policy." *Journal of Political Economy* 69 (5): 447–66.
- Gigerenzer, G., and Gaissmaier, W. (2011) "Heuristic Decision Making." *Annual Review of Psychology* 62: 451–82.
- Hetzl, R. (1985) "The Rules versus Discretion Debate Over Monetary Policy in the 1920s." Federal Reserve Bank of Richmond *Economic Review* (November/December): 3–14.
- Kahneman, D. (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., and Tversky, A. (1972) "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3 (3): 430–54.
- _____ (1982) "The Psychology of Preference." *Scientific American* 246: 160–73.
- Klein, G. (2003) *Intuition at Work: Why Developing Your Gut Instincts Will Make You Better at What You Do*. New York: Random House.
- Kocherlakota, N. (2016) "5 to 4." Kocherlakota 009 Blog (January 29). Available at sites.google.com/site/kocherlakota009/home/policy/thoughts-on-policy/1-29-16.

- Kydland, F., and Prescott, E. (1977) "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85 (3): 473–90.
- List, J. A. (2003) "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118 (1): 41–71.
- (2004) "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace." *Econometrica* 72 (2): 615–26.
- McCallum, B. (1984) "Monetarist Rules in the Light of Recent Experience." *American Economic Review* 74 (2): 388–91.
- Meade, E., and Sheets, D. (2005) "Regional Influences on FOMC Voting Patterns." *Journal of Money, Credit, and Banking* 37 (4): 661–77.
- Meehl, P. E. (1954) *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. St. Paul, Minn.: University of Minnesota.
- Orphanides, A. (2015) "Fear of Liftoff: Uncertainty, Rules, and Discretion in Monetary Policy Normalization." Federal Reserve Bank of St. Louis *Review* 97 (3): 173–96.
- Reinhart, C., and Rogoff, K. (2011) *This Time Is Different: Eight Centuries of Financial Folly*. Princeton, N.J.: Princeton University Press.
- Rotemberg, J. J. (2008) "Behavioral Aspects of Price Setting and Their Policy Implications." NBER Working Paper No. 13754.
- Samuelson, W., and Zeckhauser, R. J. (1988) "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1 (1): 7–59.
- Schulz-Hardt, S.; Frey, D.; Lüthgens, C.; and Moscovici, S. (2000) "Biased Information Search in Group Decision Making." *Journal of Personality and Social Psychology* 78 (4): 655–69.
- Seidenfeld, M. (2002) "Cognitive Loafing, Social Conformity, and Judicial Review of Agency Rulemaking." *Cornell Law Review* 87: 486–548.
- Sunstein, C. R. (2002) "Conformity and Dissent." Chicago Public Law and Legal Theory Working Paper No. 34. Chicago: University of Chicago.
- (2014) *Why Nudge? The Politics of Libertarian Paternalism*. New Haven, Conn.: Yale University Press.
- Sunstein, C. R., and Hastie, R. (2008) "Four Failures of Deliberating Groups." Chicago John M. Olin Law and Economics Working Paper No. 401; Chicago Public Law Working Paper No. 215. Chicago: University of Chicago.

- Tavlas, G. (2014) "In Old Chicago: Simons, Friedman, and the Development of Monetary-Policy Rules." Becker Friedman Institute Working Paper No. 2014-02. Chicago: University of Chicago.
- Taylor, J. (1985) "What Would Nominal GNP Targeting Do to the Business Cycle?" *Carnegie-Rochester Conference Series on Public Policy* 22: 61-84.
- Tversky, A., and Kahneman, D. (1973) "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5 (2): 207-32.
- Walsh, C. E. (2010) *Monetary Theory and Policy*, 3rd ed. Cambridge, Mass.: MIT Press.
- Woolley, J., and Gardner, J. (2009) "Does Sunshine Reduce the Quality of Deliberation? The Case of the Federal Open Market Committee." APSA Toronto Meeting Paper. Washington: American Political Science Association.
- Yellen, J. (2007) "Implications of Behavioral Economics for Monetary Policy." Paper presented at the Federal Reserve Bank of Boston Conference on "Implications of Behavioral Economics for Economic Policy" (September 28). Available at www.frbsf.org/our-district/files/0928.pdf.

THE LIMITS OF KNOWLEDGE AND THE CLIMATE CHANGE DEBATE

*Brian J. L. Berry, Jayshree Bihari,
and Euel Elliott*

Those who have knowledge don't predict. Those who do predict don't have knowledge.

—Lao Tzu

Doubt is not a pleasant condition, but certainty is absurd.

—Voltaire

The question of whether climate change is produced by anthropogenic global warming (henceforth AGW) has triggered an increasingly contentious confrontation over the conduct of science, the question of what constitutes scientific certainty, and the connection between science and policymaking. In a world in which we seek to understand complex, multifaceted phenomena such as

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Brian J. L. Berry is the Lloyd Viel Berkner Regental Professor, and former Dean of the School of Economic, Political, and Policy Sciences, at the University of Texas at Dallas. Jayshree Bihari is an independent scholar with a PhD in public affairs from the University of Texas at Dallas. Euel Elliott is Professor of Public Policy and Political Economy and the Associate Dean for Undergraduate Education in the School of Economic, Political, and Policy Sciences at the University of Texas at Dallas.

climate (and to extract from this knowledge appropriate policy responses) the enduring epistemological question arises: What do we know? Logical inquiry might be expected to help resolve this knowledge problem (Hayek 1945) but is confounded by the assertion that the “science is settled,” by condemnation of those who disagree as “deniers,” and even by proposals that they be prosecuted as RICO offenders.¹ There is increasing talk on the left—and even among Democratic state attorneys general and the highest levels of the Obama administration—of criminalizing the very effort to rebut the climate change orthodoxy (Gillis and Schwartz 2015, Moran 2016).

What could have been a fruitful, albeit perhaps contentious debate over decisionmaking when addressing highly complex phenomena has degenerated into a prolonged contest. While recognizing the problems attending denial of climate change, our purpose here is to elucidate the limitations of the now-dominant view. We ground this view within a Kuhnian framework and suggest the limitations of that framework in understanding the uncertainties of climate change and policies that flow from it. Kuhn (1962) points to an often-repeated process whereby scientific paradigms become locked in and resist challenges to their validity because knowledge production is socially controlled and deeply

¹The term “denier” seems to encompass a wide range of possible positions on global warming and climate change. It clearly includes not just those who reject altogether the idea that the planet is warming, or that it is warming as a result of greenhouse gas emissions. “Denier” also is used promiscuously to describe those who express even mild doubt or who offer certain qualifications about what may be occurring. Most important, the usage of “denier” appears to conflate these positions with the question of how, assuming anthropogenic global warming exists, we should address the problem. On this point, the possible approaches range from those of Lord Stern (2013) and Martin Weitzman (2007, 2009), who believe that dramatic action needs to be taken immediately, to those of Nordhaus (2007, 2008), who has suggested a more incremental approach. It is worth noting that even among those calling for dramatic action, there are major distinctions to be made between those who believe that catastrophic climate change is virtually certain, and those believe it is unlikely but that the precautionary principle nevertheless calls for addressing the problem through dramatic policy interventions.

invested in the political currents of the day.² Power relationships and vested interests have frequently played a critical role in determining what acceptable science is or is not. In contemporary parlance there is historical lock-in and path dependence: once there is commitment to a particular body of knowledge that relates to a particular course of action, the costs of change increase over time and even if one wishes to move to a different path, it is difficult to do so. This is not to say that it is impossible for dissenters from the standard accepted approach to get their views expressed in the standard academic journals, but it is clearly more difficult. Moreover, consistent with the concept of path dependency (Greif and Laitin 2004, Arthur 1989), once a scientific paradigm becomes locked in, it becomes increasingly difficult to challenge the status quo in the accepted scientific outlets, at least until challenges to the orthodoxy of the day become so compelling they cannot be ignored.

To be sure, sometimes change does take place in a relatively smooth fashion, as when Lavoisier's description of oxygen led to the abandonment of Becher's phlogiston theory of combustion. At other times, where long-held doctrine is at stake, the conflict over new ideas becomes brutal: Galileo was tried by the Inquisition, found guilty, and spent the rest of his life under house arrest. In all cases, time is involved and supporting facts must be provided before a new paradigm gains acceptance. Both Wegener's 1915 theory of continental drift and Milankovitch's 1912 theory of the relationship of climate cycles to earth-sun geometry were dismissed for many decades until new evidence was provided—the Wilson-Morgan-Le Pinchon-McKenzie evidence for plate tectonics that was codified in 1965–67 and the Hays-Imbrie-Schackleton spectral analysis of ice core data that reinforced the idea of orbital forcing in 1976 (Hays, Imbrie, and Shackleton 1976).

²As noted by an anonymous reviewer, discussion of climate change takes place to a considerable extent within the so-called "blogosphere." While there is no formal peer review process in that setting, there is nonetheless an informal sorting of viable versus nonviable ideas and theories. Given the resistance, consistent with Kuhnian processes involving the social construction of scientific theories, to research challenging the established scientific position, those who differ from the scientific orthodoxy often find these nonstandard means of communicating ideas very valuable.

Emergence of the AGW Paradigm

AGW theory is an example of a contemporary Kuhnian lock-in. How and why did it emerge? According to Lindzen (1992) it has its origins in the observation that CO₂ levels are increasing in the atmosphere due to the burning of fossil fuels, and that CO₂ increases are not simply correlated with rising global temperatures, but are forcing agents. Thus in Mann's view (Mann, Bradley, and Hughes 1998) the post-1970 surge of global growth has created a "hockey stick" of increased emissions, higher CO₂ levels, and therefore temperatures. The mechanism is that of the "greenhouse effect." CO₂ is one of several greenhouse gases—methane is another—that inhibit the radiation of heat from the earth's surface back to space: hence AGW. The abrupt increase in temperature supposedly captured by Mann's hockey stick led invariably to the conclusion that there was greenhouse warming, that humans were the cause, and that dramatic intervention was required to prevent runaway global warming in the future.³

This now-standard narrative, consistent with the Kuhnian vision of "normal science" as a social construct, can be seen in the emerging scientific debate over just when the Anthropocene Epoch began and humans achieved the ability to alter the nature of the environment. Most believers in AGW make the case that the Anthropocene began with the Industrial Revolution in the late 18th century (Crutzen 2002), with gradual increases in greenhouse gas emissions culminating in the sharp temperature increase portrayed in Mann's "hockey stick." Certainly, that narrative fits the needs of this camp. But there

³It is worth noting that 60 years ago many climatologists suggested that Earth was entering a new ice age, the consequences of which would have been profound for food production and its impact on human settlements and migration (Hays, Imbrie, and Shackleton 1976). At the same time, neo-Malthusians such as Ehrlich (1968) proclaimed that Earth was on a disastrous trajectory because global population was rapidly outstripping food supplies. Such predictions warned of disastrous food shortages in advanced industrial societies, and mass starvation elsewhere. Ehrlich's doomsaying struck a responsive chord with the intellectual elites of the day and helped spawn massive and draconian efforts at population control. Of course, Ehrlich's predictions proved laughably false. The Green Revolution and its progeny, as well as the amazing advances made with genetically modified crops, have led to massive increases in food production. Today, with a population more than double what it was when Ehrlich first came on the scene, the food supply situation is much better than it was. The new environmentalist mantra is that the revolution in food production is threatened by global warming.

are other narratives that may be at least as compelling and could have resulted in an alternative lock-in to a different view of global warming. Some contend that the Anthropocene began thousands of years ago with the transition from hunter-gatherer to agricultural societies (Glikson 2013, Ruddiman 2013, Balter 2013).

Researchers like Loyola (2016) evaluate the anthropogenic nature of greenhouse gas warming within the context of even longer cycles of hundreds of thousands of years marked by periods of warming that could not have been triggered by human activity. More immediately, others suggest that a date around 1610, or alternatively, 1964, should be considered the beginning (Lewis and Maslin 2015). In the latter case, advocates point to the effects of atmospheric nuclear testing and the Nuclear Test Ban Treaty that was signed in 1963 (Waters et al. 2014; also see Lewis and Maslin 2015). As for the earlier date, Nevle and Bird (2008) point to the dramatic Columbian discovery of the Americas and the resulting integration of economies that led to the collapse of indigenous civilizations by the early 1600s. European expansion resulted in a drastic decline in population, producing a noticeable drop in greenhouse gas emissions that only recovered decades later (also see Lewis and Maslin 2015).

The point here is that choosing one date or another produces a lens through which we view scientific observations. If humans were shaping the environment and engaging in activities contributing to AGW, or reducing AGW as in the case of the Columbian intervention, then the way we interpret our influence today on the climate must surely be altered. An early date leads to human impacts on climate being seen as being part of the natural order (Lewis and Maslin 2015), while a much later date produces a narrative that sees those activities that result in climate change as a violation of a pristine nature, which must be rectified by any means necessary. Today, the “consensus” notion is that the Anthropocene is a late arrival, and this fundamentally frames the debate.

The Role of Environmental Advocacy

AGW theory developed in the 1980s in lockstep with growth in environmental advocacy. In Europe, militant new Green parties emerged, committed to forestalling what they believed to be an apocalyptic crisis. In the United States large public interest advocacy groups began to use “global warming” as a fundraising instrument

and to funnel money to the climate modeling community. Money talked. “By early 1989, the popular media in Europe and the United States were declaring that ‘all scientists’ agreed that warming was real and catastrophic in its potential” (Lindzen 1992: 92). Although there were skeptics, they were becoming marginalized—sidelined in Senate hearings, having papers rejected by scientific journals, being denigrated in popular publications, and excluded from the activities of the Intergovernmental Panel on Climate Change (IPCC).

At this stage a firm marriage emerged between environmentalism and left-liberal parties’ dreams of sustainable egalitarian societies based on suppression of economic growth in favor of smaller populations, and relying to the maximum extent possible on renewables. Global warming also became a major concern for national security agencies, which further increased funding for climate modelers and drew increasing proportions of academia into a position of AGW dependence for their research support. Growing numbers of scientists, with eyes focused on their research budgets and livelihoods, were willing to sign on to the proposition that “the science is settled”—even to sign letters calling for the prosecution of skeptics, Inquisition style.⁴ To name just one instance among many, when the ecologist James Steele offered a rebuttal to the claims of a butterfly ecologist, Camille Parmesan (Parmesan 1996), that blamed climate change for threatening the Edith’s checkerspot butterfly with extinction, he was immediately attacked as a “denier” (Ridley 2015a). An insidious kind of policy-based evidence-making is taking place: as Plimer (2015: 10–11) suggests, it is based on “pre-ordained conclusion, huge bodies of evidence are ignored and analytical procedures are treated as evidence.”

⁴The most notorious such communication was a letter from Jagadish Shukla (the lead signatory), a climatologist at George Mason University; Edward Maibach, a professor of communications, also at George Mason; and 18 others dated September 1, 2015 (Richardson 2016). In this letter the authors urge the Obama administration to consider prosecuting, under RICO (Racketeering Influenced and Corrupt Organizations Act), corporations and other organizations that have “knowingly deceived the American people about the risks of climate change. . . .” Later, in response to a lawsuit brought by the Competitive Enterprise Institute (CEI), the main authors sought to keep their email exchanges from those challenging the letter. That effort failed when the CEI obtained a ruling from a federal court under the Freedom of Information Act requiring George Mason University to release the e-mails (Shukla et al. 2015). Sadly, the difficulty one would have in finding significant mainstream news coverage of these events parallels the difficulty that scientific dissenters have in publishing findings in academic journals.

A Poisoned Debate

Money aside, it may be worth considering why the arguments over climate change have become as vitriolic as they are. On the one hand, if one acknowledges the reality that climate constitutes a system consisting of interacting subsystems that behave in nonlinear and unpredictable ways, one should at least be open to the possibility of uncertainty and unpredictability, and to the limits of knowledge. Indeed, this is the very point made by Judith Curry, a leading atmospheric scientist, who refers to the vast problem of forecasting highly nonlinear systems like the climate, referring to the “uncertainty monster” (Curry and Webster 2011). But constructivist thinking, and in particular the kind of constructivism associated with a particular species of progressivism, tends to deny uncertainty. While humans play a central role in the climate change narrative, that role is oddly static. If humans are embedded within the climate system, and human institutions and technological systems are a part of it, one might expect an appreciation for how human-centered creativity responds to climate change in ways that preclude, or at least place limits upon, our ability to predict the future. There is a kind of “God’s Eye View” perspective in which climate change advocates claim a kind of omniscience. A more appropriate perspective would be to assume a “Local Eye View.” Complex systems possess the characteristic that there is no particular vantage point from which one can know everything. No observer is omniscient. Local observers have knowledge within their particular epistemic neighborhoods, but no agent has access to the entire system (Borrill and Tesfatsion 2011: 239–40).

But there are other reasons for the poisoned debate. Classical liberals from Hayek (1945) to Berlin (1969) have issued cautionary warnings about what happens when the state, or some entity seeking the blessings of the state, assumes it has a monopoly on the truth. If “they” believe *x* and others believe *y*, and if only *x* can be correct, then *y* is of necessity wrong. If one also believes that ultimate truths governing the human condition are at stake, or in this case that the very survival of the human race is in the balance, then the failure to believe *x* is enough to consign nonbelievers to the outer darkness. If one denies climate change, or even the means of accomplishing certain goals, one is committing an act against humanity because one is wrong on an existential moral issue having to do with the survival of

the human race.⁵ These are the dangers of the constructivist mindset when carried to its ultimate, logical conclusion. From the Robespierres of the French Revolution to the Stalins, Maos, and Pol Pots of the 20th century, moral certainty breeds intolerance and intolerance breeds extremists.

The Need for Popperian Epistemology

Skepticism about AGW theory continues to exist, and at high levels, because the massive computer models that have been constructed to predict global warming have consistently failed by overpredicting temperature increases, and by failing to predict the current 20-year “pause” (Shanahan 1992; also see Pindyck 2013 and Curry and Webster 2011). Of course, critics of the “pause” argument say that the data were flawed and that once corrections are made, the pause in temperature increases goes away. Needless to say, these arguments serve only to heighten suspicions among doubters, and to question the scientific integrity of the process. What may be required is full deployment of a Popperian epistemology to properly test the central propositions of AGW. Philosopher Karl Popper—an advocate of skeptical stances that encourage refutation—is perhaps unique among philosophers of science in rejecting the idea that we should seek to confirm our theories, as is typical among AGW proponents, but rather that we should seek to refute them. *Conjectures and Refutations*, one of Popper’s key works, argues that the aim of science is to offer conjectures (hypotheses) that can be subjected to testing

⁵The recent dispute over the desirability of nuclear power provides a startling example of how activists are capable of turning on each other. When four leading scientists (James Hansen among them), all long associated with the AGW argument and the need for rapid decarbonization, suggested that fossil fuels must be replaced, at least in part, by nuclear energy, they were condemned by Naomi Oreskes, a professor of the history of science and a longstanding activist. Yet their position merely acknowledged the reality that renewable energy sources such as wind and solar will not, in anything like the foreseeable future, be able to fully substitute for fossil fuels. As such, the complete elimination of nuclear energy flies in the face of everything we know about the problem of switching from a fossil fuel–dominated system of energy production to carbon-free sources of production. Former Microsoft founder and CEO Bill Gates finds himself in this company as well. Gates, who has been arguing passionately for a major effort in technological innovation, recognizes that nuclear energy would play an important role in any future carbon-free energy infrastructure.

(Popper 1963). On the basis of these tests, new conjectures are proposed and new experiments conducted in an iterative process that allows investigators to get closer to the truth. Popper never believed that theories are final, even in phases (De Bruin 2006). They are always subject to refutation. Using a Popperian approach, Michaels (2015) and Michaels and Knappenberger (2015) provide a test of whether climate models are actually simulating reality. They conclude that the models suggest a degree of sensitivity of temperature changes to CO₂ that does not exist. As a result, the models overpredict temperature increases. These findings are reinforced by research from Kirby et al. (2016).

Popper's evolutionary epistemology captures, in our view, the essence of science, but the conduct of climate science today is a far cry from the scientific process that Popperian epistemology envisages. Popperian epistemology rejects inductivism, as evidenced by the famous "Black Swan" analogy. We may observe only white swans, but that does not mean there are only white swans. If we base scientific judgment on the assumption there are only white swans, all that is required to reject the theory is the observation of a single black swan. Such a black swan may be the current "pause" in global warming that has occurred even while the presumed cause of warming, CO₂, has continued to increase. To confirm this conjecture Popperian logic would require the formulation of testable hypotheses that either refute or confirm the original theory, with the results of the experimentation used to refine or replace the hypothesis, in an iterative process.

No such process is to be found in AGW climate science. Instead, there is a confirmation bias with an emphasis on statistical significance. With a paradigm asserted to be accepted science, research findings that contradict the status quo are ridiculed and the standing of the errant scientists is questioned. Ascendant AGW theory—married to left-liberal politics—has become the basis of national and global policy initiatives. People are told to be confident in the ability of policymakers not only to have a firm grounding in all the essential facts governing the climate today, but also to be able to understand and account for those processes decades in the future, taking into account an ever-evolving human-technological-climate interaction. The Black Swan pause, if it is a pause, may belie this belief: Type I errors may have been committed involving the assumption that there has been significant human-caused climatic

change that requires dramatic action, when in fact no such action is necessary. Type I errors are consistent with constructivist rationality that assumes that we can be highly confident in models used to predict climate change and to determine appropriate policy responses (see Heimann [1993] for an excellent discussion of the practical implications of Type I versus Type II errors). The alternative Popperian approach involves acquiring knowledge through posing a hypothesis, testing it, then reformulating your hypothesis in an iterative process (Popper 1963, De Bruin 2006).

Constructivists believe that global warming is occurring and that this warming is the product of human activity: if temperature data disagree with their models it is the data that are wrong and must be adjusted to conform. Smith (2003) and others advocating for an ecological approach to decisionmaking provide a much better framework for understanding the process of knowledge acquisition. This paper argues that, given the tentative nature of knowledge, Popperian epistemology offers an important counterpoint to constructivist approaches. We suggest that data-model differences are, in effect, Black Swans that require reformulation of one's hypothesis rather than, as seems so often the case, an effort to "massage" the data.

The Forcing Question

At the heart of the debate is the question of "forcing"—what causes what. If there is "warming," is it a consequence of increased levels of CO₂ produced by the burning of fossil fuels, or are changing CO₂ levels a consequence of the rise (and fall) of temperatures due to natural fluctuations in the climate system and varying solar inputs? The current climate science paradigm presupposes high temperature sensitivity to CO₂, and its modelers embed their assumptions about this sensitivity in large-scale climate models that predict runaway temperatures. True believers predict that the probability of catastrophic change is so great that there is little option but to act in dramatic fashion. Many environmental activists dream of a "de-developed" society in which fossil fuels and automobiles are banned and in which childbearing is strictly controlled to achieve Malthusian goals in a redefined humanity—Callenbach's green "Ecotopia" (1990). Even those who doubt that the effects of climate change will yield catastrophic results (Weitzman 2009), think that the consequences of ignoring a very low-probability, high-impact event that

actually occurs are unacceptably high. For them, use of the precautionary principle inevitably suggests moving toward a society so risk-averse that the very basis of scientific and technological advance is threatened.

There are, of course, a variety of possible responses if climate scientists' predictions turn out to be valid:

- Do nothing, because efforts to reduce greenhouse gas emissions are too little and too late.
- Take incremental action, applying a “wait and see” approach combined with efforts to reduce greenhouse gas emissions in a manner consistent with continued economic growth, a gradual phase-out of carbon-based fuels, and the deployment of new energy sources.
- Take dramatic action immediately in an effort to significantly reduce greenhouse gas emissions by 2050. Such an effort presumably would be designed to stabilize the global temperature increase by 2100 in the 2.5–3 C° range.
- Concentrate on adaptation rather than mitigation, for example by employing geoengineering technologies designed both to reduce the increase in temperatures and to keep CO₂ concentrations below the critical level of 450 ppm.
- Engage in some combination of mitigation and adaptation as defined above.

Beyond these are a multiplicity of somewhat radical proposals for technological fixes. They include the use of aerosols, sprayed into the stratosphere to create greater reflectivity, thereby deflecting solar radiation back into space; the placement of vast solar mirrors in space to deflect solar radiation; genetically engineering arboreal forms designed to absorb larger amounts of CO₂ than trees and plants do in their natural form; and algae or bacterial organisms that will have similar effects in reducing the amount of CO₂ in the oceans. Less dramatic are proposals to reduce CO₂ production via advanced electric battery technology, fed by improvement in solar energy capture, and the deployment of improved nuclear reactors—even the transition to nuclear fusion, although the problem with nuclear energy is one of political will and the opposition of the radical environmental groups who are the backbone of the AGW movement.

Of course, if skeptics (e.g., Loyola 2016) are correct, the complexity of climate change—in which humans play some role but are not

the sole agent—suggests that we should not have great faith in any of our existing models. As Michaels (2015) and Michaels and Knappenberger (2015) have shown, these models exaggerate the extent to which CO₂ forces increasing temperatures.

The Realities of Complex Systems

The problem is that each of the solutions proposed above ignores the complexity of the decision that would need to be made—one which demands comprehension of a complex adaptive system like the environment. Too often, that decision is reduced to a rhetorical question: Do you or do you not believe in climate change? If you do believe, your only realistic option is to radically reduce greenhouse gas emissions in the next few decades. Yet uncertainties exist at each step. The seeming unwillingness of the AGW community to articulate the existence of any uncertainties and to reduce any and all skeptics to a kind of caricature (the “deniers”) does not seem consistent with the traditional conduct of science and notions of scientific objectivity. Pindyck (2013), certainly no global warming denier, has gone further, cautioning that using climate change models to determine policy is not a worthwhile enterprise. The models are simply not sufficiently precise to allow for effective policymaking. Moreover, it is necessary to consider the policy interdependence of potentially catastrophic events (Martin and Pindyck 2015). As noted earlier, Loyola (2016) points out the severe limitations of the current models, and considers the anthropogenic nature of climate change as being embedded within much larger cycles of variability that are produced by natural processes. There is also the critical need to address other aspects of the human condition (see Sen 2014), particularly in the developing world.

An example of the ever-evolving knowledge of the mechanisms involved in warming is provided by Kirkby et al. (2016), who find that highly oxidized organic compounds play a role in cloud formation. This finding has important implications: namely, that climate models may have underestimated the amount of cloud cover in preindustrial times. If that is the case, it means that the amount of radiative forcing from anthropogenic activities has been overestimated too. In turn, that points to global warming estimates over the next century toward the low end of the 1.5–5 C° prediction range. While certainly

not the last word, this research serves to highlight the very tentative nature of the climate change models.

Even if the AGW thesis is entirely correct, its true believers typically ignore the considerable benefits increased levels of CO₂ and global warming may bring. Greenpeace founder Patrick Moore has taken his successors to task in a powerful statement of the benefits of rising CO₂. He says:

It is a proven fact that plants, including trees and all our food crops, are capable of growing much faster at higher levels of CO₂ than present in the atmosphere today. Even at today's concentration of 400 ppm plants are relatively starved for nutrition. The optimum level of CO₂ for plant growth is about 5 times higher, 2000 ppm, yet the alarmists warn it is already too high. They must be challenged every day by every person who knows the truth in this matter. CO₂ is the giver of life and we should celebrate CO₂ rather than denigrate it as is the fashion today [Moore 2015].

Moore, of course, is making the point that higher CO₂ levels create the conditions for plant growth, thus allowing for more food to be grown. Higher CO₂ levels should be a boon for those concerned with feeding global populations. He concludes with “a challenge to anyone to provide a compelling argument that counters my analysis of the historical record . . . much of society has been collectively misled into believing that global CO₂ and temperature are too high” (Moore 2015). It is worth emphasizing that Moore is someone who has been at the forefront of the environmental movement. Whether he is correct or not, his views should at least be evaluated rather than automatically dismissed.

A Science is Never “Settled”

The AGW hypothesis hangs on the question of whether human-caused increases in global CO₂ are producing rising temperatures that are inimical to human welfare. While the great majority of the scientific community claims that the “science is settled,” Siegel (2015), taking a Popperian stance, has reviewed the evidence and concluded the opposite. Critical thinking, he says, led to his skepticism and a series of realizations: policy always involves politics; political beliefs cloud the ability to process information; forecasts

are neutral constructs—tools, not truth; and consensus is not an argument for any scientific principle. There are, he says, key questions: What are the natural drivers of temperature and its variability? What do projected increases in temperature and greenhouse gases hold for the environment and people? Is decarbonization the solution? He concludes that Mann's hockey stick is wrong, produced by cherry picking data from tree rings and not supported by any other evidence; that government agencies have rigged data to support the AGW hypothesis; that solar forcing evidence provides better explanations for temperature change; and that rigged inputs and false assumptions about feedback have guaranteed the now well-documented climate model failures. Yet despite this, support for the hypothesis remains strong and aggressive because "Think tanks, NGOs, universities, the alternative power industry, consultants, government agencies, magazines, and others switched from scientific inquiry to rent seeking" (Siegel 2015: 41). Now, perhaps Moore and Siegel are wrong. Nevertheless, it seems foolish to simply dismiss their arguments, although this is precisely what the Kuhnian paradigm of "normal science" would expect. Is it asking too much for all parties to argue in a manner more consistent with a Popperian epistemological approach, seeking through conjecture and refutation to arrive at something closer to truth?

Mann's hockey stick model remains at the core of the AGW community's beliefs, even though it has been condemned as being unreproducible, a product of faulty methodology and likely scientific misconduct (Steyn 2015). Monckton (2015) looked to the current "pause" as evidence that Mann is wrong, finding that:

- Satellite data show no global warming at all for 224 months from February 1997 to September 2015—more than half the 441-month satellite record.
- There has been no warming even though one-third of all anthropogenic forcings since 1750 have occurred since the Pause began in February 1997.
- The entire satellite dataset from January 1979 to date shows global warming at an unalarming rate equivalent to just 1.2 C° per century.
- Since 1950, when a human influence on global temperature first became theoretically possible, the global warming trend has been equivalent to below 1.2 C° per century.

- The global warming trend since 1900 is equivalent to 0.75 C° per century. This is well within the range of natural variability.
- The warming trend since 1990, when the IPCC wrote its first report, is equivalent to 1 C° per century. The IPCC had predicted close to three times as much.
- The IPCC's prediction of 4.8 C° by 2100 is four times the observed real-world warming trend, since we might in theory have begun influencing temperatures in 1950.

While many will balk at the veracity of findings by nonscientists like Steyn and Monckton, and therefore reject their conclusions out of hand, these arguments are closely paralleled by Curry. That atmospheric scientist's "lukewarm" position basically asserts that global warming is real, that it is very likely anthropogenic (at least in some very substantial percentage), but that the threat from global warming on a time scale that could be of any relevance to us, say a century or longer, is unlikely to be substantial. If this is the case, then expending vast resources now on a problem that future technological advances may readily be able to address seems extremely short-sighted and may, indeed, prove counterproductive (Curry and Webster 2011; Curry, Webster, and Holland 2006).

Curry's position is similar to Ridley's (2015b). Ridley's thesis in *The Evolution of Everything* is that the top-down perspective of most climate modelers overemphasizes the role of CO₂ in climate change at the expense of other influences.

Ridley (2015b: 272) states:

This explains, these sceptics (such as Judith Curry of the Georgia Institute of Technology) think, the failure of the climate to warm nearly as fast over recent decades as predicted. It also explains the fact that Antarctic ice cores reveal a clear relationship between temperature and CO₂ as the earth goes into and out of ice ages that is the reverse of that predicted by the theory: CO₂ levels follow temperature up and down, rather than precede them. Effects cannot precede causes, and we now know almost for sure that ice ages are caused by changes in the earth's orbit, with CO₂ playing a minor, reinforcing role, if any at all. In short, there is a tendency to over-prioritize CO₂ as a cause of global temperature, rather than just another influence among many.

A Precautionary Stance?

As Deutsch (2011) points out, it is still not known just what degree of sensitivity the atmosphere has to any given concentration of CO₂. That relationship is critical, since its value determines how serious the climate change problem is. If the value is high—that is, the atmosphere is very sensitive—then the urgency is high. If it is extremely high there is little to be done. But this leads invariably to the question of the anthropogenic origins of climate change. Deutsch says that all sides seem to assume that if it turns out that a random fluctuation in the temperature is going to have disastrous consequences the best thing to do is just grin and bear it; or, if two-thirds of the increase is anthropogenic and one third natural, we are not supposed to do anything about the natural part. Most important, the point that trying to predict what our net effect on the environment will be for the next century and then subordinating all policy decisions to optimizing that prediction cannot work. As Deutsch (2011: 440) argues,

We cannot know how much to reduce emissions by, nor how much effect they will have, because we cannot know the future discoveries that will make some of our present efforts seem wise, some counter-productive and some irrelevant, nor how much our efforts are going to be assisted or impeded by sheer luck. Tactics to delay the onset of foreseeable problems may help. But they cannot replace, and must be subordinate to, increasing our ability to intervene after events turn out as we did not foresee. If that does not happen in regard to carbon-dioxide induced warming it will happen with something else.

He goes on to note:

There is a saying that an ounce of prevention equals a pound of cure. But that is only when one knows what to prevent. No precautions can avoid problems we do not yet foresee. To prepare for those, there is nothing we can do but increase our ability to put things right if they go wrong.

Picking up on Deutsch's "ounce of prevention" theme, Martin and Pindyck (2015) also reject the precautionary principle. Although they accept the reality of climate change, Martin and Pindyck also believe that, first, climate change models are not useful for making policy

due to the complex uncertainties contained in such models. Second, they believe that even if climate change represents a serious threat, one cannot consider it in isolation. Civilization, in their view, faces many potential threats of varying likelihood—whether global pandemics, nuclear war, or economic collapse. As a result, efforts to address one problem may limit our ability to address another. Depending upon the expected scale of the problem and the likelihood of such a problem actually arising, decisionmakers at a particular time may rationally decide not to commit resources to a specific problem.

The precautionary principle might, at first, seem quite reasonable. But where does one draw the line in terms of when the precautionary principle should be triggered: at the likelihood of catastrophe being 1 in 10? 1 in a 100? 1 in a 1000? As Ridley (2015b: 73) suggests:

Blaise Pascal argued that even if God is very unlikely to exist, you had better go to church just in case, because if he does exist the gain will be infinite, and if he does not the pain will have been finite. To me this is a dangerous doctrine, which justifies inflicting real pain in the here and now on disadvantaged people on the basis of forestalling a distant possibility of doom. This was exactly the argument used by eugenicists: the noble end justifies the cruel means.

Carried to its logical conclusion, the precautionary principle could be used to completely paralyze our civilization and stymie progress for future generations—depriving our descendants of the chance to live in a much wealthier society than our own.

Conclusion

As awareness of the uncertainties of global warming has trickled out, polling data suggests that the issue has fallen down the American public's list of concerns. This has led some commentators to predict "the end of doom," as Bailey (2015) puts it. In light of this, it seems odd to keep hearing that "the science is settled" and that there is little, if anything, more to be decided. The global warming community still asks us to believe that all of the complex causal mechanisms that drive climate change are fully known, or at least are known well enough that we, as a society, should be willing to commit ourselves to a particular, definitive and irreversible, course of action.

The problem is that we are confronted by ideologically polarized positions that prevent an honest debate in which each side acknowledges the good faith positions of the other. Too many researchers committed to the dominant climate science position are acting precisely in the manner that Kuhnian “normal science” dictates. The argument that humanity is rushing headlong toward a despoiled, resource-depleted world dominates the popular media and the scientific establishment, and reflects a commitment to the idea that climate change represents an existential or near-existential threat. But as Ellis (2013) says, “These claims demonstrate a profound misunderstanding of the ecology of human systems. The conditions that sustain humanity are not natural and never have been. Since prehistory, human populations have used technologies and engineered ecosystems to sustain populations well beyond the capabilities of unaltered natural ecosystems.”⁶

The fundamental mistake that alarmists make is to assume that the natural ecosystem is at some level a closed system, and that there are therefore only fixed, finite resources to be exploited. Yet the last several millennia, and especially the last two hundred years, have been shaped by our ability—through an increased understanding of the world around us—to exploit at deeper and deeper levels the natural environment. Earth is a closed system only in a very narrow, physical sense; it is humanity’s ability to exploit that ecology to an almost infinite extent that is important and relevant. In other words, the critical variables of creativity and innovation are absent from alarmists’ consideration.

In that sense, there is a fundamental philosophical pessimism at work here—perhaps an expression of the much broader division between cultural pessimists and optimists in society as a whole. Both Deutsch (2011) and Ridley (2015b) view much of the history of civilization as being the struggle between those who view change through the optimistic lens of the ability of humanity to advance, to solve the problem that confronts it and to create a better world, and those who believe that we are at the mercy of forces beyond our control and that efforts to shape our destiny through science and technology are doomed to failure. Much of human history was under the control of the pessimists; it has only been in the last three hundred years that

⁶Even the Food and Agricultural Organization of the United Nations, itself an organization deeply prone to accepting the latest scientific orthodoxies, has acknowledged that the nine billion people expected to inhabit the earth by 2050 can be sustained indefinitely provided that necessary investments are made.

civilization has had an opportunity to reap the benefits of a rationally optimistic world view (see Ridley 2010).

Yet the current “debate” over climate change—which is really, in Ridley’s (2015a) terms, a “war” absent any real debate—has potentially done grave harm to this scientific enterprise. As Ridley documents, one researcher after another who has in any way challenged the climate orthodoxy has met with withering criticism of the sort that can end careers. We must now somehow return to actual scientific debate, rooted in Popperian epistemology, and in so doing try to reestablish a reasonably nonpolitical ideal for scientific investigation and discovery. Otherwise, the poisoned debate over climate change runs the risk of contaminating the entire scientific endeavor.

References

- Arthur, W. B. (1989) “Competing Technologies, Increasing Returns, and Lock-in by Historical Events.” *Economic Journal* 99 (394): 116–31.
- Bailey, R. (2015) *The End of Doom: Environmental Renewal in the Twenty-First Century*. New York: Thomas Dunne.
- Balter, M. (2013) “Archaeologists Say the Anthropocene Is Here—But It Began Long Ago.” *Science* 340 (6130): 261–62.
- Berlin, I. (1969) *Four Essays on Liberty*. Oxford: Oxford University Press.
- Borrill, P. L., and Tesfatsion, L. (2011) *Agent-based Modeling: The Right Mathematics for the Social Sciences?* In J. B. Davis and D. W. Hands (eds.), *The Elgar Companion to Recent Economic Methodology*, 228–58. Northampton, Mass.: Edward Elgar.
- Callenbach, E. (1990) *Ecotopia*. New York: Bantam Books.
- Crutzen, P. J. (2002) “Geology of Mankind.” *Nature* 415 (6867): 23.
- Curry, J. A., and Webster, P. J. (2011) “Climate Science and the Uncertainty Monster.” *Bulletin of the American Meteorological Society* 92 (12): 1667–82.
- Curry, J. A.; Webster, P. J.; and Holland, G. J. (2006) “Mixing Politics and Science in Testing the Hypothesis that Greenhouse Warming Is Causing a Global Increase in Hurricane Intensity.” *Bulletin of the American Meteorological Society* 87 (8): 1025–37.
- De Bruin, B. (2006) “Popper’s Conception of the Rationality Principle in the Social Sciences.” In I. Jarvie, K. Milford, and D. Miller (eds.), *Karl Popper: A Centenary Assessment: Selected*

- Papers from Karl Popper*, Vol. 3, 207–15. Aldershot, England: Ashgate Publishing.
- Deutsch, D. (2011) *The Beginning of Infinity*. New York: Viking Press.
- Ehrlich, P. R. (1968) *The Population Bomb*. New York: Ballantine Books.
- Ellis, E. C. (2013) “Overpopulation Is Not the Problem.” *New York Times* (September 13).
- Gillis, J., and Schwartz, J. (2015) “Exxon Mobil Accused of Misleading Public on Climate Change Risks.” *New York Times* (October 30).
- Glikson, A. (2013) “Fire and Human Evolution: The Deep-Time Blueprints of the Anthropocene.” *Anthropocene* 3 (November): 89–92.
- Greif, A., and Laitin, D. D. (2004) “A Theory of Endogenous Institutional Change.” *American Political Science Review* 98 (4): 633–52.
- Hayek, F. (1945) “The Use of Knowledge in Society.” *American Economic Review* 35 (4): 519–30.
- Hays, J. D.; Imbrie, J.; and Shackleton, N. J. (1976) “Variation in the Earth’s Orbit: Pacemaker of the Ice Ages.” *Science* 194 (4270): 1121–32.
- Heimann, C. (1993) “Understanding the Challenger Disaster: Organizational Structure and the Design of Reliable Systems.” *American Political Science Review* 87 (2): 421–35.
- Kirkby, J.; Duplissy, J.; Sengupta, K.; et al. (2016) “Ion-Induced Nucleation of Pure Biogenic Particles.” *Nature* 533 (7604): 521–26.
- Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lewis, S. L., and Maslin, M. A. (2015) “Defining the Anthropocene.” *Nature* 519 (7542): 171–80.
- Lindzen, R. S. (1992) “Global Warming: The Origin and Nature of the Alleged Scientific Consensus.” *Regulation* 15 (Spring): 87–98.
- Loyola, M. (2016) “Twilight of the Climate Change Movement.” *The American Interest* (March 31). Available at www.the-american-interest.com/2016/03/31/twilight-of-the-climate-change-movement.
- Mann, M. E.; Bradley, R. S.; and Hughes, M. K. (1998) “Global-Scale Temperature Patterns and Climate Forcing over the Past Six Centuries.” *Nature* 392 (6678): 779–87.
- Martin, I. W. R., and Pindyck, R. S. (2015) “Averting Catastrophes: The Strange Economics of Scylla and Charybdis.” *American Economic Review* 105 (10): 2947–85.

- Michaels, P. J. (2015) "Why Climate Models Are Failing." In A. Moran (ed.), *Climate Change: The Facts*. Melbourne: Institute of Public Affairs.
- Michaels, P. J., and Knappenberger, P. C. (2015) *Lukewarming: The New Climate Science that Changes Everything*. Washington: Cato Institute.
- Monckton, C. (2015) "There's Life in the Old Pause Yet." *WUWT* (October 8). Available at wattsupwiththat.com/2015/10/08/theres-life-in-the-old-pause-yet.
- Moore, P. (2015) "Should We Celebrate Carbon Dioxide?" Annual Global Warming Policy Foundation Lecture (October 15). Available at www.thegwpf.org/patrick-moore-should-we-celebrate-carbon-dioxide.
- Moran, R. (2016) "Democratic Attorneys General to Initiate Witch Hunt of Climate Change 'Deniers.'" *American Thinker* (March 30). Available at www.americanthinker.com/blog/2016/03/democratic_attorneys_general_to_initiate_witch_hunt_of_climate_change_deniers.html.
- Nevle, R. J., and Bird, D. K. (2008) "Effects of Syn-pandemic Fire Reduction and Reforestation in the Tropical Americas on Atmospheric CO₂ During European Conquest." *Paleogeography, Paleoclimatology, Paleoecology* 264 (1–2): 25–38.
- Nordhaus, W. D. (2007) "A Review of the Stern Review on the Economics of Climate Change." *Journal of Economic Literature* 45 (3): 686–702.
- (2008) *A Question of Balance: Weighing the Options on Global Warming Policies*. New Haven, Conn.: Yale University Press.
- Parmesan, C. (1996) "Climate and Species Range." *Nature* 382 (6594): 765–66.
- Pindyck, R. S. (2013) "Climate Change Policy: What Do the Models Tell Us?" *Journal of Economic Literature* 51 (3): 860–72.
- Plimer, I. (2015) "The Science and Politics of Climate Change." In A. Moran (ed.), *Climate Change: The Facts*. Melbourne: Institute of Public Affairs.
- Popper, K. (1963) *Conjectures and Refutations*. London: Routledge.
- Richardson, V. (2016) "Professors Pushing RICO for Climate Skeptics Have Own Problems with FOIA." Available at www.washingtontimes.com/news/2016/may/15/climate-skeptic-rico-pushers-tried-to-hide-records.
- Ridley, M. (2010) *The Rational Optimist: How Prosperity Evolves*. New York: Harper.

- _____ (2015a) “The Climate Wars’ Damage to Science.” *Quadrant* (June 19). Available at quadrant.org.au/magazine/2015/06/climate-wars-done-science.
- _____ (2015b) *The Evolution of Everything: How New Ideas Emerge*. New York: Harper.
- Ruddiman, W. F. (2013) “The Anthropocene.” *Annual Review of Earth and Planetary Sciences* 41 (4): 1–24.
- Sen, A. K. (2014) “Global Warming Is Just One of Many Environmental Threats that Demand Our Attention.” *The New Republic* (August 22).
- Shanahan, J. (1992) *A Guide to the Global Warming Theory*. Washington: Heritage Foundation.
- Shukla, J.; Maibach, E.; Dirmeyer, P.; et al. (2015) “Letter to President Obama, Attorney General Lynch, and OSTP Director Holdren.” Available at templatelab.com/LetterPresidentAG (September 1).
- Siegel, D. (2015) “What I Learned about Climate Change: The Science Is Not Settled.” Available at medium.com/@pullnews/what-i-learned-about-climate-change-the-science-is-not-settled-1e3ae4712ace#.7xtklqmie (October 16).
- Smith, V. L. (2003) “Constructivist and Ecological Rationality in Economics.” *American Economic Review* 93 (3): 465–508.
- Specter, M. (2015) “How Not to Debate Nuclear Energy and Climate Change.” *New Yorker* (December 18).
- Stern, N. (2013) “The Structure of Economic Modeling of the Potential Impacts of Climate Change: Grafting Gross Underestimation of Risk onto Already Narrow Science Models.” *Journal of Economic Literature* 51 (3): 838–59.
- Steyn, M. (ed.) (2015) *A Disgrace to the Profession*. Woodsville, N.H.: Stockade Books.
- Waters, C. N.; Zalasiewicz, J. A.; Williams, M.; et al. (2014) “A Stratigraphical Basis for the Anthropocene?” *Geological Society, London, Special Publications* 395 (1): 1–21.
- Weitzman, M. L. (2007) “A Review of the Stern Review on the Economics of Climate Change.” *Journal of Economic Literature* 45 (3): 703–24.
- _____ (2009) “Tail-Hedge Discounting and the Social Cost of Carbon.” *Review of Economics and Statistics* 51 (3): 873–82.

THE IMPACT OF CORPORATE TAX POLICY ON STATE TRADE FLOWS

Austin P. Johnson

State politicians are well aware of the importance of having a robust economy in order to be reelected. The economic vote is well documented in the literature and the intuition is naturally understood among politicians (Lewis-Beck and Stegmaier 2013). “It’s the economy, stupid,” which harkens back to the 1992 U.S. presidential election, has become a common slogan that is reiterated on media outlets in almost every election cycle. Therefore, economic growth is of great importance to elected officials, especially executives, and most states have entire departments dedicated toward advancing economic development within their boundaries (Eisinger 1988). Business tax cuts are a direct approach to economic development, although current empirical research on this topic is mixed at the state level.

A variety of mechanisms for spurring state economic growth through tax cuts have been suggested in the literature. First, it has been argued that cutting taxes can attract business from other regions, enriching the low-tax state. A potential problem is that subsequent reciprocal tax cuts in other states to attract businesses could lead to a race-to-the-bottom (Basinger and Hallerberg 2004; Grieco 1988; Plümper, Troeger, and Winner 2009). However, there is virtually no evidence for this type of excessive fiscal competition at the

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Austin P. Johnson is a doctoral student in the Department of Political Science at Texas A&M University. He thanks his former colleagues at Deloitte LLP for their past mentorship, and an anonymous reviewer for providing useful commentary on an earlier version of the article.

state level during the recent financial crisis. Second, lower taxes are thought to increase the profit retention of existing resident businesses, which in turn spurs further growth-enhancing investment (Eisinger 1988). For example, retained revenues paid out in dividends can then be reinvested into debt or equity markets, aiding other businesses. Alternatively, the retained earnings can be reinvested in the business's own operations, leading to business expansion that includes increased production and a larger payroll. A third possible economic growth mechanism is that tax cuts lead to spending cuts, which could free up resources for the private sector (Browning and Johnson 1984, Friedman 1967). These mechanisms all describe ways in which tax cuts can benefit the economy, and are therefore attractive as policy instruments.

Tax cuts are also desirable from the perspective of the businesses they affect. Businesses have become increasingly capable of voting with their feet—also known as capital mobility—and moving to a state with a more favorable business climate. Capital mobility is a well-documented issue in the political economy literature, and taxation can be a potential instigator of capital mobility (Andrews 1994, Cerny 1990, Jensen 2013, and Kurzer 1993). Capital mobility frequently has a negative connotation, but the movement of capital can prove to be the salvation of businesses teetering on the edge of bankruptcy, saving numerous jobs in the process. Incentives have been found to be effective at attracting mobile capital, such as manufacturing (Head, Ries, and Swenson 1999), but there has been a lack of adequate data to empirically test how incentives alter the location of firms with much regularity (Arauzo-Carod, Liviano-Solis, and Manjón-Antolín 2010). During my review of the literature, I have found no research that has rigorously considered capital mobility in response to tax incentives at the state level. Many corporations today are capable of acquiring information and talented consultants for the purpose of restructuring business operations. Assuming they comply with their fiduciary duty to shareholders, corporations are likely to restructure their operations in such a fashion that minimizes their effective tax rate (ETR). Relatively higher business taxes may therefore trigger businesses to leave a relatively higher-tax state for a relatively lower-tax state. Finding evidence for this behavior has proven elusive at the aggregate level until now.

In the remainder of this article, I discuss the current literature on taxation and state economic growth, argue for using trade flow data

as a measure of economic activity in the United States, test for the influence of taxation on economic activity, and conclude with a discussion of the pertinent empirical findings. My findings indicate that pro-business state tax policy matters for growing existing businesses and attracting new ones.

Taxation and Economic Growth

The literature that examines how taxation influences state economies has focused on how taxation directly influences gross state product (GSP). Over the last 10 years, six articles have examined the influence of taxation on U.S. state economic growth at the aggregate level. This body of research has produced mixed findings, and it is widely acknowledged that empirical model specification and methodology are largely responsible for this variation (Reed 2009).

Specific details on the different methodological approaches for all six articles are listed in Table 1. As can be seen from the table, one of the major factors that leads to differing results relates to how tax revenues are aggregated across studies. Studies that aggregate all taxes together when calculating the total tax burden are consistent. Bania, Grey, and Stone (2007) and Reed (2008) both found that an increase in the *total tax burden*, a measure that aggregates all state-level taxes together, can harm growth. Similarly, Atems (2015) found evidence that an increase in the total tax burden of a state can harm both internal growth and the economic growth of nearby states. The consistency of the results changes after scholars begin to dissect tax revenues into their constituent components. Ojede and Yamarik (2012) segmented taxes into income, sales, and property taxes, and found that increases in the latter two categories hurts growth (income taxes were statistically insignificant). In contrast, Prillaman and Meier (2014) split tax revenues into business and nonbusiness categories and created two sets of models. One set of models was “trimmed” for outliers (Prillaman and Meier 2014: 371), and the authors found evidence that nonbusiness taxation severely hinders economic growth but business taxation supposedly increases economic growth.¹

¹Prillaman and Meier (2014) excluded “energy-producing states” for certain years and controlled for energy prices at the consumer level, but they did not attempt to control for upstream energy prices that benefit energy-producing states. This process may exclude the very states that are the most capable of attracting businesses.

TABLE 1
RECENT STUDIES ON TAXATION AND STATE-LEVEL ECONOMIC GROWTH

Author(s)	Bania et al.	Reed	Reed	Ojede & Yamarik	Prillaman & Meier	Atems
Year Published	2007	2008	2009	2012	2014	2015
Estimator(s)	GMM	Multiple	EBA	Panel-ECM	GMM	Spatial Durbin
Taxation	Total	Total	Property	Total	Business	Total
Variable(s)			Sales	Individual Income	Nonbusiness	
			Individual Income	Corporate Income		
			Corporate Income	Total Income		
			Corporate Income	Sales		
Finding(s) of Taxation on Growth	Negative	Negative	Mixed	Property	Null	Negative
				Mixed		

NOTES: Articles may have more than one model; GMM is general method of moments (Arellano-Bond estimator); EBA is extreme bounds analysis; OLS is ordinary least squares; FGLS is feasible generalized least squares; and ECM is error correction model. Reed (2008) found consistent results with multiple estimators: a GMM model, OLS with 2-way fixed effects model, and weighted FGLS model.

With respect to their *untrimmed* set of models, Prillaman and Meier (2014) produced largely null findings for both measures of taxation, which contradicts past literature. Reed (2009) considered a wide variety of taxation measures, and most notably found that an increase in the overall tax burden hurts economic growth, but that corporate income taxes *increase* growth in a segmented model. The author cautioned that these results could have been influenced by the degree to which corporate profits can be “exported outside the state” (Reed 2009: 697), an allusion to capital mobility.²

To put it simply, capital mobility with large corporations is quite real in North America. Empirically, Prillaman and Meier (2014) found some evidence for business exits in response to higher business taxes in their empirical models, and this finding should not be surprising. Businesses are started for the purpose of generating profits, and will seek to do so as effectively as possible. Publically traded corporations frequently pay management in long-term stock options, which has been found to align management’s self-interest with that of shareholders during aggressive tax planning projects (Armstrong, Blouin, and Larcker 2012; Slemrod 2004). Tax planning is the strategic process of minimizing a corporation’s ETR. A rapid expansion in tax planning and compliance projects can be seen in the financial statements released by one of the largest accounting firms that make this data public by region; Deloitte LLP reported that consulting and tax service revenues (not including SEC required auditing services) increased by over 19.8 percent in the United States from 2013 to 2015 (Deloitte 2015). The rise of large tax consulting firms, whose services increase the accessibility of corporate tax law and enable corporations to engage in tax planning, has given many corporations the opportunity to gain a competitive advantage by restructuring their operations to avoid regions that may lead to higher ETRs (Bonner, Davis, and Jackson 1992; Hasseldine, Holland, and van der Rijt 2011; McGuire, Omer, and Wang 2012). As an example, Google Inc. reduced its ETR from approximately 35 percent to a meager 2.4 percent through aggressive tax planning with its *global* operations (Drucker 2010). In the rapidly evolving atmosphere of interstate

²It is important to note that most states employ formulary apportionment systems to tax large corporations, not transfer pricing, so profits cannot merely be shifted on paper between affiliates.

business, reorganizing supply lines and business structures according to the results of tax planning has become a viable method for businesses to respond to state corporate taxation. Previous empirical literature has largely been unable to capture this widely understood phenomenon that is crucial for understanding how states can attract business.

The principle problem with the extant literature is model specification. The literature typically uses different empirical methodologies, dependent variables, years of data, and measures of taxation. The end result is that taxation is often found to influence state economies, but there is no conclusive evidence as to *how* it might matter, for better or worse. Instead of focusing on economies as independent silos, I propose that the study of *intranational* trade flows represents a better method for evaluating the health of state economies. The macroeconomy is essentially an amalgam of microeconomic transactions. While my research design still uses aggregate data, it emphasizes aggregate transactions through trade flows that occur within states and between states, rather than attempting to directly estimate GSP with exotic methodologies and overspecified models. In short, I propose that taxation is best examined spatially since state economies spill over their borders. Current literature unequivocally finds that increased trade between economies is associated with gains in real income (Noguer and Siscart 2005), and heightened internal trade clearly reflects an increase in internal economic activity. Therefore, increased trade flows are an indicator of growth due to increasing economic activity.

Research Design and Data

In this article, I utilize a set of spatial gravity models to test my theory. Gravity models are loosely based on Newton's theory of gravity. The central idea is that trade flows are a positive function of market size (i.e., market mass) and a negative function of transaction costs, namely distance (Leblang 2010).³ My regression models reflect an

³Leblang (2010) calculates market mass by taking the logged product of GDP for the two regions involved in a transaction. To mitigate concerns of reverse causality and market volatility in the regression model, I emulate this approach but utilize lagged two-year averages of GSP (e.g., 2010–11 for 2012). Data on GSP were retrieved from the Bureau of Economic Analysis.

augmented specification of the gravity models utilized in the interstate home-bias literature. This literature seeks to explain why state borders in the United States influence the level of trade in America (Hillberry and Hummels 2003, Wolf 2000). To correct for the non-linear relationship of these variables and so that ordinary least squares (OLS) regression can be utilized, the volume of trade, market mass, and distance are logged. Logging these variables is a universal practice in the literature that uses gravity models.

Dependent Variable

The trade flow data come from the economic census carried out every five years by the Census Bureau and the Bureau of Transportation Statistics. This census of U.S. economic activity involves the mass surveying of established businesses across all sectors of the economy regarding their sales transactions and associated data, such as weight and the distance traveled for goods sold. These transactions may be passing from businesses to independent parties, from parent companies to a subsidiary, from subsidiaries to parent companies, or from subsidiaries to other subsidiaries. These data are later compiled into a large dataset on trade flows in the United States, known as the commodity flow survey (CFS). The two most recent CFS datasets are used in this article since they are consistent in their measurement techniques.

The trade flow data are directional and include all 50 states that constitute America, for 2007 and 2012. Given the recent evolution of state tax policy with respect to multistate corporations, the timeframe of this dataset is ideal. The commodities from each state in the dataset have been aggregated by dollar value. These trade flows can move within the state or to one of the other 49 states in the union, for a total of 50 possible destinations for each state. This leads to a potential sample size of up to 2,500 observations per year, for a total possible sample size of 5,000 in this dataset. Due to the system of surveying and associated responses to the CFS, there are some minor instances of missing data between small states. As a result, I have 4,424 total observations for my regression models.⁴

⁴An analysis of the missing values reveals that they are due to the withholding of cash values of goods by surveyed businesses. The distribution of these missing values appears random and should not bias the models.

Control Variables

Two control variables are included in the models to adjust for the political and geographic context of the trade flow transactions, and these are consistent with past control variables that are used in the literature. To control for regulatory boundaries and home bias, a dichotomous variable is included for whether the trade flows cross a state boundary. To control for industrial clusters that overlap state boundaries, a dichotomous variable is included for whether the destination state of a trade flow is adjacent to the trade flow's state of origin. All control variables are contemporaneous with the dependent variable, although there is no possibility of reverse causality in the regression models. In other words, trade flows in the United States cannot alter state borders or the location of states, which is an important distinction to make when attempting to determine a causal relationship. Since this is an intranational study of trade and not an international study of trade, numerous control variables for different institutions (e.g., levels of democratization) are unnecessary. These simple variables are sufficient to control for confounding factors that may underlie my models.

Taxation Variables

There are two key taxation variables that test for the influence of corporate taxation on trade flows in the United States. The first variable is a type of tax burden variable and directly tests whether corporate taxes decrease the level of economic activity. A corporate tax burden variable is included for both the state of origin and the destination state, as both endpoints are analyzed. These corporate tax burden variables utilize data that are composed of lagged two-year averages (e.g., 2010–11 for 2012), and they are calculated by dividing the sum of state-level corporate taxes and fees by the relevant state's GSP. These ratio variables are based on those used in the literature to study the influence of business taxation on economic outcomes (Prillaman and Meier 2014), and the data were retrieved from the U.S. Census Bureau. The process of using lagged two-year averages mitigates problems related to economic volatility and reverse causality. As the name implies, when the burden of corporate taxation on a state economy increases, economic activity will decrease.

The second variable is the corporate tax rate spread between the origin and destination locations. The corporate tax rate spread will

capture tax incentives that aim to induce large corporations with operations in multiple states to alter their activities away from relatively high corporate tax states. As the tax rate differential increases, the higher cost of doing business in the destination state will result in a decrease in the volume of trade flows passing to that state through corporate supply chains. This reflects the activity of large corporations moving capital assets, such as wholesaling operations, out of relatively higher tax states. This second measure of taxation was calculated by subtracting the top corporate tax rate in the origin state from the top corporate tax rate in the destination state, and the state tax rate data was retrieved from the Tax Policy Center. By using the highest rates, this tax rate spread will capture incentives for *larger firms*, the businesses that are in the best position to engage in tax planning. The tax rate spread variable is contemporaneous with the dependent variable, but was legally established prior to the relevant year.

The two gravity models are specified as follows:⁵

1. $\ln(\text{Value}_{ij}) = \alpha_0 + \beta_1 \ln(\text{Market_Mass}) + \beta_2 \ln(\text{Distance}_{ij}) + \beta_3 \text{Border} + \beta_4 \text{Adjacent_State} + \varepsilon$
2. $\ln(\text{Value}_{ij}) = \alpha_0 + \beta_1 \ln(\text{Market_Mass}) + \beta_2 \ln(\text{Distance}_{ij}) + \beta_3 \text{Border} + \beta_4 \text{Adjacent_State} + \beta_6 \text{Corporate_Tax_Burden}_i + \beta_7 \text{Corporate_Tax_Burden}_j + \beta_8 \text{Tax_Spread}_{ij} + \varepsilon$

The *i* subscript denotes an origin state variable. The *j* subscript denotes a destination state variable. The *ij* subscript denotes a directional variable. Model 1 above represents the base model. Model 2 extends Model 1 with the introduction of taxation variables. Summary statistics for all variables are located in Table 2.

Empirical Findings

The results for both models are depicted in Table 3. Model 1 shows that the gravity model is accurately specified. The presence of a state border reduces trade, and this could be due to regulatory compliance issues and other transaction costs. The dichotomous

⁵All variables are contemporaneous for 2012, except for three categories. GSP variables, corporate tax burden variables, and mobile capital variables are averages for 2010–11. Averaging of past values is intended to mitigate endogeneity in the models and economic volatility following the 2008 financial crisis.

TABLE 2
SUMMARY STATISTICS FOR VARIABLES

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
ln (Trade Transaction Values)	4,426	6.829036	1.996258	0.01479	14.18976
ln (Market Mass)	4,989	24.245040	1.453421	20.457950	29.087520
ln (Distance)	4,963	6.950687	0.883742	2.302585	9.070158
Border	4,989	0.979956	0.140165	0	1
Adjacent State	4,989	0.085789	0.28008	0	1
Origin State Tax Burden	4,989	0.377243	0.302528	0.007304	1.846524
Destination State Tax Burden	4,989	0.379506	0.308094	0.007304	1.846524
Corporate Tax Spread	4,989	0.005302	3.966293	-12	12

variable for adjacent states suggests that there could be industry clusters that overlap borders, increasing trade between states that share a common border. As the distance between the place of origin and destination of trade increases, the level of trade decreases in kind; given the cost of transportation, this is quite intuitive. In summary, specification of the gravity models has an intuitive and predictable impact on economic activity.

The findings from Model 2, with the taxation variables, display the detrimental effect that corporate taxation can have on economic activity. As the spread in the top corporate tax rate increases between the destination and origin states, there is a sharp decrease in the value of trade. According to a direct interpretation of the exponentiated coefficient in Model 2, a 1 percentage point increase in the corporate tax spread between two locations will result in approximately a 1.7 percent *decline* in trade flows. This finding provides evidence that states with relatively higher corporate tax rates may be experiencing business exits that are consistent with the findings in Prillaman and Meier (2014). This effect is compounded by how both variables for state corporate tax burdens decrease trade flows as the ratio of corporate taxes to GSP

TABLE 3
REGRESSION OUTPUTS FOR
CORPORATE TAX GRAVITY MODEL

Dependent Variable:	ln (Trade Transaction Values)			
	Model 1		Model 2	
	Coef.	T-Stat.	Coef.	T-Stat.
ln(Market Mass)	0.986***	(68.39)	0.976***	(68.50)
ln(Distance)	-1.085***	(-27.91)	-1.120***	(-29.34)
State Border	-1.072***	(-5.37)	-0.968***	(-4.76)
Adjacent States	0.547***	(6.44)	0.479***	(5.82)
Corporate Tax Spread	—	—	-0.017***	(-5.16)
Origin Tax Burden	—	—	-0.522***	(-5.24)
Destination Tax Burden	—	—	-0.271***	(-4.19)
Constant	-8.747***	(-21.06)	-8.046***	(-19.19)
Observations	4,424		4,424	
F-Statistic	1,420.94		946.05	
R-Squared	0.82		0.83	

NOTES: * denotes $p < 0.10$; ** denotes $p < 0.05$; and *** denotes $p < 0.01$ (reported p-values are two-tailed tests, although all hypotheses are directional). The use of two-tailed tests makes these results highly conservative, emphasizing the robustness of the findings. All models are estimated with standard errors clustered at the dyadic level with year fixed-effects, using the LSDV method. The dummy variable for 2012 is omitted from the table. These modeling techniques will correct for spatially correlated error terms, which may bias model results, and unaccounted-for intertemporal effects.

increases in all locations. For states where the corporate tax burden is equal to 1 percent of GSP, trade exports are 40.7 percent lower than an otherwise equal state with a corporate tax burden of zero. Similarly, states where the corporate tax burden is equal to 1 percent of GSP will have trade imports that are 23.7 percent lower than an otherwise equal state with a corporate tax burden of zero. Though these hypothetical comparisons represent immoderate

examples (see the summary statistics in Table 2 for the minimum and maximum observation values), the findings are nevertheless profound. In summary, the findings contained herein suggest that increased levels of state corporate taxation can be severely harmful to economic activity *everywhere* in the country.

Conclusion

My findings suggest that state politicians can be active players in promoting economic growth by cutting corporate taxes. Corporations are rational actors that are responsive to their shareholders and government incentives to invest and relocate. States with relatively lower tax rates appear to be in a strong position to benefit from their business-friendly tax policies.

States with relatively higher corporate taxes should look to corporate tax rate cuts to spur economic growth. Corporate tax rate cuts should of course be accompanied by other growth-enhancing policies, such as increasing the flexibility of a state's workforce. Pro-business policies are multidimensional and not composed of a single policy domain. However, corporate tax rate cuts are a solid starting point for improving a state's business environment.

The enactment of pro-business policies in a state will not go unnoticed. There are an abundance of well-trained consultants to help corporations find ways to retain as much revenue as possible. These consultants decrease informational asymmetries that can spur corporate responsiveness. This article provides evidence for some conventional beliefs among policymakers about spurring state economies—namely, that corporate taxation matters.

References

- Andrews, D. (1994) "Capital Mobility and State Autonomy: Toward a Structural Theory of International Relations." *International Studies Quarterly* 38 (2): 193–218.
- Arauzo-Carod, J. M.; Liviano-Solis, D.; and Manjón-Antolín, M. (2010) "Empirical Studies in Industrial Location: An Assessment of Their Methods and Results." *Journal of Regional Science* 50 (3): 685–711.
- Armstrong, C. S.; Blouin, J. L.; and Larcker, D. F. (2012) "The Incentives for Tax Planning." *Journal of Accounting and Economics* 53(1–2): 391–411.

- Atems, B. (2015) "Another Look at Tax Policy and State Economic Growth: The Long-Run and Short-Run of It." *Economics Letters* 127: 64–67.
- Bania, N.; Gray, J. A.; and Stone, J. A. (2007) "Growth, Taxes, and Government Expenditures: Growth Hills for U.S. States." *National Tax Journal* 60 (2): 193–204.
- Basinger, S. J., and Hallerberg, M. (2004) "Remodeling the Competition for Capital: How Domestic Politics Erases the Race to the Bottom." *American Political Science Review* 98 (2): 261–76.
- Bonner, S. E.; Davis, J. S.; and Jackson, B. R. (1992) "Expertise in Corporate Tax Planning: The Issue of Identification Stage." *Journal of Accounting Research* 30: 1–28.
- Browning, E. K., and Johnson, W. R. (1984) "The Trade-Off between Equality and Efficiency." *Journal of Political Economy* 92 (2): 175–203.
- Cerny, P. G. (1990) *The Changing Architecture of Politics*. London: Sage Publishing.
- Deloitte LLP. (2015) "2015 Transparency Report." (August).
- Drucker, J. (2010) "Google 2.4% Rate Shows How \$60 Billion Lost to Tax Loopholes." *Bloomberg* (October 21).
- Eisinger, P. K. (1988) *The Rise of the Entrepreneurial State: State and Local Economic Development Policy in the United States*. Madison: University of Wisconsin Press.
- Friedman, M. (1967). "Fiscal Responsibility." *Newsweek* (August 7): 68.
- Grieco, J. M. (1988) "Realist Theory and the Problem of International Cooperation: Analysis with an Amended Prisoner's Dilemma Model." *Journal of Politics* 50 (3): 600–24.
- Hasseldine, J.; Holland, K.; van der Rijt, P. (2011) "The Market for Corporate Tax Knowledge." *Critical Perspectives on Accounting* 22 (1): 39–52.
- Head, C. K.; Ries, J. C.; and Swenson, D. L. (1999) "Attracting Foreign Manufacturing: Investment Promotion and Agglomeration." *Regional Science and Urban Economics* 29 (2): 197–218.
- Hillberry, R., and Hummels, D. (2003) "Intranational Home Bias: Some Explanations." *Review of Economics and Statistics* 85 (4): 1089–92.
- Jensen, N. M. (2013) "Domestic Institutions and the Taxing of Multinational Corporations." *International Studies Quarterly* 57 (3): 440–48.

- Kurzer, P. (1993) *Business and Banking: Political Change and Economic Integration in Western Europe*. Ithaca, N.Y.: Cornell University Press.
- Leblang, D. (2010) “Familiarity Breeds Investment: Diaspora Networks and International Investment.” *American Political Science Review* 104 (3): 584–600.
- Lewis-Beck, M. S., and Stegmaier, M. (2013) “The VP-Function Revisited: A Survey of the Literature on Vote and Popularity Functions after over 40 Years.” *Public Choice* 157 (3–4): 367–85.
- McGuire, S. T.; Omer, T. C.; and Wang, D. (2012) “Tax Avoidance: Does Tax-Specific Industry Expertise Make a Difference?” *Accounting Review* 87 (3): 975–1003.
- Noguer, M., and Siscart, M. (2005) “Trade Raises Income: A Precise and Robust Result.” *Journal of International Economics* 65 (2): 447–60.
- Ojede, A., and Yamarik, S. (2012) “Tax Policy and State Economic Growth: The Long-Run and Short-Run of It.” *Economic Letters* 116: 161–65.
- Plümpert, T.; Troeger, V. E.; and Winner, H. (2009) “Why Is There No Race to the Bottom in Capital Taxation?” *International Studies Quarterly* 53 (3): 761–86.
- Prillaman, S. A., and Meier, K. J. (2014) “Taxes, Incentives, and Economic Growth: Assessing the Impact of Pro-business Taxes on U.S. State Economies.” *Journal of Politics* 76 (2): 364–79.
- Reed, W. R. (2008) “The Robust Relationship between Taxes and U.S. State Income Growth.” *National Tax Journal* 61 (1): 57–80.
- _____ (2009) “The Determinants of U.S. State Economic Growth: A Less Extreme Bounds Analysis.” *Economic Inquiry* 47 (4): 685–700.
- Slemrod, J. (2004) “The Economics of Corporate Tax Selfishness.” *National Tax Journal* 57 (4): 877–99.
- Wolf, H. C. (2000) “Intranational Home Bias in Trade.” *Review of Economics and Statistics* 82 (4): 555–63.

REMOVING DIRECT CONTRIBUTION LIMITS: FREE SPEECH AND DISCLOSURE

Adam G. Byrne

Carol is a politician. She is a state legislator, to be more specific, and a very good one. Quickly rising through the ranks of her party, Carol has come to be revered by both her constituents at home and her colleagues at the Capitol. Propelled by her popularity, Carol seeks to unseat a member of Congress in an upcoming election. As a challenger, Carol faces the typical disadvantages that come with running against an incumbent: she has less experience, less recognition, and less clout. One area where Carol might be able to outperform her opponent, however, is raising money. Carol is a master fundraiser. Her charm and charisma have enabled her to befriend some of the wealthiest individuals in her district.

Dora is a wealthy businesswoman who happens to be close friends with Carol. The two share similar views when it comes to political, economic, and social issues. Dora hopes Carol gets elected and, in order to fully support her campaign efforts, she offers Carol a \$25,000 check. As beneficial as that money would be for Carol, she must refuse it. Under federal contribution limits, she can accept only \$2,600 per election from any one individual.

Contribution limits, therefore, impose significant difficulties. Not only do they burden Carol's efforts to raise money, but they also restrict Dora's interest in spending money the way she desires to express her political beliefs.

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Adam G. Byrne is a student at Loyola Law School in Los Angeles. He thanks Professor Jessica Levinson and an anonymous referee for helpful comments on earlier versions of this article.

The purpose of this article is twofold. First, I will demonstrate why direct contribution limits placed on campaigns ought to be removed. To do this, I will argue that money is in fact speech, and therefore is entitled to strict constitutional protection. Second, I will argue that disclosure requirements can act as an adequate safeguard against corruption.

Buckley's Basics

Buckley v. Valeo, decided in a 1976 per curiam opinion, is the wellspring of campaign finance law. Many of the Supreme Court cases dealing with campaign finance revolve around this decision, and an abundance of academic literature also draws upon it in one way or another. This section examines the *Buckley* Court's decision and its implications.

Giving and Spending Money Equals Speech

The idea that money is equal to speech underlies the holding of the Supreme Court in its seminal campaign finance case, *Buckley v. Valeo* (424 U.S. 1 [1976]). The *Buckley* Court repeatedly affirmed its belief that contribution and expenditure limitations violate the First Amendment rights of political expression and association (424 U.S. at 14, 23, 54). The Court is led to conclude that money is equal to speech by the fact that "virtually every means of communicating in today's mass society requires the expenditure of money" (424 U.S. at 19). For example, imagine that Senator Chuck Schumer (D-NY), as part of his 2016 reelection campaign, wanted to send out a mailer boasting of all his accomplishments since he took office. To do this, he would have to spend tens of thousands of dollars to mail the flyer to his constituents in New York City, let alone the entire state. Of course, he has alternative options. He could knock on doors or drive a soundtruck around Manhattan, opting not to spend any money communicating. That this alternative gives him any chance of winning the election, however, strains credulity. He could also rely on free social media platforms to disseminate his speech, but many successful political campaigns end up paying for social media ads or outsourcing their accounts to help increase their presence on these platforms. For a New York senatorial candidate, as for any viable contender, spending money is a necessary part of communicating.

To give another example, imagine that a wealthy actor wants to become more politically active and, in order to be seen as a stalwart Democrat, makes a sizeable donation to the California Democratic Party. He could have simply drafted a press release, e-mailed it to a tabloid, and had them blog about it free of charge. But by actually donating money, he is making a statement that he truly stands behind the Democratic cause. Regardless of whether one thinks money is or is not speech, the old maxim still stands: “Actions *speak* louder than words.”

In equating money with speech, the *Buckley* Court gives an analogy of its own: “Being free to engage in unlimited political expression subject to a ceiling on expenditures is like being free to drive an automobile as far and as often as one desires on a single tank of gasoline” (424 U.S. at 19). Many who oppose the idea that money is equal to speech take this metaphor and deconstruct it. For instance, even after the car runs out of gas, the driver is still able speak, for speech is unlimited (Levinson 2013: 898). Arguments such as this, however, misconstrue the metaphor. The car is what symbolizes speech, not the driver of the car. So when the car (speech) runs out of gas (hits expenditure ceiling), the car can no longer drive (no more speech).

Despite these examples and analogies, some believe that, because money is merely the antecedent to speech, it does not merit the same level of First Amendment protection as speech itself (Levinson 2013: 899). A brief search in any ordinary dictionary would support this position. According to the *American Heritage Dictionary*, “speech” is defined as “the faculty or act of speaking; the act of expressing or describing thoughts, feelings, or perceptions by the articulation of words.” No connection with money can be drawn from this definition, so it is quite obvious that money is not speech in the literal sense. However, it is the *act* of giving and spending money that equals speech.

Even then, some claim that although the act of giving and spending money does serve an expressive function, it is not sufficient to fall within the protection afforded by the First Amendment. In explaining this position, Professor Deborah Hellman (2011: 967) notes how money facilitates the exercise of other fundamental rights and yet spending money in connection with some of these rights is not protected. As an example, she brings up the fundamental right to vote and how that does not include the right to buy or sell votes (Hellman 2011: 976). This is true because money has nothing to do with the

right to vote. Giving and spending money, however, has become an integral part of the right to speak freely in the context of political campaigns. One's donation or receipt of a political contribution constitutes a form of political expression and therefore it merits constitutional protection. In other words, giving and spending money is so closely linked to the right to free speech that, unlike buying and selling votes, it should fall within that right's penumbra. Therefore, "the right to speak necessarily encompasses the right to pay for the speech, just as the right to counsel encompasses the right to pay for a lawyer and the right to free exercise of religion includes the right to contribute to a place of worship," as Levy (2010) argues.

Justice White's opinion in *Buckley* touches directly on the question of whether giving and spending money equals speech. In his mind, "the argument that money is speech and that limiting the flow of money to the speaker violates the First Amendment proves entirely too much" (424 U.S. at 262). Money alone, he argues, is not speech. Only speech is speech. Money simply facilitates someone's speech. To support his argument, Justice White points out that many campaign activities are not themselves communicative or even remotely related to speech. After all, spending campaign money on a candidate's travel and lodging expenses surely does not amount to that money being used for political expression. Justice White goes on, however, to discuss how all campaigns differ, with some spending less money while still managing to communicate more. Yet to suggest that the level of efficiency with which candidates use their money has any sort of impact on the constitutional protection afforded to their speech is to stray away from the issue at hand. Whether candidates spend every penny wisely or waste a million dollars on an unsuccessful television ad, their political speech still deserves to be free from any government restriction limiting its amount.

Justice Stevens echoes Justice White's concerns in *Nixon v. Shrink Missouri Government PAC* (528 U.S. 388 [2000]), but offers an alternative reason for why money is not equal to speech. In his view, "Money is property; it is not speech" (see Sharma 2008). Money, whatever its worth, is simply a medium of exchange. Speech, on the other hand, has nouns, verbs, and adjectives. It has the "power to inspire volunteers to perform a multitude of tasks on a campaign trail, on a battleground, or even on a football field," tasks that money can only pay people to do (528 U.S. at 398). Generally, deprivations of

property violative of the Fifth Amendment receive a lower level of constitutional scrutiny than infringements on speech that contravene the First Amendment. The Constitution, therefore, protects the right to “use one’s own money to hire gladiators,” but a deprivation of this property interest by enacting campaign finance restrictions deserves a lower level of review than an act depriving one of the right to say what one pleases. Mixing money and speech, it seems, would be a category error.

Unfortunately for Justice Stevens, the argument that money is completely distinct from speech quickly falters. As the late Justice Scalia stated in a recent interview, “You cannot separate speech from the money that facilitates the speech. It is utterly impossible.”¹ A glimpse at newspapers proves his point. If the government were to tell publishers that they could only spend so much money in publishing their newspapers, the government would certainly be abridging their freedom of speech, not just depriving them of property. Even Justice Breyer, who has consistently voted in favor of upholding campaign finance restrictions, has admitted that there is an identifiable link between money and speech. In his opinion, “a decision to contribute money to a campaign is a matter of First Amendment concern—not because money is speech (it is not); but because it *enables* speech” (528 U.S. at 400). Under this line of reasoning, permitting restrictions on the amount of money is a way for the government to *disable* speech.

For the government to wield such control over speech, any use of that power must be subject to strict constitutional scrutiny. Otherwise, it would be as if the government could tell an author that he has the right to publish so long as his book does not exceed 1,000 pages, or tell an artist that he has the right to paint so long as he spends no more than \$1,000 on materials.

Contribution and Expenditure Limits

The *Buckley* Court established a number of things after it was asked to look at amendments made to the Federal Election Campaign Act (FECA). For one, the Court differentiated its treatment of contribution limits and expenditure limits (424 U.S. at 23).

¹CNN interview with Justice Scalia (July 18, 2012). A transcript of the interview is available at transcripts.cnn.com/TRANSCRIPTS/120718/pmt.01.html.

Contributions, it was found, are a type of speech by proxy. The act of giving money to a person and enabling that person to spend it for campaign purposes warrants a lower level of constitutional protection than the act of spending money directly for campaign purposes (424 U.S. at 20–21). The existence of an intermediary, then, renders limitations on the activity susceptible to a lesser form of judicial scrutiny. Expenditures, on the other hand, involve no such intermediary. An expenditure limit simply restricts the amount of overall spending a person can make relative to a clearly identified candidate. It was thought that this type of action is more reflective of the spender's expression, and therefore it merits a higher standard of review. By distinguishing contributions and expenditures, the Court essentially upheld limitations on the former and struck down restraints on the latter.

The Court needed to put forth clear reasoning in order to justify why it was gutting half of FECA's amendments. In regards to contribution caps, it found only "a marginal restriction" was imposed upon the contributor's interests as a speaker (424 U.S. at 20). Such interferences are appropriate if the government has a sufficiently important interest. Here, the Court viewed the government's interests, namely, curbing corruption or the appearance of corruption, as sufficiently important. After all, "to the extent that large contributions are given to secure a political quid pro quo from current and potential officeholders, the integrity of our system of representative democracy is undermined" (424 U.S. at 26–27). Although opponents might have argued that such a limitation is too sweeping and that bribery laws were already in place, the Court paid deference to Congress' decision that bribery laws alone do not suffice. It was believed that the Court "has no scalpel to probe" in determining which contribution level works best (424 U.S. at 30).

With respect to expenditure limitations, the Court found a "severe restriction" was placed on the spender's interests as a speaker (424 U.S. at 23). Given this greater burden, the Court applied a heightened standard of review. It held that the government's interest in curbing corruption or its appearance was insufficient to justify the expenditure ceilings it had imposed. Furthermore, the Court found that the expenditure limits were insufficiently tailored. A candidate, it was reasoned, was less susceptible to corruption when spending money than when receiving money. Challengers to FECA argued

that the law's \$1,000 limitation on expenditures "relative to a clearly identified candidate" was overly vague and ambiguous. The Court responded by interpreting the phrase "relative" as meaning "advocating the election or defeat of a candidate" (424 U.S. at 42). The Court further narrowed this interpretation as meaning any "explicit words of express advocacy" such as "vote for," "elect," "support," "defeat," or "reject" (424 U.S. at 44). The result of this magic-words test was that campaigns could still put out ads without being subject to the \$1,000 spending limit by simply not including any of these magic words. Consequently, the law limited only a narrow scope of spending—that is, spending on ads using a magic word—and was in the end ineffective in carrying out the government's goal of deterring corruption.

Rejecting Alternative Frameworks

The Court's complex bifurcation and disputed metaphors aside, it must be remembered that fundamental rights are never absolute. One's freedom of expression, for example, can be limited in several respects. Consider *U.S. v. O'Brien* (391 U.S. 367 [1968]), where a man was convicted for burning his draft card on the steps of a Boston courthouse. While his actions had an expressive component—namely, protesting the war in Vietnam—the Court upheld his conviction because it found the law to be targeted at the nonexpressive component of his actions. Symbolic speech, therefore, does not always receive complete constitutional protection.

The *Buckley* Court explicitly addressed *O'Brien* and considered whether or not it could take a similar approach. It ultimately concluded that *O'Brien* could not apply to the FECA amendments because there is a distinct difference between the expenditure of money and the destruction of a draft card. Unlike the actual burning of a card, which introduces a nonexpressive component that is subject to government restriction, it has never been suggested that "the dependence of a communication on the expenditure of money . . . introduce[s] a non-speech element or reduce[s] the exacting scrutiny required by the First Amendment" (424 U.S. at 16). In other words, because the purpose of the FECA amendments was to limit potentially harmful speech, as opposed to regulating conduct wholly unrelated to the interest of suppressing speech, the *O'Brien* framework proved inappropriate and inapplicable.

Speech can also be regulated by time, place, and manner restrictions. These limitations must be content-neutral, meaning they apply to speech regardless of who is speaking and the topic they are speaking about. For example, a Los Angeles municipal ordinance prohibiting any protests on Hollywood Boulevard at 8 a.m. on Mondays would most likely be upheld as constitutional. The regulation is not aimed at any specific speaker or topic; rather, it serves the government's interest in ensuring freedom of movement on public roads. The *Buckley* Court opted not to treat the expenditure ceilings of the FECA amendments as time, place, or manner restrictions. It reasoned that unlike time, place, or manner regulations, the expenditure limitations imposed "direct quantity restrictions on political communication" (424 U.S. at 18).

Those defending the FECA amendments compared money with a sound truck: the more money involved, the higher the decibel level of the sound truck. The Court rejected this analogy, finding that requiring a sound truck to reduce its decibel level still leaves other channels of communication available to the speaker, while the same is not so with expenditure limitations. It has been argued that the sound truck analogy still holds water because even after FECA's enactment, speakers remain free to say as much as they want (Levinson 2013: 896). This assertion, however, assumes giving money is not speech, something that *Buckley* explicitly rejected. If speakers truly remained free to say as much as they wanted under FECA's limits, then they would be able to spend any amount of money to support whatever candidate or cause they chose (424 U.S. at 19–20).

The Court's Mishandling of Contribution Limits

As previously explained, the *Buckley* Court struck down FECA's expenditure limits but upheld its contribution limits. Recognizing the "opportunities for abuse inherent in a regime of large individual financial contributions," the Court found the contribution limits to be only marginal restrictions on freedom of speech, which were sufficiently justified by the government's interest in preventing corruption and the appearance of corruption (424 U.S. at 26–27). It reasoned that an individual or organization's contribution to a specific candidate only communicates a general expression of support that does not contain a specific message. In the Court's view, whether a donor contributed \$10 or \$1,000, the message of support remains the same. "At most," the Court explained, "the size of the

contribution provides a very rough index of the intensity of the contributor's support for the candidate," making any limit on that amount have "little direct restraint on his political communication" (424 U.S. at 21).

However, the Court's suggestion that the amount a donor contributes matters much less than the simple fact of the contribution is untenable. To illustrate, imagine a donor with an average income gave \$1,000 to a city council candidate who was a Democrat bent on education reform. This donor is likely one who cares significantly about local affairs, favors Democratic causes, and wants to see a change in the way education is run. Now imagine that same donor gave \$50 to an incumbent state senator. This shows that the donor is likely content with the way things are run at the state level and cares far more about the composition of her city council. The way people spend their money reveals an abundance of information about them and giving money to a campaign is no exception. Donors, therefore, can convey a number of different messages simply through the magnitude of their donations. The "intensity of the contributor's support," to use the Court's words, should be free from government restriction because any limit placed upon it imposes a substantial restraint on the various political messages conveyed. Contributing money is not just a mere "general expression of support;" rather, it is one way that an individual engages in what should be constitutionally protected political speech.

Nevertheless, the Court tried to support its argument by pointing out how contributions are a type of speech by proxy, requiring the presence of an intermediary to convey the communication that is sought to be expressed. After one contributes money to a candidate or association, "the transformation of contributions into political debate involves speech by someone other than the contributor" (424 U.S. at 21). The Court's analysis here may seem accurate, but it is mistaken. Just because the money is one step removed from entering political debate has no bearing on whether or not the contribution should be protected. Once the contribution has been made, the political expression of the donor has been established. What actually happens to the money after that point in time is irrelevant. Whether it is used to pay for consultants, yard signs, or coffee, donors have still told the recipient and disclosed to the world at large that they stand in support of that cause. Furthermore, even expenditures usually require the use of an intermediary to get the message across, be it a

television station or a newspaper publisher (528 U.S. at 413). The existence of an intermediary therefore does little to distinguish contributions and expenditures, yet the *Buckley* Court relied on this slender reed to justify its bifurcation.

Finally, the *Buckley* Court looked at political action committees (hereinafter “PACs”) and struck down the FECA amendments’ ceiling on independent expenditures. Given the ability of PACs to make unlimited independent expenditures, the Supreme Court has often used this option to justify contribution limits (see *Austin v. Michigan Chamber of Commerce*, 494 U.S. 652, 660 [1990]). Once individuals reach the contribution limit, the thinking was that they could still speak by making an unlimited amount of independent expenditures. As the Supreme Court has come to recognize, however, the PAC option is not as easy as once made out to be. There are in fact a number of administrative costs involved with establishing and maintaining these types of committees. PACs are often required to appoint a treasurer, keep records of all contributions, file a statement of organization, send frequent reports to the Federal Election Commission (FEC), and make known any changes to the above information (see *FEC v. Mass. Citizens for Life, Inc.*, 479 U.S. 238, 253 [1986]). In his majority opinion in *Citizens United v. FEC* (558 U.S. 310, 337 [2010]), Justice Kennedy held that a prohibition on corporations and labor unions from using their general treasury funds to make independent expenditures was an outright “ban on corporate speech notwithstanding the fact that a PAC created by a corporation can still speak.” Under this view, PACs do not alleviate the First Amendment problem because they are entirely separate entities. Their formation still does not allow the original contributor to speak. The existence of PACs, therefore, provides no justification for preserving contribution limits.

Overall, the contribution limits that were upheld in *Buckley* have proven to be largely ineffective. They have forced a substantial amount of political speech underground as contributors and candidates have devised elaborate methods of avoiding them. Along with this rise in “soft money,” *Buckley’s* framework has also created an indirect system of accountability that confuses voters (528 U.S. at 408). Instead of rooting out corruption, *Buckley’s* bifurcation has created an endless cycle of fundraising where incumbents must solicit money from as many donors as possible instead of spending that time fulfilling their duties in office.

REMOVING DIRECT CONTRIBUTION LIMITS

It is also important to ask why contribution limits are set the way they are. Chief Justice Burger's opinion in *Buckley* raised this issue and he noted that "it is clearly arbitrary Congress has imposed the same ceiling on contributions to a New York or California Senatorial campaign that it has put on House races in Alaska or Wyoming" (424 U.S. at 242). Why the majority in *Buckley* thought that a \$1,000 contribution limit was enough to prevent corruption but not \$900 or \$1,100 is also a question left unanswered. However, the Court reasoned, as it almost invariably has since, that it has no scalpel to probe which contribution level works best, for such a determination ought to be made in the galleries in Congress (424 U.S. at 30). Justice White, a staunch proponent of contribution limits, believed that Congress was entitled to determine that bribery and disclosure laws were not enough. He accepts the Congressional judgment that "the evils of unlimited contributions are sufficiently threatening to warrant restriction regardless of the impact of the limits on the contributor's opportunity for effective speech" (424 U.S. at 260).

Such deference to Congress, though, is fraught with danger. It must not be forgotten that many of the legislators setting these limits will run for reelection in the future. Naturally, they may feel inclined to manipulate the contribution limits to their advantage and protect their seats. Incumbents enjoy a variety of advantages by virtue of already being in office. They also tend to disfavor large contributions because a challenger who is able to receive an unlimited amount in contributions has a much better shot at beating them than a challenger who can only take in a certain amount per donor. It is no coincidence that over half of the legislators in Congress have a net worth of over \$1 million (see Choma 2014); those with substantial amounts of money will win more often than those without such funds. In light of the potential for incumbency protectionism, the fact that contribution limits are set by legislators themselves casts more doubt on the validity of such restrictions.

In sum, the act of giving and spending money on political activities is equal to political speech and deserves the full protection of the First Amendment. Contribution limits "infringe as directly and as seriously upon freedom of political expression and association as do expenditure limits" (*Colorado Republican Federal Campaign Committee v. FEC*, 518 U.S. 604, 640 [1996]). Being "two sides of the same First Amendment coin," contribution limits should be

subject to the same heightened level of scrutiny as expenditure limits (424 U.S. at 241).

Disclosure Requirements

If the current election system is to function properly without contribution limits, then the weight of democracy is put on the back of disclosure requirements. Compelling disclosure of political activity requires a careful balancing of the individual's interest in privacy and anonymity with the public's interest in access and transparency. As will be demonstrated, this balancing will generally, but not always, tip in favor of the public's interest in information. An aware and well-informed polity has always been an indispensable part of what makes the United States government work. In 1788, American patriot Patrick Henry remarked that "the liberties of a people never were, nor ever will be, secure when the transactions of their rulers may be concealed from them."² Over two centuries worth of elections, scandals, and reforms later, his words remain relevant.

Sunlight Is the Best Disinfectant

Along with contribution and expenditure limits, *Buckley v. Valeo* considered an as-applied challenge to FECA's disclosure provisions. FECA required that each political committee register with the FEC and keep detailed records of all the funds it was receiving and spending (2 U.S.C. § 431 *et seq.* [1970 ed., Supp. IV]). Committees and candidates were also required to file quarterly reports with the Commission that included information such as the names, addresses, and occupations of those involved. Opponents of the law claimed that it was overbroad in its application to minor-party and independent candidates and that it should not have included contributions as small as \$11. While the Court recognized that compelled disclosure can seriously infringe on rights guaranteed by the First Amendment, it ultimately found that the asserted governmental interests were sufficiently important to outweigh the possibility of such infringement (424 U.S. at 64–66).

The Court found three governmental interests underlying FECA's disclosure requirements. One was the public's interest in

²Patrick Henry, "Speech On the Expediency of Adopting the Federal Constitution," Delivered in the Virginia Ratifying Convention (June 9, 1788).

information. If the public had access to data displaying all the exchanges of political money, it would be able to judge who was indebted to whom and get a better sense of what a candidate or organization truly stood for. Another interest was the deterrence of corruption or the appearance of corruption. Whether money was being used improperly, or special favors were being granted, would become more perceptible to the public eye and therefore less likely to occur. Lastly, the recordkeeping requirements would serve the government's enforcement interest and make violations easier to detect. "Electric light," as Justice Brandeis (1914: 92) aptly put it, is "the most efficient policeman."

With these three interests in mind, the Court next looked at the burdens that FECA's disclosure requirements place on individuals and associations. It conceded that public disclosure would likely deter some individuals from contributing and may even subject some donors to harassment or retaliation (424 U.S. at 68). The Court had previously held that such threats of retaliation can be sufficient to exempt a candidate or association from disclosure, but only where there has been an uncontroverted showing that such threats have occurred on past occasions (see *NAACP v. Alabama*, 357 U.S. 449, 462 [1958]). In *Buckley*, however, there was no evidence in the record of such threats or reprisals occurring or even being likely to occur. Instead, these potential threats against contributors were found to be "highly speculative," and therefore the disclosure requirements did not seriously infringe on First Amendment rights (424 U.S. at 70). With respect to the provision's low threshold, the Court shed light on the fact that Congress did not focus on setting the perfect level but rather adopted the threshold existing some 65 years prior. Nonetheless, because of the complexity of the legislation, the Court paid deference to Congress in making this determination (424 U.S. at 83). In the end, the Court facially upheld the law's disclosure requirements, seemingly hesitant to tamper with them after having already eviscerated half of FECA.

The merits of *Buckley* aside, there are many advantages in requiring disclosure. An open and transparent campaign finance system enables the public to hold candidates and committees accountable for their fundraising. To illustrate, imagine the executive of a natural gas company donated the maximum amount of direct contributions to every Republican senator. Each of these senators then voted against an anti-fracking bill. By virtue of disclosure, the public would

be able to judge for themselves the motivations of their legislators, whether it be serving constituents or pleasing special interests. This would remain just as true in a world without contribution limits. If a wealthy woman named Claire were to contribute \$3 million directly to a candidate's campaign efforts, that candidate would be identified by the public as "Claire's Candidate" and the public would be much more sensitive to any special favors Claire might one day receive.

Disclosure requirements can also serve to help contributors. Donors who want to make it known to the public who they support and how much they support them can use disclosure as a tool to amplify their political expression. It is one thing for individuals to claim that they stand behind certain candidates or issues; it is an entirely different thing to be able to examine the records and see the numerous contributions those individuals have made in the past.

Mandatory Disclosure's Impact on Individual Autonomy

For all of the benefits inherent in disclosure requirements, there remain a number of encumbrances that they impose on contributors. Chief Justice Burger raised several of these burdens in his separate opinion in *Buckley*. For instance, he agreed with the Court that these requirements would deter some individuals who might otherwise contribute. To illustrate, he cites rank-and-file union members and rising junior executives who might not want to run the risk of being caught supporting causes unpopular with upper management (424 U.S. at 237). The Chief Justice claimed that the "public interest" was too vague and ill-defined to risk breaching the historic safeguards of the First Amendment.

What the Chief Justice overlooked, however, is that the First Amendment not only encompasses the value of self-expression, but it also furthers the public's strong interest in knowing where speech is coming from. If the Chief Justice's examples were taken at full value, then anyone could donate a substantial amount of funds to candidates or committees but then claim exemption from disclosure to conceal their actions because it might reflect negatively on the donor. Speakers cannot expect their expressions to be locked in a box when matters of national magnitude are concerned. For as-applied challenges that make an uncontroverted showing that threats and reprisals have occurred and are likely to keep occurring, the Court has recognized that exemptions can be made. For anything falling short of that standard, the public's interest on this front outweighs that of the individual.

REMOVING DIRECT CONTRIBUTION LIMITS

Disclosure requirements also raise the potential for public confusion. Committees making independent expenditures have been known to use ambiguous names like “Organization for a Better America” or “U.S.A. All the Way,” leaving voters with little to no information about what the group’s mission is or what it stands for. The premise of the First Amendment, though, is that “the American people are neither sheep nor fools, and hence fully capable of considering the substance of the speech presented to them” (*McConnell v. FEC*, 540 U.S. 93, 258 [2003]). So every time a campaign commercial airs on television, viewers are presumed to be able to judge for themselves the substance of the ad, regardless of the ambiguous disclaimer at the end.

Yet another topic of concern related to compulsory disclosure is timing. As mentioned above, one of the governmental interests underlying disclosure is the deterrence of corruption, but exactly *when* corruption is prevented is a difficult question. FECA’s penalty for violation of its record-keeping and reporting provisions was a fine not more than \$1,000 or a prison term of not more than one year, or both. When used in the context of the First Amendment, criminal sanctions like these create many problems. For instance, consider donors who were unsure about whether they needed to disclose. They would have to go through the courts and seek a declaratory ruling in order to avoid possible criminal sanctions. A chilling effect like this runs afoul of the First Amendment.

Further problems arise in the time period just before an election. If in the last week of the campaign donors were to violate the disclosure requirements, their wrongdoing would likely not be revealed until after the election was over and the damage was already done. The fear is that candidates may accept undisclosed funds at the tail end of their campaigns. Here, however, more than just criminal sanctions act as a deterrent. The political repercussions of using unlawful, undisclosed money also help ensure that candidates run proper campaigns in accordance with the law (424 U.S. at 56).

Another controversial facet of disclosure requirements concerns the advent of the Internet. The disclosure provisions examined in *Buckley* took place in a context remarkably different than today. When *Buckley* was decided in 1976, the reports that were sent to the FEC were stored on location, available to anyone willing to travel there and have a look. In contrast, today these reports are compiled in a database and uploaded onto the Internet just a few mouse clicks

away from anywhere in the world. As a result, modern technology has enabled the public to make prompt evaluations of political spending. Critics of disclosure requirements raise an important issue, though, and that is the fact that most Americans do not go to the FEC website and use the information it offers to help them choose who to support and what to vote for. But that was not how disclosure was intended to work. What matters is that any individual who does want such information will be able to access it easily. Nonprofit organizations and media outlets, for instance, put great effort into investigating and analyzing the information, looking for anything worth sharing with the general public.

Still, there are a number of alarming concerns raised by the Internet that are less easy to address. With all of this information readily available to the public, individuals can collect and use it however they would like. Nosey neighbors and personal enemies have access to the names, addresses, and occupations of people who have contributed. Justice Thomas has demonstrated how real this danger is. In his separate opinion in *Citizens United v. FEC*, he described how opponents of a state ballot measure were able to compile the disclosed information and use it to inflict property damage and make death threats targeted at those donors who supported the measure (558 U.S. at 480). Horror stories like this reflect the chilling effect that disclosure requirements can have on speech. The hard-won right to privacy is undeniably impacted by compulsory disclosure. This is especially true for groups who may have unpopular views or unfashionable ideas, the exact groups that the First Amendment was designed to protect.

Many tools, from hammers to knives, can be used as weapons when in the wrong hands, but that does not detract from their utility. Similarly, disclosure is a tool that must be used with great care within the realm of the First Amendment. Ultimately, because the Court has held that disclosure laws appear to be “the least restrictive means of curbing the evils of campaign ignorance and corruption” (424 U.S. at 68), and because exemptions can be made for cases involving an uncontroverted showing of threats or reprisals, some of the most common arguments against mandatory disclosure are unavailing (Hasen 2012: 559).

In addition to the negative aspects of disclosure mentioned above, the true costs of mandated disclosure, according to one commentator, are higher than will ever be known (see Samples 2010). This is because the rate at which disclosure leads citizens to decide not to

donate is something that cannot be accurately measured. Furthermore, some critics believe that compulsory disclosure directs public attention toward the speech's source of funding and away from its content. By promoting the notion that identity matters more than ideas, mandatory disclosure "moves us away from the ideal of popular government" (see Samples 2010). Berkeley Professor Bruce Cain believes that these immeasurable costs could be prevented by "adopting a narrower tailoring of disclosure practices" (see Cain 2010). He recommends a regime of "semi-disclosure," which involves the full reporting but only partial disclosure of campaign donor information. Regardless of whether this suggestion would be successful, proposals like this lend optimism to the belief that a proper balance can be struck between mandatory disclosure and the preservation of individual political rights.

Disclosure in the Modern Campaign Finance Landscape

In the wake of *Citizens United* and its progeny, American elections have experienced an influx of outside spending in the form of independent expenditures. Media outlets and campaign finance reformers have focused a lot of attention on "SuperPACs" and the massive amounts of money they have poured into recent elections. The proliferation of SuperPACs, however, does not pose a problem in a system devoid of contribution limits. This is because SuperPACs are legally required to publicly disclose who their donors are, making them suitable within a regime that would wholly rely on mandatory disclosure. If anything, the presence of SuperPACs argues against the continued use of contribution limits. If donors want to exceed the maximum amount levied upon them by a contribution limit, they now enjoy the option of donating to an outside, independent organization that will use the money to further the donors' political causes.

The real threat of independent expenditures in this post-*Citizens United* world comes from organizations that generally do not have to disclose who their donors are (Heerwig and Shaw 2014). Nonprofit organizations registered under § 501(c)(4) of the Internal Revenue Code are allowed to receive unlimited donations from corporations, unions, and individuals and have been spending that money to influence elections. To maintain their 501(c)(4) status, these organizations must be "primarily engaged in promoting in some way the common good and general welfare of the people of the community." Unlike SuperPACs, which are required to disclose their donors, 501(c)(4)

organizations are permitted to carry out their political activities without having to publicly disclose the identity of their donors.

To completely depend on disclosure, public light must be cast on the money that is being spent by these 501(c)(4) groups. This is especially true given that recent spending by 501(c)(4) organizations has been significant, amounting to 4.7 percent of total election spending in 2014.³ The seemingly obvious solution to this problem is to amend the tax code so that these groups are required to share with the public who their donors are. However, a number of commentators suggest that merely expanding disclosure will not be enough. In their view, the existing disclosure regime is deeply flawed in ways that neither the courts nor reformers have yet acknowledged (Heerwig and Shaw 2014). To give just one example, candidates and political committees are currently not required to use standardized disclosure forms; this obfuscates the information that gets disclosed and makes enforcement of reporting requirements more difficult.

To remedy these existing flaws and create a disclosure regime that offers accessible, comprehensible, and credible information, revisions must be made to the mechanisms by which data on election spending are collected, maintained, and disseminated (Heerwig and Shaw 2014). These changes may require time, as well as some trial and error, but with a determined approach they can ultimately confront and eliminate the controversial problems caused by 501(c)(4) spending.

The More Speech, the Better

When the drafters of the Bill of Rights proposed that “Congress shall make no law . . . abridging the freedom of speech,” they recognized how fundamental it was for individuals to be able to speak their views, beliefs, and ideologies free from governmental restraint. Admittedly, not all types of speech enjoy First Amendment protection (e.g., fighting words). Still other types of speech enjoy protection, but to a lesser degree (e.g., commercial speech). Political expression, however, is at the heart of the First Amendment. Self-government depends on the open exchange of ideas, in which citizens can criticize

³Author calculation based on figures from www.OpenSecrets.org. 501(c)(4) spending has played an outsized role in a number of competitive races. For example, Congresswoman Joni Ernst (R-IA) was backed by over \$17 million of 501(c)(4) spending in support of her 2014 senate campaign.

their leaders and propound unpopular beliefs, all without facing retaliation from the majority. Any laws limiting such speech must be subject to strict constitutional scrutiny.

The premise of democracy requires that “there is no such thing as *too much* speech” (540 U.S. at 259). Therefore, speech must be allowed access to the marketplace of ideas without any barriers to entry. Not everyone agrees, of course. For example, Professor Wayne Batchis posits that a marketplace lacking enforceable rules is no marketplace at all. “Even the most ardent defenders of the free market,” he asserts, “acknowledge the need to limit freedom to avoid monopolistic behavior” (Batchis 2007: 30). Yet while his analysis is critical of a system without contribution and expenditure limits, he omits any reference to disclosure requirements. Compelling candidates, committees, and donors to disclose their political activity would better inform the public and prevent secretive spending, thus ensuring that the marketplace of ideas remains both healthy and competitive.

Another argument put forth by Batchis (2007: 30) is that “more spending on speech can often mean less diversity of opinion and a diminished quality of ideas.” Restrictions on political money, then, ought to promote political speech by giving more breathing room to hear from a greater diversity of speakers (Levinson 2013: 897). This reasoning sounds dangerously close to the principal of equalization, a suggestion that the Court has invariably refused to accept. As was plainly stated in *Buckley v. Valeo*, “the concept that government may restrict the speech of some elements of our society in order to enhance the relative voice of others is wholly foreign to the First Amendment” (424 U.S. at 48–49). That amendment was designed to “assure unfettered interchange of ideas for the bringing about of political and social changes desired by the people” (424 U.S. at 49). Just because one speaker has a louder voice than another does not mean that the former must quiet down for the benefit of the latter. Competition, after all, is what makes the marketplace of ideas work; it should not be inhibited. Some may complain that an unequal playing field built on unlimited funds only results in attack ads devoid of substantive content, but “it is not the proper role of those who govern us to judge which campaign speech has ‘substance’ and ‘depth’ . . . and to abridge the rest” (540 U.S. at 261). An open and robust debate on public issues benefits everyone, large and small donors alike.

In his new book *Plutocrats United*, Professor Richard L. Hasen argues that, in lieu of justifying campaign finance restrictions in terms of the government's anticorruption interest, the courts should construe these restrictions as a means of promoting and preserving political equality. Citing the "inevitable tension between free economic markets and voter equality," he posits that democracy in America will soon be too far skewed toward the interests of the wealthy (Hasen 2016: 11). The accuracy of his prediction aside, the actions necessary to address Hasen's concerns ought to be taken in the political arena, not in the courtroom. Should citizens feel that their representatives are beholden to big donors and special interests, they can voice their feelings in the ballot box and elect better representatives. It is not the judiciary's duty, however, to ensure that all citizens exercise their free speech rights to the fullest extent possible.

The above arguments notwithstanding, there remains the frequently mentioned fear of "drowning out." The thinking is that with an unlimited amount of money entering the marketplace of ideas, well-funded speech is "capable of drowning out other voices and diminishing the likelihood that less generously financed ideas will have any reasonable chance of winning favor with the majority of voters" (Batchis 2007: 30). Justice Stevens gave voice to this concern in his dissent in *Citizens United*. Worried that the opinions of many will be marginalized by large campaign war chests, he believes that uninhibited money would "decrease the average listener's exposure to relevant viewpoints" and "diminish citizens' willingness and capacity to participate in the democratic process" (558 U.S. at 472). To him, there *is* such a thing as too much speech.

In making his arguments, Justice Stevens contradicts himself in two ways. First, he expresses his concern for the integrity of the electoral process and preserving the public's confidence in it, but at the same time he likens the American public to sheep who are unable to judge for themselves the quality and content of the speech they are listening to. Second, he emphasizes his concern that Americans do not have the time of day to sit down and think through all of the speech that is being thrown at them. At the same time, though, he feels the public would be better off if it could hear from a diverse group of speakers with varied perspectives. An ideal world for

Stevens, it seems, is one where everyone's voice can be heard, but in such a world there is a far less chance that Americans will be able to thoroughly contemplate everyone else's speech. Given these two contradictions, the reasoning of Justice Stevens proves rather weak and unconvincing.

Finally, regardless of whether one equates money with speech, there is the risk that more money being given and spent for campaign purposes will only lead to more corruption. "One would be blind to history to deny that unlimited money tempts people to spend it on whatever money can buy to influence an election" (424 U.S. at 265). What makes disclosure requirements so valuable, however, is their ability to root out corruption. "Money, like water, will always find an outlet" (540 U.S. at 224). Instead of trying to plug all of the holes in contribution and expenditure limits, legislators should focus more on keeping track of where the money is coming from and where it is headed to. The American public would be much better off participating in a system with this information available than in a system rife with so-called "dark money."

Conclusion

Ending direct contribution limits and focusing on disclosure are concomitant ideas. A system without donation caps causes a number of potential problems, but compulsory disclosure readily answers them. Under this proposed scheme, Dora the donor would be able to give Carol the candidate her \$25,000 so long as they properly disclose this transaction. That way, Dora can fully exercise her First Amendment right to political expression and the public can be fully informed about her political donation.

Ever since its landmark decision in *Buckley v. Valeo*, the Supreme Court's campaign finance jurisprudence has been in a state of flux. Several cases have reflected the Court's approval of government actions seeking to prevent corruption (e.g., *Austin* and *McConnell*). More recent cases, on the other hand, have demonstrated the Court's commitment to the First Amendment's protection of political expression (e.g., *Citizens United* and *McCutcheon*). If this current trend continues, direct contribution limits may soon be subject to deregulation. Taking the above comments into consideration, this would be a wise move for the Court to make.

References

- Batchis, W. (2007) "Reconciling Campaign Finance Reform with the First Amendment: Looking Both Inside and Outside America's Borders." *Quinnipiac Law Review* 25 (1): 27–72.
- Brandeis, L. D. (1914) *Other People's Money and How the Bankers Use It*. New York: Frederick A. Stokes Co.
- Cain, B. (2010) "Shade from the Glare: The Case for Semi-Disclosure." *Cato Unbound: A Journal of Debate* (November 8). Available at www.cato-unbound.org/2010/11/08/bruce-cain/shade-glare-case-semi-disclosure.
- Choma, R. (2014) "Millionaires' Club: For First Time, Most Lawmakers Are Worth \$1 Million-Plus." *OpenSecrets.org* (January 9). Available at www.opensecrets.org/news/2014/01/millionaires-club-for-first-time-most-lawmakers-are-worth-1-million-plus.
- Hasen, R. L. (2012) "Chill Out: A Qualified Defense of Campaign Finance Disclosure Laws in the Internet Age." *Journal of Law and Politics* 27: 557–73.
- (2016) *Plutocrats United: Campaign Money, the Supreme Court, and the Distortion of American Elections*. New Haven, Conn.: Yale University Press.
- Heerwig, J. A., and Shaw, K. (2014) "Through a Glass, Darkly: The Rhetoric and Reality of Campaign Finance Disclosure." *Georgetown Law Journal* 102: 1443–1500.
- Hellman, D. (2011) "Money Talks But It Isn't Speech." *Minnesota Law Review* 95: 953–1002.
- Levinson, J. A. (2013) "The Original Sin of Campaign Finance Law: Why *Buckley v. Valeo* Is Wrong." *University of Richmond Law Review* 47: 881–937.
- Levy, R. A. (2010) "Campaign Finance Reform: A Libertarian Primer." Washington: Cato Institute. Available at www.cato.org/publications/commentary/campaign-finance-reform-libertarian-primer.
- Samples, J. (2010) "The Costs of Mandating Disclosure." *Cato Unbound: A Journal of Debate* (November 10). Available at www.cato-unbound.org/2010/11/10/john-samples/costs-mandating-disclosure.
- Sharma, M. (2008) "Money as Property: The Effects of Doctrinal Misallocation on Campaign Finance Reform." *University of Michigan Journal of Law Reform* 41: 715–43.

BARBER LICENSURE AND THE SUPPLY OF BARBER SHOPS: EVIDENCE FROM U.S. STATES

Joshua C. Hall and Shree B. Pokharel

The Institute of Justice, a public interest law firm in Washington, D.C., focuses part of its litigation on issues of the ability of citizens to enter and compete in markets unburdened by unnecessary regulations. In a study titled *License to Work: A National Study of Burdens from Occupational Licensing* (Carpenter et al. 2012), the Institute reports that “in the early 1950s, only one in 20 U.S. workers needed the government’s permission to pursue their chosen occupation.” In 2008, that number was estimated to be one in three (Kleiner and Krueger 2013).

Given its pervasiveness, occupational licensing has long been a subject of debate as to whether it serves to protect the public interest or the interests of special interest groups by acting as a barrier to entry. Proponents of occupational licensing argue occupational licensing enables better quality services to consumers that would otherwise not have been provided (Arrow 1971). It has also been argued that occupational licensure encourages prospective entrepreneurs to accumulate human capital in their occupation of choice (Akerlof 1970, Shapiro 1986). Opponents, however, argue that occupational licensing gives rise to regulatory capture (Stigler 1971) and results in barriers to entry that disproportionately affect the poor and

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Joshua C. Hall is Associate Professor of Economics at West Virginia University and Shree B. Pokharel is a Graduate Research Assistant.

disadvantaged (Dorsey 1983, Bernstein 1994). Supporting the claim of regulatory capture is Kleiner (2000), who reports that more often than not members of licensing boards are chosen from the occupations being licensed.

The literature on occupational licensure has typically focused on the effects of licensure on wages and safety. A few articles focus on licensure as a barrier to entry, but those studies largely deal with high-skilled labor markets. Carpenter and Stephenson (2006), for instance, find that 150 hours of college course work necessary to sit for the CPA exam reduces the number of candidates sitting for the CPA exam by 60 percent.

In this article, we focus on occupational licensure as a barrier to entry for one relatively low-skilled occupation—barbering. The barbering profession was one among many professions to be licensed early in the United States, with Minnesota passing the first barber licensing law in 1897 (Thornton and Weintraub 1979). Alabama was the last state to license barbers in 2013 (Burkhalter 2014). Today all states and the District of Columbia regulate barbering. In 1976, barbering was heavily regulated with average education and experience requirements of 1,460 hours and a mean apprenticeship period of approximately 18 months (Thornton and Weintraub 1979). By 2012, average education and experience requirements were 890 days and average fee requirements were \$330 (Carpenter et al. 2012).

While many studies have focused on occupational regulation and economic outcome variables, such as changes in earnings and employment (Kleiner 2000) and migration (Mulholland and Young 2016), few studies have examined the impact of occupational regulatory burdens on low-income professions such as barbering. Some previous studies have estimated the relationship between regulatory burdens and the supply of barbers (Fuchs and Wilburn 1967, Maurizi 1974, Thornton and Weintraub 1979). Thornton and Weintraub (1979) find that average minimum grade level affects the supply of barbers. Timmons and Thornton (2010) find that state barber licensure has increased barber earnings by between 11 and 22 percent.

In this study, we estimate the relationship between the state-level regulatory burden on the practice of barbering and the number of barber shops in a state. Since many barber shops are one-or two-chair shops, restrictions on the profession of barbering are

restrictions on the number of barber shops. We hypothesize that states with higher regulatory burdens on becoming a barber should have fewer barber shops per capita. Utilizing the one year of regulatory data on barbering from Carpenter et al. (2012), we find that the number of exams required to become a barber in a state is negatively related to the number of barber shops per capita in that year. Conversely, we find that fees, minimum grade levels, and minimum age requirements do not explain state variation in the number of barber shops per capita.

Data

We use barber shops per 100,000 inhabitants for all 50 states and the District of Columbia in 2011 as our measure of entrepreneurial barber activity. Our data come from the U.S. Census Bureau's Nonemployer Statistics database, and we use the North American Industry Classification System (NAICS) code for barber shops (812111) to identify "establishments known as barber shops or men's hair stylist shops engaged in cutting, trimming, and styling boys' and men's hair; and/or shaving and trimming men's beards." State population in 2011 was obtained from the U.S. Census Bureau. The dependent variable is the authors' calculation with scores ranging from 14.0 (Utah) to 92.8 (Alabama) for each state. The score is calculated by dividing each state's total barber shop establishments by the state population. For example, Alabama's score means that, on average, there are approximately 93 barber shops for every 100,000 residents.

There are three categories of explanatory variables in this study which might affect the number of barber shops per capita: Measures of Occupational Regulation, State Controls, and Attributes of Entrepreneurs. Our major variables of interest fall in the Measures of Occupational Regulation category and consist of variables representing governmental burdens imposed by state governments on prospective barbers. The variables included in this category are average number of exams, average fees, average minimum grade level, and average minimum age imposed by states on barbers to acquire a license. These variables are reported from *License to Work: A National Study of Burdens from Occupational Licensing*. While Carpenter et al. (2012) provide regulatory information for 102 occupations in which the average income is below the national

average, we use only their measures of occupational licensure for barbering.¹

All variables in this category are reported in their original form. *Fees* are in dollars and represent the payments necessary to achieve an initial license. Continuing education fees and renewal fees are not included. *Number of Exams* represent the number of written and practical exams required in a state to get a license. *Minimum Grade* is the minimum education level necessary to apply for a license. States without a minimum grade level receive a 0; states with an eighth-grade minimum receive an 8, with a high school minimum a 12, and so on. For barber licensure, no state requires more than a 12th grade education. *Minimum Age* is the minimum age an individual in the state must be to apply for barber licensure and varies across states from 0 to 18. Many states, such as Iowa, have both a minimum grade and age requirement.

In addition to regulatory burdens, state-specific variables related to the economic or social environment might also influence the decision to become a barber and open a barber shop. We primarily draw on the entrepreneurship literature as the motivation for these controls, which are all measured for 2011. For example, the *Unemployment Rate* is found to negatively affect self-employment across OECD countries (Blanchflower 2000). At the level of U.S. states, however, the results are mixed. Unemployment is found to have an insignificant relationship with new business starts (Carree 2002) and a negative relationship with latent entrepreneurship (Gohmann 2012). However, Gohmann and Fernandez (2014) find that unemployment Granger-causes proprietorships. In addition, Coomes, Fernandez, and Gohmann (2013) find that the unemployment rate is positively related to proprietorships at the MSA level. *Unemployment Rate* is obtained from the Bureau of Labor Statistics. The role of median income in influencing entrepreneurship is unclear (Yago, Barth, and Zeidman 2007), but is generally thought to positively affect the number of new businesses as individuals seek greater diversity in consumption. *Median Household*

¹Carpenter et al. (2012) also provide a measure of the days of education and experience necessary to achieve a license. Doing so requires a number of assumptions, however, and we prefer to focus on the directly comparable features of barber regulation listed such as fees and number of exams. Inclusion of the number of days of education and experience does not qualitatively affect our empirical results.

Income for each state and the District of Columbia was obtained from the U.S. Census Bureau. Crime has been found to negatively affect entrepreneurship (Rosenthal and Ross 2010), and we use *Property Crimes* from the FBI's *Uniform Crime Reports*.

In addition to regulatory burden variables and state control variables that affect entrepreneurship, we also include demographic controls to capture the attributes of those most likely to start a business. The variables in this category include percentage of labor force that is *Male*, percentage of labor force that is *White*, and *Median Age* of state residents. All demographic data are for 2011 and were obtained from the Bureau of Labor Statistics and the U.S. Census Bureau. Kreft and Sobel (2005) and Hall and Sobel (2008) find that the percentage of labor force that is male and white, as well as the median age within the state, affect entrepreneurship. Similarly, Langowitz and Minniti (2007) find that the probability of men being entrepreneurs is higher than women. Table 1 presents summary statistics for all variables employed in our article.

The dependent variable, *Barber Shops per 100,000 Residents*, has a mean of 38.80. This means that, on average, there were

TABLE 1
SUMMARY STATISTICS

Variable	Mean	St. Dev.	Min.	Max.
Barber Shops per 100,000 Residents	38.80	20.87	14.0	92.8
Number of Exams	2.20	0.69	0.0	4.0
Fees	127.63	70.25	0.0	330.0
Minimum Grade	6.78	5.21	0.0	12.0
Minimum Age	13.31	7.08	0.0	18.0
White	79.73	13.33	26.03	95.48
Male	49.33	0.79	47.3	51.9
Income	50,686.47	7,475.12	39,856.0	68,876.0
Unemployment Rate	8.17	1.93	3.5	13.1
Property Crimes	2,863.06	670.07	1,395.2	4,795.5
Median Age	37.73	2.38	29.9	43.5

NOTES: N = 51 (all U.S. states and the District of Columbia). For sources of data, see discussion in text.

approximately 39 barber shops per 100,000 residents per state throughout the United States in 2011. There is significant variation across states, however, in the number of barber shops. While Utah had only 14 barber shops per 100,000 residents, Alabama had approximately 93 barber shops per 100,000 of its residents.

There is a lot of variation in terms of *Number of Exams*. Alabama requires no exams while Minnesota and Nevada require four exams. *Fees* varied considerably as well from \$0 in the District of Columbia to \$330 in Kentucky. *Minimum Grade* requirements also varied across states with many requiring no educational attainment level while others specifically require at least a high school or equivalent degree. In terms of *Minimum Age*, many states do not have a minimum age requirement to be a barber while in other states one has to be at least 18 years old.

Unemployment rates vary across states as well. North Dakota had the lowest unemployment rate of 3.5 percent while Nevada had the highest unemployment rate of 13.1 percent. Property crime rates across states also vary notably. Rhode Island had the lowest property crime rate with approximately 1,395 crimes per 100,000 inhabitants, while Washington, D.C., had the most property crimes per 100,000 inhabitants. There is almost a 1.5 times difference between the state with the lowest median income and the state with the highest median income; Kentucky has median income of \$39,856 while Maryland has a median income of \$68,876. Percent of male population in a state also varies with 47.3 percent in District of Columbia to 51.9 percent in Alaska. There is a significant difference within states in terms of racial composition as well. While approximately 26 percent of Hawaii's population is white, 95.48 percent of Vermont's population is white.

Empirical Approach and Results

Since we are limited in terms of numbers of observations in our data set, we employ a simple linear OLS regression model for our empirical analysis. Our model is represented as follows:

$$\text{BARBSHOPS} = \beta_0 + \beta \text{REGULATION} + \gamma \text{STATE} \\ + \delta \text{ENTREPRENEUR} + \epsilon$$

where β , γ , and δ are row vectors and *REGULATION*, *STATE*, and *ENTREPRENEUR* are column vectors. *BARBSHOPS* represents

total barber shops per capita in U.S. states. As mentioned in the previous section, *REGULATION* represents barber-specific regulatory variables; it consists of *Number of Exams*, *Fees*, *Minimum Grade*, and *Minimum Age*. *STATE* represents state controls and therefore includes *Unemployment Rate*, *Property Crimes*, and *Income*. Attributes of entrepreneurs are represented by *ENTREPRENEUR* and consist of *Male*, *White*, and *Median Age*.

Table 2 shows the effect of state-level barber regulations on the total number of barber shops per capita in 2011. Specification 1 represents a parsimonious specification containing only the primary variables of interest. While this specification does not explain the full effect of the explanatory variables on the dependent variable, it helps to outline the basic relationship between them. The signs of *Minimum Grade*, *Fees*, and *Number of Exams* are as expected, with *Number of Exams* statistically significant at the 1 percent level. The sign on *Minimum Age* is positive although not statistically significant.

In Specification 2, we add basic entrepreneur characteristics controls standard in the literature. We find that *Number of Exams* is still significant at the 1 percent level. *Fees* continue to be negatively associated with the dependent variable, but is still statistically insignificant. *Male* exhibits a strong negative relationship at the 1 percent level on the number of barber shops. *White* also is negatively related to the number of barber shops at the 5 percent level. However, the signs for *Male* and *White* exhibit opposite signs than what was previously found for other measures of entrepreneurship (Kreft and Sobel 2005, Hall and Sobel 2008).

Specification 3 adds basic state controls standard in the literature—median household income and the unemployment rate. The key result is that *Number of Exams* continues to exhibit a significant negative effect on the level of barber shops in a state at the 1 percent level. *Income* leads to fewer barber shops per capita, although the economic magnitude is small. The sign on *Unemployment Rate* is positive but statistically insignificant (Blanchflower 2000).

Finally, in Specification 4, we add *Property Crimes* and *Median Age*. *Number of Exams* continues to be negatively related to the number of barbershops per capita at the 1 percent level. *White*, *Male*, and *Income* are statistically significant as well as *Median Age*. *Property Crimes* are not significant. This full specification explains 71 percent of the variation in barber shops per capita in 2011 across U.S. states.

TABLE 2
STATE-LEVEL BARBER REGULATIONS AND NUMBER OF
BARBER SHOPS PER CAPITA

Variable	(1)	(2)	(3)	(4)
Number of Exams	-13.195*** (3.97)	-9.079*** (3.14)	-8.811*** (2.76)	-8.141*** (2.74)
Fees	-0.025 (0.04)	-0.003 (0.03)	-0.010 (0.03)	-0.020 (0.03)
Minimum Grade	-0.431 (0.56)	0.308 (0.48)	0.308 (0.43)	0.408 (0.44)
Minimum Age	0.622 (0.41)	0.317 (0.32)	-0.093 (0.30)	-0.078 (0.30)
White		-0.414** (0.19)	-0.394** (0.17)	-0.387** (0.18)
Male		-13.243*** (2.82)	-11.255*** (2.58)	-13.914*** (2.90)
Income			-0.001*** (0.00)	-0.001*** (0.00)
Unemployment Rate			1.834 (1.12)	1.452 (1.14)
Property Crimes				-0.004 (0.00)
Median Age				-1.713* (0.96)
R-squared	0.23	0.57	0.68	0.71

NOTE: Dependent variable is the number of barber shops per 100,000 state residents.

N = 51 in all specifications. *, **, and *** denote statistical significance at the 10, 5, and 1% levels, respectively. Numbers in parentheses are absolute standard errors. Constant included but not reported.

Conclusion

Given the growth in occupational licensure and the importance of barriers to entry for low-income workers, we analyzed the effect of barber licensure on the number of barber shops across U.S. states. We find that the number of required exams is robustly associated in

a negative way with the number of barber shops per capita in a state. However, we find that other restrictions such as age requirements and fees have no consistent relationship with the number of barber shops. This might be the result of our limited data set. We feel that this exploratory look at the issue of barber licensure opens up future research in this area, especially research that can establish more of a causal link.

Further research could focus on the origins of these laws, especially since historically many of these laws have their roots in discrimination. Bernstein (1994), for example, details how licensing laws have historically been used to reduce the number of African-Americans in certain occupations such as barbering. As Kuznicki (2009) points out, government power exercised through things like occupational licensure is never neutral when it comes to race. Our results also say nothing about the efficacy of the restrictions in terms of the quality of haircuts received in states with more stringent regulations. Carpenter (2012) is a good example of the type of applied research that could be done in this area, as he finds no difference between licensed and unlicensed florists. Finally, it would be fruitful to further investigate the vast differences across states in the amount of regulation of certain industries, such as barbering, from a political economy perspective, in order to better understand the various special interests at play.

References

- Akerlof, G. (1970) "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.
- Arrow, K. J. (1971) The Theory of Risk Aversion. In K. J. Arrow (ed.), *Essays in the Theory of Risk-Bearing*, 90–120. Chicago: Markham.
- Bernstein, D. E. (1994) "Licensing Laws: A Historical Example of the Use of Government Regulatory Power against African Americans." *San Diego Law Review* 31 (1): 89–104.
- Blanchflower, D. G. (2000) "Self-Employment in OECD Countries." *Labour Economics* 7 (5): 471–505.
- Burkhalter, E. (2014) "Barbers Facing New State Regulations." *The Anniston Star* (February 5). Available at www.annistonstar.com/news/barbers-facing-new-state-regulations/article_7ccec4f5-832a-5848-acb1-4695d14ad821.html.

- Carpenter, C. G., and Stephenson, E. F. (2006) “The 150-Hour Rule as a Barrier to Entering Public Accountancy.” *Journal of Labor Research* 27 (1): 115–26.
- Carpenter, D. M. (2012) “Testing the Utility of Licensing: Evidence from a Field Experiment on Occupational Regulation.” *Journal of Applied Business and Economics* 13 (2): 28–41.
- Carpenter, D. M.; Knepper, L.; Erickson, A. C.; and Ross, J. K. (2012) *License to Work: A National Study of Burdens from Occupational Licensing*. Washington: Institute for Justice.
- _____ (2015) “Regulating Work: Measuring the Scope and Burden of Occupational Licensure among Low-and Moderate-Income Occupations in the United States.” *Economic Affairs* 35 (1): 3–20.
- Carree, M. A. (2002) “Does Unemployment Affect the Number of Establishments? A Regional Analysis for U.S. States.” *Regional Studies* 36 (4): 389–98.
- Coomes, P. A.; Fernandez, J.; and Gohmann, S. F. (2013) “The Rate of Proprietorship among Metropolitan Areas: The Impact of the Local Economic Environment and Capital Resources.” *Entrepreneurship Theory and Practice* 37 (4): 745–70.
- Dorsey, S. (1983) “Occupational Licensing and Minorities.” *Law and Human Behavior* 7 (2–3): 171–81.
- Fuchs, V. R., and Wilburn, J. A., eds. (1967). *Productivity Differences within the Service Sector*. New York: Columbia University Press.
- Gohmann, S. F. (2012) “Institutions, Latent Entrepreneurship, and Self-Employment: An International Comparison.” *Entrepreneurship Theory and Practice* 36 (2): 295–321.
- Gohmann, S. F., and Fernandez, J. M. (2014) “Proprietorship and Unemployment in the United States.” *Journal of Business Venturing* 29 (2): 289–309.
- Hall, J. C., and Sobel, R. S. (2008) “Institutions, Entrepreneurship, and Regional Differences in Economic Growth.” *Southern Journal of Entrepreneurship* 1 (1): 69–96.
- Kleiner, M. M. (2000) “Occupational Licensing.” *Journal of Economic Perspectives* 14 (4): 189–202.
- _____ (2014) “Life, Limbs, and Licensing: Occupational Regulation, Wages, and Workplace Safety of Electricians, 1992–2007.” *Monthly Labor Review* 137: 1–31.

- Kleiner, M. M., and Krueger, A. B. (2013). "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." *Journal of Labor Economics* 31 (2): 173–202.
- Kreft, S. F., and Sobel, R. S. (2005) "Public Policy, Entrepreneurship, and Economic Freedom." *Cato Journal* 25 (3): 595–616.
- Kuznicki, J. (2009) "Never Neutral State: American Race Relations and Government Power." *Cato Journal* 29 (3): 417–54.
- Langowitz, N., and Minniti, M. (2007) "The Entrepreneurial Propensity of Women." *Entrepreneurship Theory and Practice* 31 (3): 341–64.
- Maurizi, A. (1974) "Occupational Licensing and the Public Interest." *Journal of Political Economy* 82 (2): 399–413.
- Mulholland, S. E., and Young, A. T. (2016) "Occupational Licensing and Interstate Migration." *Cato Journal* 36 (1): 17–31.
- Rosenthal, S. S., and Ross, A. (2010) "Violent Crime, Entrepreneurship, and Cities." *Journal of Urban Economics* 67 (1): 135–49.
- Shapiro, C. (1986) "Investment, Moral Hazard, and Occupational Licensing." *Review of Economic Studies* 53 (5): 843–62.
- Stigler, G. J. (1971) "The Theory of Economic Regulation." *Bell Journal of Economics and Management Science* 2 (1): 3–21.
- Thornton, R. J., and Weintraub, A. R. (1979) "Licensing in the Barbering Profession." *Industrial and Labor Relations Review* 32 (2): 242–49.
- Timmons, E. J., and Thornton, R. J. (2010) "The Licensing of Barbers in the USA." *British Journal of Industrial Relations* 48 (4): 740–57.
- Yago, G.; Barth, J. R.; and Zeidman, B. (2007) *Entrepreneurship in Emerging Domestic Markets: Barriers and Innovation*. Springer: New York.

RENT SEEKING AND ENTREPRENEURSHIP: INTERNET STARTUPS IN CHINA

Tain-Jy Chen and Ying-Hua Ku

Against all odds, China has developed one of the most vibrant Internet industries in the world. According to Atomico (2015), which tracked venture capital (VC)-funded startups in the world, there were 156 Internet startups that had been founded in 2003–14 and that had become billion-dollar companies (based on market capitalization) by the end of 2014 after initial public offerings (IPOs). The United States leads the list with 86 companies, followed by China’s 30, and Sweden’s 5. All Chinese billion-dollar startups are consumer-related, while billion-dollar startups in other countries include business applications, games, and others. Similarly, the *Wall Street Journal* tracked unlisted VC-funded startups and identified 78 of them whose market valuation (measured by financing terms in the most recent round of funding) had exceeded one billion dollars in March 2015 (Table 1). The list includes startups in the Internet as well as other sectors. Again, the United States leads the list with 50 ventures, followed by China’s 8. All Chinese ventures are Internet-related, if Xiaomi, which tops the list of all startups and sells smartphones on the Internet, is also counted as an Internet company (Dow Jones Venture Source 2015). In short, Chinese startups are numerous, vigorous, and most successful in Internet-based consumer business.

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Tain-Jy Chen is Professor of Economics at National Taiwan University. Ying-Hua Ku is an economist with the Chung-Hua Institution for Economic Research in Taipei.

TABLE 1
CHINESE STARTUPS VALUED AT \$1 BILLION OR MORE,
MARCH 2015

Company	Industry	Last Valuation (\$ billion)	Total Equity Funding (\$ billion)
Xiaomi	Smartphone	46.0	1.4
Meituan	Group purchase	7.0	1.1
Didi Dache	Taxi app.	3.5	0.828
VANCL	Apparel	3.0	0.512
Kuaidi Dache	Taxi app.	2.8	0.88
Dianping	Restaurant app.	2.0	0.569
Koudai Shopping	Mobile shopping	1.5	0.364
Mogujie	Apparel	1.0	0.2

SOURCE: Dow Jones Venture Source (2015).

Startups are manifestations of entrepreneurship and innovation. Vibrant startups indicate that the Chinese business environment is conducive to entrepreneurial and innovative activities. This is at odds with the general impression that China is nowhere near a business-friendly country. The Chinese institutions are considered inadequate or hostile to entrepreneurial activities. For example, the costs of starting a business in China are high, protection of property rights is inadequate, contract enforcement is lax, and financial market development is immature. Moreover, China ranks as one of the worst among the emerging countries in terms of corruption (La Porta et al. 2004), which fuels rent-seeking activities. It has been demonstrated in the literature that rent seeking undermines entrepreneurship (Baumol 1990; Murphy, Shleifer and Vishny 1990).

If rent seeking is prevalent in China, as suggested by the high-profile anticorruption campaigns in recent times, why are innovative activities so vibrant in the Internet industry? We argue in this article that rent seeking in the real sector actually encourages innovations in the Internet sector to uncover the hidden opportunities unrealizable in the real sector. Regulations designed by rent seekers always create distortions in the market, from which extra profits are generated. In China, such distortions are often reflected in above-normal prices, which benefit producers, especially large producers, while

depressing consumption. A stylized Chinese startup in the Internet industry creates a new business model that offers goods or services to satisfy unfulfilled consumer demand in the real sector. Low prices are a common feature of their business models. In China, Internet trade is analogous to the underground economy in other countries where small traders escape regulations. Despite the siphoning of talent and resources into rent-seeking activities, the fact that China is an open economy allows the local startups to tap entrepreneurship from international sources to undertake innovative activities. The startups also leverage international institutions to protect the value of their innovations. The sheer size of the Chinese market heightens the value of innovations, which in turn offsets the high risk of startup ventures.

Scholars often asked why private business activities remain robust in China despite weak institutions. The conventional explanation for this paradox is that informal institutions supplant the formal ones. For example, informal financial instruments make up for the weaknesses of the formal financial institutions in allocating financial resources (Allen, Qian and Qian 2005), and personal relations and informal contracts make up for the weak institutions in contract enforcement (Yu and Zhang 2008; Kwock, James and Tsui 2013). Internet startups are not a part of this story, however. The vitality of the Chinese Internet startups is not a manifestation of the working of informal institutions. Instead, the Chinese Internet industry functions under a set of private rules installed by the platform operators, who work as national champions to protect the domestic market from foreign penetration. The bureaucrats support and endorse such rules, which govern Internet trade more effectively than public regulations. More importantly, these rules reinforce rather than undermine the power of the state. The platform operators also invest in public goods that underpin the growth of the Internet industry. The Chinese Internet startups thrive in spite of weak institutions because they work under an umbrella provided by the platform operators. With a commitment to protect state interests, the platform operators are able to keep the government's hands off the market and therefore create a safe haven for startups. The umbrella contains a set of private rules that, while safeguarding the state interests, promote entrepreneurship and competition to enrich the platforms.

The remainder of this article is organized as follows. First, a theoretical background is provided for the relationship between the

state and entrepreneurship. Second, we discuss the unique combination of weak institutions and strong organizations that underlie the development of China's Internet industry. Next, we present a case study of Chinese Internet startups in taxi app, retail, and O2O (online to offline) businesses, highlighting the Internet economy as a complement to real sector deficiencies. Finally, we explain how the Internet platform operators create a favorable business environment for the Internet startups while compromising with the state interests.

The State and Entrepreneurship

It is well documented in the literature that the state has an important role to play in promoting entrepreneurial activities. The state has to establish and maintain good institutions that reduce transaction costs and risks in business undertakings. For example, institutions that protect property rights and enforce contracts are considered critical for the development of a market economy in which private enterprises play a major role in resource allocation (North 1990). While establishing and maintaining good institutions is important for the functioning of the markets, the state has to refrain from intervening in the markets if private enterprises are to thrive. The inclination of the state to intervene encourages rent-seeking activities that undermine private entrepreneurship. Rent-seeking activities are likely to be rampant when the state is potent but prone to the influence of private interests due to weak institutional constraints (Lin 2009). Rent seeking occurs not because of a lack of laws and regulations, but because laws and regulations are not institutionalized and can be applied in a discretionary way by the authorities (Schneider 2002).

Baumol (1990) has demonstrated that entrepreneurship can be directed to productive or unproductive activities. When more entrepreneurship is directed toward unproductive activities, such as rent seeking, less entrepreneurship will be available for productive activities, such as starting new businesses. Whether entrepreneurship is directed toward productive or unproductive activities depends on the reward system in the society. Widespread corruption indicates that rent seeking yields good returns in a society. In the same vein, Murphy, Shleifer, and Vishny (1990) have argued that when more of a nation's talent is allocated to distributive functions, such as rent-seeking-oriented

regulations, less talent will be devoted to creative functions, thus impeding the nation's growth potential.

According to the above theoretical assertions, we should not expect private entrepreneurship to thrive in China given its weak institutions. For example, institutions for property rights protection and investor protection are apparently inadequate (Allen, Qian, and Qian 2005). Although privately owned business firms have already been recognized by the PRC Constitution, state expropriation of privately owned firms remains a real risk (Hsia 2007). Market entry is strictly regulated in certain industries to protect established firms (Djankov et al. 2002). Contract enforcement remains difficult in general (Kwock, James, and Tsui 2013). The weaknesses in property rights protection and contract enforcement not only discourage risk-taking business ventures, but also prompt the investors to seek political connections (Li et al. 2008), which nurture rent-seeking behavior.

Economic growth is still possible under weak institutions if a strong state has sufficient capacity to chart the course of economic growth by coordinating private actions (Wade 1990). However, state-led growth is normally not conducive to private entrepreneurship, although the bureaucrats themselves may act like entrepreneurs as in the case of China (Oi 1995, Chen 2016). Baumol, Litan, and Shramm (2007) classified capitalism into four kinds: state-led, oligarchic, big firm-led, and entrepreneurial. The taxonomy depends on who dictates the resource allocation in the economy. China could easily fall into the category of state-led capitalism, in which the state rather than private entrepreneurs dictates the allocation of resources. There are observations that, in recent years, state and state-controlled enterprises have advanced their power in allocating resources at the expense of private firms (*guo jin min tui*) (Adams 2011). Huang (2008) has demonstrated that in the post-1978 Chinese economic development course, when the state increased its power in resource allocation, private entrepreneurship receded. With a similar argument, Wu and Huang (2008) portrayed a see-saw game between rent-seeking and innovative activities in the process of China's economic reforms. When the market was given more power in allocating resources, innovative activities were promoted and rent-seeking activities suppressed. Conversely, when the government, or its proxy, was given more power in allocating resources, rent seeking was fueled

at the expense of innovative activities. The boom in Internet startups in recent years appears to be a counter-example of this line of argument, and thus explanations are called for.

Weak Institutions and Strong Organizations

We argue that Internet startups thrive in China because of weak institutions coupled with strong organizations. Weak institutions allow strong organizations in the Internet sector to establish their own rules that encourage private entrepreneurship but also protect state interests, especially the interests of local government in the real sector. A delicate balance is struck between the state-led real sector and a market-led Internet sector. The balance is possible because the Chinese economy is half-open and half-closed. A half-open economy in the real sector allows Internet startups to leverage international institutions for development. International institutions enable access to foreign resources including financial, technological, and entrepreneurial resources. International institutions even extend property rights protection to local startups. A half-closed economy in the Internet sector creates national monopolies that provide a platform for the startups. The national monopolies act as agents of the state in regulating the Internet industry, but their profits have to come from the innovative activities of the startups that exploit the deficiencies in the real sector that is prone to rent seeking. The Internet sector is a shadow of the real economy. Rent seeking in the real economy creates distortions and hidden opportunities that nourish innovations in the Internet sector. This explains why rent seeking in the real sector promotes Internet startups. The platform operators have a strong incentive to maintain the state's power of rent seeking, because by doing so they also preserve the roots of profits. More distortions in the real sector imply more business opportunities in the Internet sector. Regulations originating from rent seeking often favor large producers at the expense of consumers (Stigler 1971). In response to this, the platform operators create an Internet environment that serves small producers and consumers.

In the following, we shall discuss some important features of Chinese (weak) institutions that significantly affect the behavior of business firms. These institutions fuel rent seeking and impede entrepreneurial activities in the real sector, but the institutional weaknesses are overcome in the Internet sector with the structures provided by

the platform operators and others. As a result, the Internet sector becomes a complement to the real sector deficiencies.

Foreign Investment Regime

China has offered favorable conditions to foreign direct investment since the early 1990s. Foreign companies bring new products and technologies to China, which may inspire indigenous innovations. However, favorable conditions afforded to foreign-invested companies may have crowded out the business opportunities of domestic enterprises (Huang 2008). Contrary to the favorable policy toward foreign investment in the real sector, multinational firms are almost completely blocked out in the Internet sector due to the concerns of national security. Aghion and Griffith (2005) have argued that trade protection may encourage rather than impede innovation if it enhances the potential value of innovation. The Chinese experience indicates that trade protection of the Internet industry has indeed encouraged innovations because innovation is the only way to enter the Internet markets, which never existed before. However, the innovation may have originated from imitating foreign products or business models. The imitation of foreign business models is especially visible in the Chinese Internet sector. Copy to China (a homonym of C2C) with modifications to fit the Chinese trading habits and business environment appears to be a quick success formula. Trade protection has produced a few monopolies in the Chinese Internet industry, all of which are privately owned. National monopolies impose self-regulation in order to prevent government interventions. Whereas the primary interest of the central government in the Internet sector is political—namely, maintaining the order of trade and controlling public opinion—the primary interest of the local government is economic. The platform operators impose private regulations to serve the central government interests on the one hand, and adopt business models that protect the local government interests on the other. By doing so, they create a free space for startups that fuels the rapid growth of the Internet sector, thereby justifying their legitimacy as national champions (Nolan 2001).

Insecure Property Rights

The risk of state predation on private properties remains imminent in China. For example, the cases of private businesses being

nationalized continue to exist. Even publicly listed companies are not completely immune from the risk of nationalization.¹ Building political connections is thus a popular way for business enterprises to hedge against state expropriation (Kung and Ma 2014, Duan and Chik 2012). Since profitable and growing companies are subject to higher risks of state predation than nonprofitable and stagnant ones, because the former command more rents to be extracted by the state, a good strategy for the owner-entrepreneur of a startup is to build a business venture as quickly as possible, making it profitable (or seemingly profitable) and getting it listed on the stock exchange or acquired by established companies. Upon the IPO or acquisition, the entrepreneur relinquishes (or dilutes) the ownership and realizes the entrepreneurial gains. In other words, insecure property rights prompt the startup founders to create business ventures without the intention of owning or managing them on a long-term basis. This strategy fits perfectly with the VC model, and it works especially well when foreign-based venture capital is involved as the latter facilitates overseas IPOs. State predation is unlikely to occur before an IPO. In addition to avoiding state expropriation, the Chinese Internet ecosystem works well to protect the property rights of Internet innovators as the value of an innovation can only be realized through the platforms owned by the national monopolies. Founders of the startups are entrepreneurs who sense the business opportunities and are willing to take action to realize such opportunities (Kirzner 1979). Platform operators serve as gatekeepers of their property rights. Platform operators also operate their own VC funds and angel funds that can be used to invest in and acquire promising startups.

Market Entry Barriers

It is not easy to start a private business in China. The World Bank (2015) ranked China 128th out of 189 countries surveyed in terms of the ease of starting a business. The ranking is measured by the number of days needed to complete a business registration. In fact, in

¹For example, two listed semiconductor companies on the NASDAQ, Spreadtrum and RDK, were recently acquired by a state-backed fund and delisted to become companies under state control.

addition to business registration, various licenses may be required before a business can be operated legally. As an indication of the difficulty in obtaining a license, the World Bank ranked China, in a separate category of the same survey, 179th out of 189 countries in terms of the time needed to obtain a construction permit. Most licensing controls are in the hands of local governments that typically prefer large businesses over small ones. Local governments would rather provide incentives to attract subsidiaries of large, established companies, whether domestic or foreign, instead of helping startups or small and medium-sized enterprises (SMEs), because the former are more effective in terms of quickly increasing the production base in their jurisdictions. They expend more resources on attracting investments from outside the regions (including foreign countries) than on incubating startups within their territories. Therefore, no help is provided in starting a private business in China, especially in rich cities that are resourceful in subsidizing outside investors. In contrast to the hassles in the real sector, starting an Internet business avoids all the troubles associated with licensing controls. All that is needed is an Internet business registration, which can be expedited through the platform operators. A business registration can even be avoided in the case of C2C trade. Once the company starts selling its products or services on the Internet, the platform operators provide an umbrella of business licenses. The platform operators welcome new products and services to enhance the value of the platforms. Most importantly, there is no need for Internet startups to acquire a piece of land, which is owned by the government and is the bastion of rent seeking. While an Internet business is a landless operation, its success may contribute to the demand for land, such as transportation and logistics operations derived from online trade, to benefit the land owners, namely, the state. In fact, the booming Internet startups in recent years have created a demand for incubation bases in urban areas, giving the local government a new outlet for land development.

Land-Based Fiscal Policy

Land is owned and disposed of by the local government in China, and land has become an important source of fiscal revenue for local governments since the tax reform in 1994 (Yang 2012). Local governments allocate land usage in a way that the land's value is maximized. For example, local governments prefer land users that build high-rise

office buildings and luxurious shopping malls rather than road-side stores. Even in allocating land use to road-side stores, local governments prefer renowned international brands over little-known local brands (Huang 2008). Local governments control the number of department stores or hypermarts in order to create monopoly rents that can then be captured by land value. This practice makes the entry charges for retail channels extremely high in Chinese cities. In other countries, this would give rise to booming road-side vendors or other forms of underground economy. However, Chinese city governments often deploy semi-police forces, or *cheng-guan*, to eliminate road-side vending activities. Against this backdrop, the Internet provides a sanctuary for small brands or counterfeits in China that would have presented themselves in underground trade in other countries. Local governments are happy to see the would-be road-side vendors retreat from the streets instead of competing with their patrons. Better yet, they can be regulated through the platform operators. In essence, the Chinese Internet economy is a special zone that keeps the underground economy at bay. Private regulations are imposed by platform operators to harness Internet trade just like the informal institutions that regulate the underground economy in other countries. The only difference is that these private regulations are subject to the ultimate control of the state and they always serve state interests, whereas informal institutions are out of the hands of the government. Reconciliation of business interests with state interests through private regulations is pivotal to the stability of the Chinese Internet industry.

Local Protectionism

Local protectionism has its origins in the dual-track price system and fiscal decentralization in the early stage of the economic reforms in China (Wedeman 2003). Despite the fact that the dual-price system has now been eliminated and that fiscal power is more centralized than in the past, local protectionism lingers on. Local protectionism is likely to occur in the industries that generate good tax revenues or business earnings for local governments, and in regions where state ownership is high (Bai et al. 2004). Local protectionism produces a segmented market, which limits industrial specialization and precludes the benefits of scale economies of a single market (Young 2000). It also produces local monopolies whose interests are aligned with local government interests. While

the central government does not like local protectionism, local government officials continue to embrace it because their prospects for promotion are linked to local economic performance (Li and Zhou 2005). Internet trade is a natural enemy of local protectionism as it spans regional boundaries and exploits the price differences between regions. The platform operators have to unite with the central government to deter interventions by local governments. They also have to refrain from entering businesses in which vested local interests lie and, if possible, they will form alliances with local monopolies. This strategy has worked well. In May 2015, the State Council of the central government issued an ordinance (no. 24), “Promoting E-Commerce as a New Economic Force” (*guanyu dali fazhan dianzi shangwu jiaokuai peiyu jingji xin dongle de tongzhi*), instructing local governments to reduce regulatory controls on Internet trade. The ordinance advises the local governments to issue business licenses to Internet stores before any field-specific approvals are to be processed (*xian zhao hou zheng*).

Case Studies

In this section, we will study the cases of Chinese Internet startups that are listed as Dow Jones Venture Source’s 2015 one-billion dollar startups to elucidate the following propositions: (1) Internet innovations are inspired by the opportunities arising from the distortions caused by rent seeking in the real sector, and (2) rules imposed by the platform operators and business models followed by the Internet startups reinforce rather than weaken the power of the state. The cases will be grouped into three categories: taxi app, Internet retailing, and O2O services.

Taxi App

As urban populations have grown, the demand for taxi services has increased. In most Chinese cities, taxicabs are operated by a limited number of licensed companies, many of which are owned by the city government or state enterprises. In many large cities, taxi licenses are frozen despite the population growth because of concerns over traffic congestion. Normally, since individual licenses are not granted, a taxi driver is required to be affiliated with a taxi company to become a qualified taxi operator. The taxi company leases the taxicabs to

drivers and the leasing fee goes by month or day. Because of controls over licensing, the leasing fees are high, and taxi drivers have to use the cabs as efficiently as they can to make a living. They consciously avoid the destinations that will result in empty cars on the return trip or will force them to enter the congested traffic zones. The drivers are particularly choosy during the rush hour or when they are approaching the deadline of their shift. Difficulty in obtaining taxi services is a common headache for travelers in Chinese cities. By controlling taxi licenses, the local government benefits directly from the inflated leasing fees charged by the state-owned taxi companies or indirectly from the rent-seeking activities of private companies. Local governments can also use the purchasing power of the taxi fleet as an industrial policy tool to support the local automobile manufacturers. In response to the taxicab shortages, unlicensed taxicabs (black cabs) have emerged in Chinese cities, particularly during the rush hour. Against this backdrop, many taxi apps started to emerge after 2012, probably inspired by the successful model of Uber in the United States. Among them, Didi Dache and Kuaidi Dache are the largest ones.

Founded in 2012, Didi Dache is an Internet startup providing a mobile app which matches taxicabs with the riders. When a rider requests a taxi service on the mobile phone, cab drivers within a certain distance of the call will receive the request signal and they can offer to deliver the service simply by responding to the request, and a match is completed. If no response is made by any cab driver, say, because the trip is too short to attract a deal, the rider can offer a premium on top of the meter price. The system operator may also jump in to offer a subsidy if the riders have waited a long time. Because of the price flexibility, the Didi Dache app has claimed a matching rate exceeding 90 percent. Tencent, one of the largest Internet platforms in China, soon found the potential of the app and invested in the startup venture. Other VCs, both domestic and foreign, joined the investment in later rounds of fund raising. In an attempt to build up its user network, Didi Dache offered subsidies to both drivers and riders in matched transactions, and provided a bonus credit for new users. Tencent also applied its mobile payment facilities to enable online payments to be made for the taxi ride. Online settlement saves the trouble of cash transactions and makes subsidies (or commission charges) easy to execute.

The business model of Didi Dache was soon imitated by Kuaidi Dache, which was founded one year later in 2013 and subsequently

invested in by Alibaba, the archrival of Tencent. Didi and Kuaidi then engaged in a subsidy war and eventually drove out all the other competitors. According to Didi's own account, by the end of 2014, it had more than 100 million registered riders, and more than one million registered taxicabs, covering more than 300 cities, with daily matches exceeding 5.2 million trips (Wu and Piao 2014). Kuaidi was about the same size. Despite the huge numbers of members, both startups are still losing money today. Nevertheless, both of them have already attracted several rounds of investment from VC funds, with a total market value exceeding one billion U.S. dollars each.

The app services offered by Didi and Kuaidi have increased the revenues of taxi drivers without hurting the profit base of the taxi companies.² The service is also welcomed by local governments whose interests are tied to local taxi companies, because the app, which is offered only to licensed drivers, narrows the operating room of black cabs. It also improves the safety of taxi riding as both drivers and riders are registered users and all rides are tracked throughout the trip. However, when Didi started to offer rental car services to corporate customers, many local governments opposed the move.

Although Didi was started by an individual entrepreneur who made the initial innovation or imitation, the strategy of the company was shaped by its major investor, Tencent. Its competition with Kuaidi is tantamount to a competition between two Internet platforms. The rules that the platform operators have established are often complementary to government regulations. For example, if a taxi driver fails to carry out the matched deal for the first time, the driver's account will be suspended for three days. A one-month suspension will be imposed for a second-time violation, and the account will be removed permanently for a third-time violation. These rules are imposed on the basis of a private contract. Complaints about the taxi services or riders are easy to convey on the mobile Internet, so malicious drivers or riders can be disciplined. The riders are allowed to rate the service quality of the taxi, which affects the subsidy offered to or commission levied on the driver. The Chinese Internet platform

² In a few cities where private cars have been allowed to offer taxi services by adopting the Didi Dache app, such as Shenzhen where Tencent is headquartered, the profits of taxi companies have been hurt and a regulation is being deliberated to protect the interests of taxi service providers.

operators are willing to impose self-regulation because they are conscious of the risk of government intervention. Didi is even collaborating with the city government of Shanghai to track down black cabs and clone (falsely licensed) cabs in the city.

Internet Retailing

Along with economic growth, modern shopping facilities, such as department stores, shopping malls, or chain stores, have mushroomed in the Chinese cities. Leasing land to commercial developers is an important source of fiscal revenue for local governments. Local governments prefer large corporatized retailers to small individual- or family-owned stores (Wang and Song 2008). The Chinese retail industry is fragmented and characterized by high entry barriers. Compared to developed countries, retail space in urban areas in China is in relatively short supply. Chin and Chow (2012) reported that retail space per capita in China was only 12.9 square feet, which was much lower than the 45.2 square feet in the United States or the 16.4 square feet in Japan despite the high population density in Chinese cities. A shortage of retail space and the concentration of modern shopping facilities in city centers translates into high rental costs that screen out small vendors. Even for street stores, local governments may prefer big brands to small ones as lessees of state-owned property in order to enhance the land value. Huang (2008) reported that the Shanghai city government, in an effort to inflate the land price, chased small local brands out of the prestigious Huaihai Street, the most popular shopping district in the city, in favor of foreign brands. Because owning shopping space is such a privilege, it is a typical practice in the Chinese retail industry for channel owners to only rent out space to brand vendors without purchasing the commodities. Channel owners typically charge multiple fees to vendors, including a store entry fee, shelf-display fee, promotion fee, advertisement fee, etc. New charges can be added to the list at any time (Zhen 2007: 166–67). The brand vendors bear the inventory costs and the marketing responsibilities. This structure makes it very difficult for small local brands to enter the modern shopping facilities, not to mention the startups. Unlike in other developing countries, road-side vending is not a viable alternative in China because of strict government regulations and interference.

Shopping online has provided an alternative to shopping in physical stores. China has developed one of the largest online retail

industries in the world. The volume of online shopping exceeded 10 trillion RMB in 2014, the largest in the world. Most Chinese online shoppers use the Internet as an alternative to physical stores instead of purchasing genuine online products that are unavailable offline, such as games or online services. McKinsey Global Institute (2013) reported that 61 percent of Chinese consumers who shop online treat the Internet as a substitute for physical store purchases. This suggests that the high costs in physical stores have fueled the growth of Chinese online business. Small brands or unbranded products see the Internet as an alternative marketing channel for which the entry charge is low and predictable. Therefore, unlike in Western countries where electronics products dominate Internet shopping, the top selling product on the Chinese Internet is clothing, which is normally more conveniently sold in physical stores as an experience product. The most successful Chinese online brand for apparel products, VANCL, allows consumers to try on the clothes upon delivery without an obligation to buy. Even after the purchase, the consumers can still return the products within 30 days without explanation. This arrangement essentially allows consumers to experience the products after delivery.

The leading e-commerce platforms, such as Alibaba's Taobao (primarily C2C) and Tmall (primarily B2C), provides virtually free entry for retail vendors. Alibaba makes money by providing delivery and payment services, in addition to advertisements. Alibaba has built one of the most efficient logistics networks across the nation. The logistics services in China are segmented because of locally erected trade barriers and heterogeneous local regulations. No logistics operator holds a nationwide service network except China Post, which is state owned and notoriously inefficient. Alibaba extensively made use of the storage and transportation facilities in different locations, instead of establishing its own facilities. It works with many logistics service providers, including China Post, to construct a nationwide network. The shoppers can choose their preferred logistics service providers when purchasing. Alibaba only combines the capabilities of different logistics companies in different locations and standardizes the service procedures. Alibaba has claimed that more than one million small logistics operators have worked for its parcel delivery service, which accounted for two-thirds of the nation's total (Inside 2013). According to a central government regulation, all parcel service providers have to obtain a license from China Post. With the rise

of online shopping, the parcel service industry has boomed throughout the country but has also remained segmented and locally embedded. Alibaba is an integrator of local service networks without intruding into the vested interests of local operators, many of whom are tied to the local government interests.

Partly owing to local protectionism, the Chinese retail industry is fragmented despite the proliferation of retail chains that transcend provincial borders. In 2013, the top 100 retail chains in China accounted for only 9 percent of retail sales (Statista 2015). Two top retailers, Suning and Gome, both of which have specialized in consumer electronics products, have built nationwide channels to break down local barriers. Four leading foreign retail chains, RT-Mart, Wal-Mart, Lotus, and Yum, have all specialized in grocery and food products (Table 2). Domestic retailers making the top-10 list are all affiliated with local states, except for Gome, which is privately owned. Online-shopping giants such as Alibaba, which dominates the Internet trade with a more than 50 percent market share, present little threat to retail chains owned by local states as Alibaba offers a platform primarily for small brands, while the state-owned retail channels host big-name products. For example, in the offline market for clothing, foreign brands such as Uniqlo, Zara, and H&M have relied on department stores and specialty stores for marketing,

TABLE 2
LEADING RETAIL CHAIN OPERATORS IN CHINA, 2013

Name	Sales (billion RMB)
Sunning Commerce Group	138.0
Gome Electrical Appliances	133.3
China Resource Vanguard	100.4
RT-Mart Shanghai	80.7
Wal-Mart (China)	72.2
Lianhua Supermarket (Lotus)	68.8
Shandong Commercial Group	61.1
Shanghai Friendship Group	60.8
Chongqing General Trading Group	60.3
Yum! Brands Inc. (China)	50.2

SOURCE: Statista (2015).

whereas small local brands have relied on street stores, supermarkets, or hyper-marts as outlets. As of 2013, department stores and specialty stores dominated the retail sales of clothing with market shares of 36.3 percent and 29.7 percent, respectively (Lu 2014). The department stores and specialty stores, which are normally located in city centers, are closely tied to the interests of local government through land-leasing contracts or direct ownership. The rise of online shopping has forged an alliance between small brands and Internet portals, which present more of a threat to street stores, supermarkets and hyper-marts than to department stores and specialty stores.³

O2O Business

O2O is a more recent business model on the Internet, which allows consumers to purchase services online while consuming the services offline. It may apply to services such as restaurants, movies, fitness clubs, tourism, and the like, which cannot be packed and sent to consumers in the form of parcels. O2O operators offer services like matching, payment collection, discount promotion (vouchers), or group purchases. They create value that resembles an agglomeration effect in the real sector by putting together a group of service providers in the Internet space for consumers to choose from. Along with the provision of information, the O2O operators also offer a location directory, store ratings, and so on to consumers and undertake product promotion campaigns for service providers. In essence, O2O facilitates trade by reducing information asymmetry. The largest O2O operator in China, Dazhong Dianping, boasts 200 million registered users on its website and 14 million service providers located in 2,500 cities. However, all services are local. The users can choose only from a list of service providers in a certain location where their consumption is to take place. Therefore O2O is a good technical tool to promote consumption in a specific location, serving the interests of the local government.

³Conflicts with local interests are more likely to arise if online shopping is modeled as a simple substitute for offline transactions. For example, the second largest online shop in China, JD.Com, adopts a business model similar to Amazon as a reseller rather than a platform, and consumer electronics is its major item of sales. It competes directly with the major retail chains of consumer electronics, notably Suning and Gome, which also hold online marketing divisions to parallel their offline operations.

The service industry has long been overlooked in the course of economic development in China. Until recently, China's industrial policy had consistently been in favor of the manufacturing industry. Services, especially consumer services were largely ignored or even discriminated against by local officials. Compared to other industries, it is relatively difficult for consumer service providers to access land or finance, or to obtain business licenses as they are considered low-tech and add little value to the land. However, that policy stance has changed in recent years as manufacturing-led growth has slowed. As a manifestation of the policy change, the central government called for the promotion of "high-value" consumer services in the 12th Economic Plan (covering 2011–15).

The relatively underdeveloped Chinese service industry can be explained by a lack of marketization (division of labor), a lack of innovation, and inadequate demand from consumers (Cheng 2013). It is partially caused by strict regulations in certain service sectors such as education and medical services. Although the share of service expenditure in consumer demand has increased in recent years, the increase has mainly been accounted for by an increase in the price of services rather than quantity (Cheng and Blanchard 2009). This has occurred because the demand for education and medical services is inelastic, a reflection of the Baumol disease (Baumol and Bowen 1966). As consumers have been forced to spend a larger share of their income on such services, they have become increasingly sensitive to the prices of other consumer services and the O2O app has become a handy tool to help them make smart shopping decisions.

Most successful Chinese O2O operators focus on modern consumer services such as restaurants, movies, tourism, and fitness clubs, for which the demand tends to rise with personal income. The O2O services target young consumers, or the so-called post-1980 and post-1990 cohorts residing in urban areas. This group of consumers depends on mobile devices for information collection and exchange. Therefore, the O2O business is essentially a mobile business where the service providers help the sellers disseminate information (e.g., in the case of new shop openings or excess capacity, or to explore new consumers). The established service providers are willing to collaborate with O2O operators as they see this as an expansion of their marketing channels, while new entrants use O2O channels to promote consumer awareness.

The development of the modern consumer services industry is in the interests of urban governments that are looking for a new growth engine where the manufacturing-centered growth model has imposed excessive environmental and population burdens on the cities. Modern consumer services that cater to the younger generation are often co-located with high-end retailers in modern shopping malls. In recent years, local governments have had a strong craze for shopping malls, which have become favorite investments for the purpose of land-value enhancement. For example, the real estate research firm CB Richard Ellis tracked new shopping mall construction in 2012 in 180 major cities around the world and found that more than half of the new malls under construction were in China (Rapoza 2013). Second-tier cities like Chengdu, Tianjin, and Shenyang top the list in terms of shopping mall investments. O2O services, which promote the offline consumption of modern services, also enhance the value of modern shopping complexes and are therefore welcomed by the local governments.

Making Peace with the State

The Chinese Internet industry has developed largely without government initiatives, such as policy guidance, financial subsidies, or tax incentives. The Internet industry is dominated by a few large firms, notably Baidu in search engines, Alibaba in online shopping, and Tencent in messaging and online games, which are together known as the BAT. The BAT has driven foreign competitors, including Google, Amazon, and eBay, out of China, thereby materializing the country's aspiration for national champions (Nolan 2001). Unlike in other industries in China where government intervention is prevalent, in the Internet industry government intervention is conspicuously absent either in terms of market entry restrictions or regulations on business operations. What, then, makes the Internet industry so different from the others?

The explanation lies with the existence of national champions in BAT, which act as policemen and rule makers in Chinese cyberspace. The BAT is a surrogate of the state in the enforcement of public regulations, including opinion controls, which top the list of government interests. The BAT offers domestic substitutes for search engines like Google, and social media like Facebook, Line, and Twitter, allowing the government to control and monitor information dissemination on

the Internet. While performing these important social functions for the state, the BAT exploits its monopoly power over the Internet for business benefits. As traffic is the king in Internet competition, three BAT firms compete with one another in amassing large numbers of users, and offer new products and services through innovation to do so. To maximize the number of users, the BAT firms position themselves as platform operators rather than store owners. They create public goods to fuel the industry growth and establish rules to maintain orderly trade on the Internet. They follow the policy winds closely and always act before the government to foreclose possible intervention. When disputes arise with the government, they compromise to preserve their core business interests. Although such compromise often constitutes a barrier to business expansion, they innovate in order to break through the barrier. In fact, innovation under institutional constraints has been the engine of growth in the Chinese Internet industry.

With this “peace making” strategy, the central government has until now imposed only a few regulations on the Internet industry. In fact, there is no basic law in China that governs Internet transactions such as electronic signatures or electronic money transfers. Piecemeal regulations have been introduced by some local states that are interested in developing the Internet industry in their regions. For example, Guangdong Province issued an e-commerce protocol (*Guangdong sheng dianzi shangwu tiaoli*) in 2004, giving electronic documents the same legal status as written documents. Zhejiang Province issued an ordinance on e-commerce (*dali tuidong wangshang shichang kuaisu jiankang fazhan de ruogan yijian*) in 2008, instructing the subprovincial government agents to lower the entry barriers to e-commerce, including making offline business registration automatically valid for online trade. It is no coincidence that Guangdong and Zhejiang respectively host two members of the BAT, namely, Tencent and Alibaba.

It is not unusual for industry-leading firms to impose self-regulation in order to avoid government regulations, to differentiate themselves from their peers, or to enhance the legitimacy of private firms (Haufler 2001). Self-regulation is often imposed to protect public interests, such as the environment or labor rights, which are vulnerable to private abuse because of inadequate public regulations. The reason why the government does not impose sufficient regulations is often because of the conflict of interests between different

parties that wield political influence. Private regulators have the discretion to side with particular interest groups, and it is desirable to do so if this stance creates a competitive edge in business-to-business or business-to-consumer relations (Sorsa 2010). The competitive edge enhances the profitability and survival of self-regulated firms.

Self-imposed rules of the BAT are often directed toward the consumers whose interests in offline trade are inadequately protected under the government regulations. The BAT sets liberal return policies, subsidizes consumers from time to time, gives an upper hand to consumers in the case of a dispute, and allows consumers to choose their own logistics service providers when shopping, and to rate the quality of service to determine the service price, and so on. On the other hand, vendors or service providers are obliged to cooperate with the platform operators in promotion campaigns and will be disciplined for unethical behavior. Siding with consumers enhances consumer loyalty. The rules are aimed at boosting the Internet traffic that ultimately determines the outcome of Internet competition. The rules therefore serve the BAT's own interests. The government may be forced to endorse the private rules if they have become industry standards. Indeed, the BAT's return policy had been practiced long before the Chinese government amended the Consumer Rights Protection Law in 2014 which mandated all retailers to accept product returns within seven days of the purchase, and Internet retailers have to do so without asking for reasons. The case suggests that private regulations, if successful, may define the course of public regulations.

Private regulations may also preempt government regulations by making them unnecessary. This happens when private regulations incur lower costs than government regulations. Rules that enforce online trading contracts are primary examples. National monopolies like the BAT can enforce trading contracts more effectively than the government because they have better information about the traders, and because they control the payment mechanisms. They also control market entry and may discipline violators without administrative or court procedures. In order to maintain trading order, it is actually more effective for the government to control a few platform operators like BAT than millions of small traders. For example, the city of Hangzhou, where Alibaba is headquartered, issued an ordinance on Internet transactions in March 2008 (*Hangzhou shi wanglu jiaoyi guanli zanxing banfa*), requiring all online stores to operate on a

third-party platform instead of establishing their own portals. The platform operators bear the responsibilities of examining product authenticity, conducting credit assessments of traders, and preventing malicious competition or unwarranted disturbances of consumers by vendors through the Internet.

Private regulations, while serving the public interest, may also constitute an entry barrier for latecomers (Curran 1993), thus protecting the market position of leading firms. For example, by setting up consumer-biased trading rules and amassing a large consumer base, it is hard for latecomers to compete with a smaller consumer base. Scale is a natural barrier to entry for Internet businesses. With a large consumer base at hand, vendors have either to go along with the existing rules set by BAT or to establish their own portals if they are to enter online trade. Large vendors and famous brands may choose the latter option, but they will not become platform operators to rival the BAT. In other words, private regulations as such serve the purpose of market differentiation. The BAT has targeted small vendors as the main customers and has consciously avoided competing with large vendors whose interests are likely to be tied to the state.

However, it is also possible that private regulations serve the public interest only superficially. An organization's compliance with government regulations or social norms may only be ceremonial thereby entailing resource costs without real effects. Chinese consumers have often complained about the rampant counterfeit products in online trade, and the SAIC regulations require that the platform operators enter a contract with the vendors so that the latter are held liable for the counterfeits. In reality, however, no private contracts may actually ensure the authenticity of the products. The platform operators of the BAT simply impose a liberal return policy to ease the grievances of consumers if they are unhappy with the products. In other words, consumers are asked to decide whether the products are authentic or not. In January 2015, SAIC issued a report (*guanyu dui Alibaba jituan jinxing xingzheng zhidao gongzuo qingxing baipishu*), indicating that only 36.25 percent of products sold on Taobao (of Alibaba) are authentic, but no actions were taken to penalize Alibaba. Legally, it is the vendors that are liable for the counterfeits, not the platform operators, who only cooperate with the authorities in law enforcement. The extent of cooperation is negotiable, and the SAIC report appeared to be a bargaining ploy. In fact, Internet trade offers more policy tools for the government to control counterfeits if the

issue is really taken seriously. Alibaba has organized a 700-member counterfeit-fighting task force which polices the platform to spot and remove counterfeits. A formal structure is a convenient tool that serves the ceremonial purpose of compliance to social norms or public regulations to earn legitimacy for the organization (Meyer and Rowan 1977).

In short, private regulations imposed by the BAT have set the standards for public regulation, have made the public regulations redundant, and have differentiated the BAT from its peers. All these serve the business interests of self-regulators and solidify their monopoly position. On the other hand, the BAT's compliance with public regulations is selective. In areas where state interests are non-challengeable, such as controlling public opinion, they comply with good efforts. In areas where the compliance cost is high, they conform to the regulations by structures rather than actions.

In addition to rule making, the BAT also creates public goods for the Internet industry. The most important public goods created by the BAT are online payment facilities, such as Alipay, which offer third-party escrow accounts to enable transactions. The third-party accounts protect consumer interests and guarantee payments to vendors at the same time. Alipay may have been an imitation after PayPal of the United States, but its success was possible only with Alibaba's long-term investment which gradually built up the public confidence in the mechanism. In 2010, six years after the inauguration of Alipay and many subsequent mimics, the central bank, People's Bank of China (PBOC), decided to regulate third-party payment accounts such as Alipay. The PBOC required account operators to obtain a license from the PBOC and to deposit their balances in commercial banks.⁴ This allowed the banks to share the benefits of online trade without having to offer online accounts to Internet shoppers or vendors for whom the service costs might have been uneconomical because of the difficulty in performing credit assessments on them. As a result, third-party payment account operators like Alipay now work as deposit collectors for the commercial banks, which are mostly state-owned. Commercial banks make money mainly from the interest rate spreads between deposits and loans; fees charged on

⁴As of July 2013, the PBOC had issued 250 licenses for third-party payment accounts.

payment collections like credit cards constitute only a small fraction of bank earnings.

The separation between third-party payment accounts and bank accounts was apparently a compromise between the state and the BAT. While requiring third-party payment account balances to be deposited in commercial banks, the PBOC did not stop the BAT from offering short-term credit to Internet traders. By doing so, Alipay and its counterpart of Tencent, Weixin, have become giant microfinancing organizations that service Internet traders. Traders' balances in the online payment accounts bear interest, and while this is at a rate that is lower than the bank rate, the balances can be applied to online and offline purchases at any time. As more and more offline shops have accepted Alipay and Weixin for payments, they have become China's dominant mobile payment instruments, foreclosing the business chances of foreign competitors like Apple Pay or Google Pay. Initially Alipay and Weixin were invented to facilitate online payments to supplant the institutional weakness arising from the lack of credit cards in circulation in China. In the end, these instruments work like interest-bearing debit cards in offline transactions. Today, the PBOC still prohibits the BAT from issuing online credit cards, which are reserved for commercial banks to defend the latter's interests.

As more and more traders hold Alipay or Weixin accounts, these payment mechanisms have become true public goods. Individual account holders can use these accounts to send gifts or make unilateral transfers without engaging in trade. C2C traders can also enact a transfer of funds from the buyer's account to the seller's without going through the platform's trading portals, thus avoiding the entry fees or commission charges. In other words, they can engage in an "underground" transaction by leveraging the online payment facilities. Public goods cannot avoid free-riding, which has been tolerated by the BAT. Some scholars argue that China now tops the world in Internet finance in terms of the volume of transactions or diversity of financing arrangements (Ma et al. 2014: 14). Many innovations in Internet finance have been precipitated by the weaknesses of the banking industry, including a lack of credit cards, an inability to perform credit assessments, credit rationing, and so on. Unlike shadow banking activities that threaten the stability of the banking system in China, Internet financing benefits the banking system by enlarging the deposit base of the commercial banks. Small young traders, who

are normally outsiders to the banking system because of a lack of creditworthiness, are now engaged in the system through Internet finance. Better still, all financial transactions take place before the eyes of the regulatory authorities, who may choose to intervene at any time.

Conclusion

The Chinese economy is half-open and half-closed. The real sector is open to both foreign trade and foreign investment. However, domestic business activities are heavily regulated, giving rise to rent-seeking behavior. Rent is created through the government ownership of land and licensing controls on business activities. Rent seeking is accompanied by high entry barriers, which suppress domestic entrepreneurship. On the contrary, the Internet sector is closed to foreign operators in favor of domestic monopolies, who offer a platform for entrepreneurial activities. Through innovations, Internet entrepreneurs uncover the market opportunities hidden behind the distortions created by rent seekers in the real sector.

All Internet innovations seem to have one common thread, which is to remedy the deficiencies of the real-sector economy. Notable Internet innovations include those that explore new market frontiers, reduce transaction costs, exploit scale economies, eliminate information asymmetry, and so on. However, none of these innovations have engendered a destructive effect on the real-sector economy as the government, or its proxy, always controls complementary assets for the realization of Internet innovations. With these complementary assets, it gains rather than loses from Internet innovations. This explains why the government has, until now, kept its hands off the Internet industry.

The Internet creates a new platform for trading which is characterized by virtually free market entry and low transaction costs. Internet trade in China constitutes an economic sector that is analogous to the underground economy in other developing countries. However, unlike an underground (informal) economy that survives by evading taxes and circumventing government regulations, the Chinese Internet sector is taxed and regulated. To the extent that Internet trade is subject to the same tax and regulatory burdens, it reduces the threat to the formal sector, and preserves the opportunity for rent creation by the local government. The Internet sector

sustains itself by low entry and low transaction costs, and therefore represents a true free-market economy in China.

Like the informal sector in other countries, the Chinese Internet sector is separated from the formal sector in terms of products offered and factors employed. Internet products are relatively unknown brands characterized by small-scale production, whereas the real sector favors large vendors and famous brands. The Internet sector employs its own resources for branding, manufacturing, marketing, financing, transportation, and after-sales services that are distinct from the resources used in the real-sector production. Differences in product and factor prices between the real and Internet sectors persist and are not to be arbitrated.

Similar to the informal sector which is subject to the rules of informal institutions, the Chinese Internet sector is regulated by rules established by the platform operators. The platform operators have a strong incentive to self-regulate to avoid government intervention. They are conscious of the central government priorities such as controlling public opinion and maintaining trade order. Platform operators also create public goods to facilitate online trade, which is otherwise infeasible under the real-sector institutions, like payment mechanisms and microfinancing. Self-regulation and public goods serve the public interest as well as the platform operators. The Chinese Internet is therefore a case of strong organizations coupled with weak institutions. While strong organizations are necessary to make up for the weaknesses of institutions, strong organizations cannot survive without state patronage. This structure underscores the stable relationship between the Internet operators and the state.

References

- Adams, P. (2011) "Nationalizing China: Private Sector Is Being Squeezed Out as 'the State Advances.'" *Financial Post* (June 23). Available at <http://business.financialpost.com/fp-comment/nationalizing-china>.
- Aghion, P. N., and Griffith, R. (2005) *Competition and Growth: Reconciling Theory and Evidence*. Cambridge, Mass.: MIT Press.
- Allen, F.; Qian, J.; and Qian, M. J. (2005) "Law, Finance, and Economic Growth." *Journal of Financial Economics* 77: 57–116.

- Atomico (2015) “Atomico Report on VC.” Available at <http://atomico.com/explore-d3>.
- Bai, C. E.; Du, Y. J.; Tao, Z.; and Tong, S. (2004) “Local Protection and Regional Specialization: Evidence from China’s Industries.” *Journal of International Economics* 63 (2): 397–417.
- Baumol, W. (1990) “Entrepreneurship: Productive, Unproductive, and Destructive.” *Journal of Political Economy* 98 (5): 893–921.
- Baumol, W., and Bowen, W. (1966) *Performing Arts, the Economic Dilemma: A Study of Problems Common to Theater, Opera, Music, and Dance*. New York: Twentieth Century Fund.
- Baumol, W.; Litan, R.; and Schramm, C. (2007) *Good Capitalism, Bad Capitalism, and Economics of Growth and Prosperity*. New Haven, Conn.: Yale University Press.
- Chen, T. J. (2016) “The Development of China’s Solar Photovoltaic Industry: Why Industrial Policy Failed.” *Cambridge Journal of Economics* 40 (3): 755–74.
- Cheng, D. Z. (2013) “The Development of the Service Industry in the Modern Economy: Mechanisms and Implications for China.” *China Finance and Economic Review* 1 (3): doi:10.1186/2196-5633-1-3.
- Cheng, D. Z., and Blanchard, J. M. (2009) “Is the Chinese Economy Moving in the Direction of Tertiarization?” *China World Economy* 17(6): 88–105.
- Chin, H., and Chow, A. (2012) “The Case for China’s Retail: Issues and Opportunities.” Available at http://research.Prudential.com/documents/rp/PRU_China_Retil_0312.pdf.
- Curran, C. (1993) “The American Experience with Self-Regulation in the Medical and Legal Professions.” In M. Faure, J. Finsinger, J. Siegers, and R. Van den Bergh (eds.), *Regulations of Professions: A Law and Economics Approach to the Regulation of Attorneys and Physicians in the US, Belgium, the Netherlands, Germany and the UK*, 47–87. Antwerp, Belgium: Maklu.
- Djankov, S.; La Porta, R.; Lopez-de-Silanes, F.; and Shleifer, A. (2002) “The Regulation of Entry.” *Quarterly Journal of Economics* 117 (1): 1–37.
- Dow Jones Venture Source (2015) “The Billion Dollar Startup Club.” Available at <http://graphics.wsj.com/billion-dollar-club>.
- Duan, H., and Chik, A. R. (2012) “Institutional Environment, Political Connection and Financial Constraints: Evidence from Private

- Enterprises in China.” *Business and Management Research* 1 (1): 133–40.
- Haufler, V. (2001) *Role for the Private Sector: Industry Self-Regulation in a Global Economy*. Washington: Carnegie Foundation for International Peace.
- Hsia, L. S. (2007) “From the Phenomenon of ‘SOEs Withdraw PREs Expand’ and ‘SOEs Expand PREs Withdraw’: Examining the Development of Privately Run Enterprises.” *Zhanwang Yu Tansuo Monthly* (Taiwan) 5 (10): 18–47.
- Huang, Y. (2008) *Capitalism with Chinese Characteristics: Entrepreneurship and the State*. Cambridge, U.K.: Cambridge University Press.
- Inside (2013) “Ma Yun: Chuantong Lingshou Ye Yu Wanglu De Jingzheng, Jiuxiang Zai Jiqiang Mianqian, Shenme Quan Dou Meiyong (“Ma Yun: Competition between Traditional Retail and Internet, All Martial Arts Are Useless before Machine Guns”). Available at www.inside.com.tw/2013/01/24/jack-ma.
- Kirzner, I. (1979) *Perception, Opportunity, and Profit: Studies in the Theory of Entrepreneurship*. Chicago: University of Chicago Press.
- Kung, J. K., and Ma, C. (2014) “Friends and Benefits: How Political Connections Help Sustain Private Enterprise Growth in China.” Working Paper. Hong Kong University of Science and Technology, Institute for Advanced Study.
- Kwock, B.; James, M.; and Tsui, A. (2013) “Doing Business in China: What is the Use of Having a Contract? The Rule of Law and Guanxi When Doing Business in China.” *Journal of Business Studies Quarterly* 4 (4): 56–67.
- La Porta, R.; Lopez-de-Silanes, F.; Pop-Eleches, C.; and Shleifer, A. (2004) “Judicial Checks and Balances.” *Journal of Political Economy* 112: 445–70.
- Li, H., and Zhou, L. N. (2005) “Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China.” *Journal of Public Economics* 89 (9):1743–62.
- Li, H.; Meng, L. S.; Wang, Q.; and Zhou, L. A. (2008) “Political Connections, Financing, and Firm Performance: Evidence from Chinese Private Firms.” *Journal of Development Economics* 87 (2): 283–99.
- Lin, Y. M. (2009) “The Institutional Context of Rent Seeking in Economic Transition.” In T. W. Ngo and Y. P. Wu (eds.), *Rent Seeking in China*, 59–78. London: Routledge.

- Lu, S. (2014) "China Apparel Market (Updated December 2014)." Available at <http://tmd433.worldpress.com/2014/12/21/china-apparel-market-updated-in-december-2014>.
- Ma, M.; Zhu, X. M.; Zhou, J. H.; Ji, J. Y.; and Chen, Y. (2014) *Zhifu Geming (Payment Revolution)*. Beijing: China CITIC Press.
- McKinsey Global Institute (2013) "China's E-Tail Revolution: Online Shopping as a Catalyst for Growth." Report by R. Dobbs, Y. Chen, G. Orr, J. Manyika, M. Chui, and E. Chang. Available at www.mckinsey.com/global-themes/asia-pacific/china-e-tailing.
- Meyer, J., and Rowan, B. (1977) "Institutional Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83 (2): 340–63.
- Murphy, K.; Shleifer, A.; and Vishny, R. (1990) "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics* 106 (2): 503–30.
- Nolan, P. (2001) *China and the Global Economy: National Champions, Industrial Policy and the Big Business Environment*. New York: Palgrave Macmillan.
- North, D. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge, U.K.: Cambridge University Press.
- Oi, J. (1995) "The Role of the Local State in China's Transition Economy." *China Quarterly* 144 (December): 1132–49.
- Rapoza, K. (2013) "China Tops World's Shopping Mall Craze." *Forbes* online: www.forbes.com/sites/kenrapoza/2013/07/11/china-tops-worlds-shopping-mall-craze.
- Schneider, F. (with D. Enste) (2002) "Hiding in the Shadows: The Growth of the Underground Economy." *IMF Economic Issues* No. 30. Available at www.imf.org/external/pubs/ft/issues/issues30.
- Sorsa, K. (2010) "Private-Regulation in Global Value Chain: A Trade Barrier or an Opportunity for Public-Private Cooperation." Working Paper, Turku University of Applied Sciences, Finland.
- Statista (2015) "Retail Industry in China: Statistics and Facts." Available at www.statista.com/topics/1839/retail-in-China.
- Stigler, G. (1971) "The Theory of Economic Regulation." *Bell Journal of Economics* 2 (1): 3–21.
- Wade, R. (1990) *Governing the Market: Economic Theory and the Role of the Government in East Asian Industrialization*. Princeton, N.J.: Princeton University Press.

- Wang, E., and Song, J. (2008) "The Political Economy of Retail Change in Chinese Cities." *Environment and Planning C: Government and Policy* 26 (6):1197–1226.
- Wedeman, A. (2003) *From Mao to Market: Rent Seeking, Local Protectionism and Marketization in China*. Cambridge, Mass.: Cambridge University Press.
- World Bank (2015) *Doing Business 2015*. Washington: World Bank.
- Wu, J., and Huang, S. (2008) "Innovation or Rent Seeking? The Institutional Environment and Entrepreneurial Behavior during China's Economic Transformation." *China and World Economy* 16 (4): 64–81.
- Wu, Z., and Piao, V. (2014) "Didi Dache Rongzi Qiyi Meiyuan Tuozhan Xiongxin" ("Didi Dache Raised 700 Million U.S. Dollars to Show Ambition"). *New York Times*, Internet Version in Chinese (December 11): <http://cn.nytimes.com/business/20141211/c11didi/zh-hant>.
- Yang, J. J. (2012) "Research Reviews on 'Chinese Land Finance.'" *Advances in Applied Economics and Finance* 3 (4): 630–33.
- Young, A. (2000) "The Razor's Edge: Distortions and Internal Reform in the People's Republic of China." *Quarterly Journal of Economics* 115: 1091–1135.
- Yu, G., and Zhang, H. (2008) "Adaptive Efficiency and Financial Development in China: The Role of Contracts and Contractual Enforcement." *Journal of International Economic Law* 11 (2): 459–94.
- Zhen, Y. (2007) *Globalization and the Chinese Retail Revolution: Competing in the World's Largest Emerging Market*. Oxford, U.K.: Chandos Publishing.

THE MAKING OF A STATE: TRANSITION IN MONTENEGRO

Igor Lukšić and Milorad Katnić

The first Montenegrin state started to take shape in the 8th century with the arrival of the Slavs and their mingling with the local population. Originally it was called Doclea, whose ruler received a royal insignia by the Pope Gregory VII in 1078 (Andrijašević and Rastoder 2006). Montenegro fell under the Ottomans in the late 15th century, but acted as a de facto independent state until formal recognition came at the Berlin Congress in 1878. Despite being on the victors' side in the Balkan Wars and in World War I, it was annexed by Serbia and lost its sovereignty in 1918. After the Second World War it became a part of socialist Yugoslavia, where it remained until 1992.

Montenegro's political transition started in earnest after the Belgrade Agreement signed in March 2002. Montenegro held an independence referendum in 2006 and was subsequently admitted to the United Nations and other international organizations. Today Montenegro is engaged in accession talks with the European Union (EU).

Political Aspects of Transition in Montenegro

At the beginning of the 1990s, the process of opening a socialist society began in Montenegro, as it did in most other countries

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

Igor Lukšić is Associate Professor at the University of Donja Gorica Montenegro. He is former Prime Minister, Deputy Prime Minister, and Foreign and European Affairs and Finance Minister. Milorad Katnić is Associate Professor at the University of Donja Gorica Montenegro and Chairman of the Societe General Montenegro Board of Directors. He is former Finance Minister and Vice Finance Minister. The authors thank Sladjana Pavlović for contributing to earlier versions of this article.

undergoing transition. The first multiparty ballot elections and the establishment of the first democratic government took place. Since the beginning of transition, the Democratic Party of Socialists (DPS), successor of the League of Communists, has been the main political driver in the country. One can say that the transition in Montenegro has been characterized by program changes of the DPS, which also defined the country's path.

At the beginning of transition, the political and economic processes in Montenegro were under the dominant influence of the transition processes in the bigger member of the federation, Serbia. At the same time, the character of the Montenegrin society, which was predominantly a traditional society, made for a slow opening up to Europe. This stage of transition is characterized by a clash within the communist elite, as a pro-reform opposition never took form (Darmanović 2003). Shortly after the changes, the Democratic Forum was established, a roundtable which prompted the establishment of multipartism.

In order to understand Montenegro's institutional setting, it is important to remember that the Yugoslav republics enjoyed significant rights stemming from the 1974 federal Constitution. Even prior to 1992, Montenegro had a number of institutions such as the constitutional court, parliament, government, Ministry of Foreign Affairs, and central bank. This institutional history makes Montenegro's situation somewhat different from other newly independent countries. Slovakia, for example, started creating institutions from scratch after the Velvet Divorce. So did the three Baltic countries after 1991.

Still, most of the political decisions were made in Belgrade, and Montenegro's institutions were quite weak. Additionally, the political process of transition in Montenegro was weighed down by civil wars in the former Yugoslavia and sanctions imposed by the United Nations in 1992. Thus, the political transition in Montenegro truly started after the clash in the ruling Democratic Party of Socialists in 1996. The DPS split into two factions: one that remained loyal to the Milošević regime in Belgrade with the then President Momir Bulatović at the helm; and the other, a pro-Western faction, led by Prime Minister Milo Djukanović. These intraparty clashes created conditions for a new stage in transition and resulted in the Agreement on Minimum Principles for Development of a Democratic Infrastructure in Montenegro. It provided guarantees to the opposition that the electoral process would be fair and to the

pro-reform part of the ruling party that it would receive the necessary support to oppose the official stance of Belgrade. The result was the election of Djukanović as the President of Montenegro in 1997.

Not only did Djukanović oppose Milošević at a great personal risk, but he also steered the country toward economic reform. Montenegro was able to pursue many reforms including those in the second wave (mid- to end 2000s) that we were in the position to propose and implement.¹ On many occasions, such as the decision to introduce the German mark as a legal tender in late 1999, President Djukanović was under significant pressure from the international community to reconsider, but he did not waver.

However, this period did not see a democratic consolidation of Montenegro's political regime due to the unresolved issue of sovereignty. The political clash between Podgorica and Belgrade at times threatened to turn into a military clash. The defeated candidate in the 1997 elections, Bulatović, was later appointed federal prime minister in an attempt to discipline Montenegro—a position he retained until the fall of Milošević in 2000. This stage lasted until a political agreement was reached on transforming the Federal Republic of Yugoslavia into a new entity called Serbia and Montenegro in March 2002. The agreement established a union between Montenegro and Serbia for a three-year term. This was a crucial interim step toward independence. However, the implementation of the union took a full year following the signing of the Belgrade Agreement.

The watershed year was 2006, at which time referendum rules were defined with the intermediation of the international community. Those rules included the provision that more than 55 percent of those casting their vote must opt for independence for Montenegro to restore its statehood. The referendum was held on May 21, 2006, bringing victory to supporters of independence by a margin of 55.5 percent to 44.5 percent, with an almost 90 percent turnout.

The newly established independence and the smooth divorce from Serbia meant that the process of institutional development could start for real.² Immediately following the declaration of

¹“We” refers to Igor Lukšić, Milorad Katnić, and Vladimir Kavarić, the current economy minister and earlier a colleague of Lukšić at the Ministry of Finance.

²We were very quick to divide assets and liabilities thanks to the fact that Serbia and Montenegro already operated as economically independent states. Thus, it took only two rounds of talks (June/July 2006) to reach the agreement over independence, which Lukšić signed with Serbian Finance Minister Mladjan Dinkić.

independence and after gaining international recognition, two key strategic state objectives were defined: membership in the European Union and in NATO. Those two goals largely act as an anchor for political reforms. Montenegro also signed the Stabilization and Association Agreement with the EU in October 2007, and shortly thereafter adopted the new Constitution, while acquiring EU candidate status at the end of 2010.

Entry into the EU meant progress in six institutional and legal areas as a condition for opening accession negotiations: the electoral process, judicial independence, fundamental human rights, the fight against corruption and organized crime, cooperation with the civil sector, and strengthening the independence of the media. By mid-2012, sufficient progress was made and Montenegro opened membership negotiations with the EU, thus entering the latest stage of economic transition. By June 2016, out of the total of 35 negotiating chapters 22 were opened.³ Assuming a successful completion of negotiations by 2019 and the ensuing ratification of the future membership agreement, Montenegro could become a full member of the EU in 2021.

Concurrently with the process of European integration, Montenegro is taking steps toward NATO membership, at first in the form of the Partnership for Peace, and then through the Membership Action Plan. At the end of 2015, NATO made the decision to invite Montenegro to join the alliance, with full-fledged membership to occur in 2017.

The further development of democracy and economic reform in Montenegro should be geared toward developing a modern state. According to Freedom House (2015), Montenegro today is a “partially free country.” The rule of law needs to be further developed by encouraging a political culture of dialogue, tolerance, transparency, and accountability. Countries in transition that have steered their political system in this direction have built functioning democracies and successfully implemented economic reforms—in particular, we studied the experience of Poland, Estonia, and Slovakia.

³After taking over the position of foreign minister in December 2012 until he stepped down in April 2016 to run for the position of the Secretary General of the UN, Lukšić led Montenegro’s delegation in the intergovernmental conferences between the EU and Montenegro to open negotiating chapters.

Economic Aspects of Transition in Montenegro

The economic transition in Montenegro, like the political one, is a process taking place in several stages. The first stage is related to the period since the beginning of the 1990s, with the introduction of multipartism. The second relates to the period 1998–2002 when a definite discord took place between Podgorica and Belgrade. The third stage started with the election of the new government in January 2003 and lasted until the declaration of independence in 2007. The fourth stage occurred during the period between the restoration of independence and the opening of EU membership negotiations in 2012. The current and final stage will last until EU membership is obtained, hopefully in 2021.

Even though economic transition in Montenegro started in the same period as in other countries of the former eastern bloc, it was soon thwarted due to irregular political conditions in which the economic transformation was taking place. Politics came to dominate the economic process, and progress toward greater economic freedoms came to a halt. This period is characterized by the economic blockade imposed on Montenegro due to international sanctions introduced by the United Nations on the Federal Republic of Yugoslavia in 1992. Gradual withdrawal of sanctions began only in 1996. At the same time, due to the loss of the common market of socialist Yugoslavia on which Montenegro's economy predominantly depended, as well as wars in the neighborhood, the GDP took a dramatic plunge. Using 1989 as the base year (100), GDP fell to 89 in 1990, 70 in 1991, 61 in 1992, and 39 in 1993. Only in 1995 did the economy start growing again. This respite was cut short after the bombing of the Federal Republic of Yugoslavia in 1999, and GDP started falling again, thus significantly retarding the postcommunist recovery (ISSP 2004: 8).⁴

The war-torn period also led to a large inflow of refugees. At a certain stage, they made up more than 20 percent of Montenegro's population. Enormous increases in the number of inhabitants, mostly

⁴The Institute for Strategic Studies and Prognoses (ISSP) in Podgorica was the first proper economic think tank in Montenegro. Katnić worked at ISSP for a time and Lukšić drew upon ISSP's policy studies. Other prominent members of the government or administration also worked with the ISSP to promote economic liberalism.

vulnerable ones, exerted intense pressure on the already exhausted budget and social funds. Survival policies were implemented in this environment, and deeper economic reforms aimed at increasing economic freedoms were postponed.

The second stage of economic transition in Montenegro was dominated by the political conflict between Belgrade and Podgorica. That period can be divided into two subphases: 1998 until October 2000, when political changes occurred in Serbia; and October 2000 until January 2003, when Montenegro's new government was established. In this period, the functional and institutional sovereignty that preceded the formal restoration of independence was established. Functional sovereignty meant independent decisionmaking, while institutional sovereignty implied freedom and the possibility of establishing new institutions. The issue of state sovereignty dominated the period 1998–2002, thus slowing the process of economic transformation.

The establishment of an independent economic system began as well, in addition to distancing from the official politics of Belgrade. Montenegro adopted measures affecting monetary, financial, and fiscal policies, as well as foreign trade. Although nominally Montenegro remained part of the federation, the pursuit of economic independence gained pace. This process was strongly backed by President Djukanović who was the personification of the process in which the government was taking back different competences from the federal level step by step. At the same time, Montenegro took over the process of defining ownership rights and privatization. The key person in this process was Veselin Vukotić, a prominent economist, who inspired and implemented the overall economic reforms and the privatization agenda (see Vukotić 2002, 2003, 2005; also Vukotić and Pejovich 2002).⁵

⁵Vukotić (currently leading private university UDG) was the vice prime minister in the last Yugoslav reformist government led by Ante Marković that lasted until the dissolution of the SFRY. He was in charge of implementing the privatization agenda. After returning to Montenegro, he reemerged as the key personality in late 1990s helping the government adopt the policy of economic liberalism, privatization, and entrepreneurship. Lukšić and Katnić were Vukotić's students in the 1990s, and Prime Minister Djukanović was a student in the 1980s. Vukotić was also pivotal in putting together and later implementing the Economic Reform Agenda that was the centerpiece of the Djukanović government (2003–06), when Lukšić was finance minister and Katnić was deputy finance minister in charge of international cooperation and financial policies.

Monetary Reforms as the Cornerstone of Other Reforms

The most important change in the economic system occurred in the monetary sphere, in November 1999, when the German mark was introduced as parallel legal tender to the Yugoslav dinar. The German mark was previously often used in Montenegro and Serbia for both savings and exchange, due to the long experience of hyperinflation. However, it was not an easy decision because it took a lot of effort to get well-prepared, and there was political risk. After the NATO intervention in 1999, which ended in June, there was tangible risk of hyperinflation. The government and local experts led by Vukotić (assisted by some foreign experts like President Djukanović's advisor Steve Hanke, who had assisted in establishing the currency board in Bulgaria) started to consider and reflect on different options to ensure a sound currency. Two options were considered: a currency board and direct introduction of the German mark as the legal tender. The first option was eventually dropped, because it assumed backing of the IMF and World Bank, which was difficult under the circumstances. The second option was opposed by Belgrade, because it was a step in the direction of economic independence for Montenegro. It was also opposed by the international community, which feared it would lead to a new conflict in the Balkans and eventually the dissolution of what remained of socialist Yugoslavia. In the winter of 1999–2000, tensions were high but Montenegro's leadership endured. As of 2001, the German mark became the only official currency in Montenegro, which has been replaced by the euro as of March 2002.

Emphasis then shifted toward fiscal policy and control of public spending. It was no longer possible to print and to borrow money from the central bank for the purpose of bridging the budget gaps and resolving solvency issues. Economic imbalance became more visible, and there was no possibility of shifting responsibility to the federal level. The budget deficit was reduced, and the first fiscal surplus was produced in 2006. This reform led toward deeper economic changes, but politically it went even further. Both "one country, two systems" and confederation were eventually dropped in favor of full independence.

The process of economic liberalization took place simultaneously with monetary, financial, and fiscal reform. Administrative price controls were removed, trade constraints were eased (prohibition, quotas,

contingents, duty rates), and capital flows were liberalized, while the privatization process was stepped up. Voucher privatization, during which all adult Montenegrin citizens became owners of vouchers that could be invested into six privatization funds or 221 companies, took place during 2001 and 2002, concurrently with the process of sale of shares in state-owned enterprises.⁶

With the benefit of hindsight, the results of the privatization process were less positive than expected. Privatization led to the transformation of ownership and greater efficiency of companies that survived transition. Citizens became shareholders overnight and acquired free ownership in companies. However, a number of big socialist-era companies ceased to operate or substantially reduced their business activities and the number of employees. In addition, following a stock market boom in the years of expansion, the share prices fell dramatically in the majority of enterprises and funds. These processes, particularly a large number of older people who lost their jobs and who could not find new ones, still cause social pressure and make most citizens think of privatization in negative terms.

Economic Reform Agenda

Following the elections in January 2003, the government adopted the Economic Reform Agenda, which prescribed the tasks that should be fulfilled in order to continue economic transition. The most important task was to promote entrepreneurship as a moving force of economic growth. Simultaneously, a number of structural reforms and reforms aimed at improving the business environment were initiated.

One of the main tasks mentioned in the Economic Reform Agenda is reducing the gray economy. The issue of the gray economy is typical for all economies in which social-economic relations were

⁶In the first phase of economic transition (1990–98), the only important investments were privatization of Trebjesa Brewery at the end of 1997 and introduction of mobile telephony through the company Promonte in 1996–97. Only with the process of voucher privatization in 2001, and the parallel sale of the state property, did substantial foreign direct investment (FDI) occur in a number of companies after 2003. Privatization of state property included the oil distribution company, national telecommunication company Telekom, large industrial systems, as well as firms in the financial sector. The last large privatization of state property occurred with the sale of a minority package of shares in national Electric Power Company of Montenegro during 2009. After these actions, almost 90 percent of the economy was privatized.

established in irregular conditions. The introduction of international sanctions directly influenced the dramatic growth of the gray economy. The subsequent opening of Montenegro toward Europe influenced the reduction, but not the elimination, of the gray economy. The gray economy was notably present in the field of distribution of excise products, as well as the labor market, which was burdened with rigid regulations and high taxes and contributions that were at the level equal to or higher than paid salaries (see Ivanović and Kuchta-Helbling 2003).

The main topics of the Economic Reform Agenda were the substantial reduction in duty rates, liberalization of the foreign trade regime and its harmonization with the World Trade Organization (WTO) and EU standards, along with adoption of the new antitrust law. Participation in the global integration processes became even more important after the introduction of the “twin track” in the EU accession process, and after approval of individual membership of Montenegro in the WTO by the WTO assembly at the end of 2004. An important element of liberalization is the removal of all remaining quantitative import restrictions, as well as the implementation of new regulations in the field of accounting and auditing, in line with international practice.

Important tasks were also defined in the field of local self-governance. An important segment of overall reforms is the structural adjustment of the system, which implies the reform of extra-budgetary funds and the pension system—through reform of the intergenerational solidarity system and introduction of the third pillar of voluntary pension insurance.

Montenegro entered the 21st century with relatively high public spending. The largest part of public spending consisted of funds allocated for salaries, pensions, health care, and social benefits. The period from 2002 to 2006 was also characterized by the three-year arrangement with the IMF, which contributed to macroeconomic stabilization and acted as a catalyst for a number of structural reforms. The budget deficit was tightly controlled and tended to decrease. At the end of 2004, for the first time in its history, Montenegro received a credit rating from Standard & Poor's.

Tax Policy

After introducing the VAT in 2003, the next major reform of tax policy was initiated in 2004 through the gradual reduction of payroll

taxes. There was a lot of apprehension about a loss of revenue once tax rates were lowered. As finance ministers, we still vividly remember fighting with the IMF on the decision to cut payroll taxes by 10 percent. It proved to be a good decision. Every tax cut led to a higher level of revenues and a reduction in the share of the gray economy. The corporate profit tax rate was also lowered and was successful. In addition, in 2005, a flat tax was put in place. Flat taxation was quite popular and we were very keen to transfer experience from Slovakia, signing a cooperation agreement with Slovak Finance Minister Ivan Miklos in 2004. After that, Montenegro promoted the concept of single-digit taxation on income and profit, gradually reducing the income tax rate to the same level of 9 percent. This made Montenegro's tax regime one of the most competitive in Europe.

Macro Policy

One of the important transition reforms that occurred in late 2007 was transformation of the right to use land into proper property rights. This was a precondition for a series of foreign direct investments (FDIs) in the tourism sector. Time suggested weaknesses in our approach. Montenegro's lesson of transition in this segment is that the state property should be identified during the very early stage and the issue of restitution should be regulated under reasonable conditions, while voucher and other types of privatization should take place subsequently.

The implementation of economic reforms including the adoption of the euro, the flat tax of 9 percent, the removal of capital controls, and equal treatment of foreign and national investors made Montenegro a favored investment destination. A large inflow of capital through the FDI and banking channels led to a strong growth of the economy (Table 1). A comprehensive process of introducing European standards was also initiated, primarily through the implementation of the Stabilization and Association Agreement with the EU.

In only six years, 2003 to 2008, nominal GDP per capita nearly doubled, from approximately EUR 2.5 thousand to EUR 4.9 thousand. The real economic growth rate in this period averaged 6.2 percent per year. Economic growth was particularly dynamic during the last three years of the period before the eurozone crisis (2006–08) when the average real growth rate was nearly 9 percent.

TABLE 1
BASIC MACROECONOMIC INDICATORS
2000–2015

	2000	2001	2002	2003	2004	2005	2006	2007
GDP (EUR mil)	1,065.7	1,295.1	1,360.3	1,510.1	1,669.8	1,815.0	2,148.9	2,680.5
Real GDP Growth	3.1	1.1	1.9	2.5	4.4	4.2	8.6	10.7
GDP per Capita	1,768.7	2,149.5	2,257.7	2,506.3	2,771.4	3,012.3	3,443.0	4,280.0
Inflation	25.7	27.9	9.4	6.2	4.3	1.8	2.0	8.0
Unemployment Rate (end of the period)	32.7	31.5	30.4	25.8	25.4	21.1	16.8	13.6
FDI (% of GDP)	n/a	0.8	6.4	2.6	3.0	21.0	21.7	21.2
Loans (% of GDP)	10.7	9.6	9.2	13.3	16.9	20.7	50.2	93.1
Current Account Balance (% of GDP)	14.3	23.6	12.9	-6.7	-7.2	-8.5	-24.7	-39.5
Own-Source Revenues (% of GDP)	n/a	32.0	31.7	39.8	39.2	39.9	45.5	49.5
Public Spending (% of GDP)	n/a	39.5	40.9	42.1	40.6	41.5	42.3	43.3
Budget Balance (% of GDP)	n/a	-7.5	-9.2	-2.3	-1.3	-1.6	3.2	6.3
Public Debt (% of GDP)	n/a	n/a	84.5	47.1	44.5	38.6	32.6	27.5

(Continued)

TABLE 1
 BASIC MACROECONOMIC INDICATORS
 2000–2015 (Continued)

	2008	2009	2010	2011	2012	2013	2014	2015
GDP (EUR mil)	3,085.6	2,981.0	3,104.0	3,234.0	3,148.9	3,327.0	3,457.9	3,594.9
Real GDP Growth	6.9	-5.7	2.5	3.2	-2.5	3.3	2.0	3.2
GDP per Capita	4,908.0	4,720.0	5,011.0	5,211.0	5,074.0	5,356.0	5,561.0	5,781.0
Inflation	7.3	1.5	0.7	2.8	5.1	0.3	-0.5	1.0
Unemployment Rate (end of the period)	12.2	13.0	13.8	13.2	13.5	14.9	14.9	17.2
FDI (% of GDP)	18.9	35.8	17.8	12.0	14.7	9.7	10.4	11.6
Loans (% of GDP)	95.9	88.7	81.1	72.9	74.4	72.6	69.8	66.4
Current Account Balance (% of GDP)	-49.8	-27.9	-22.9	-17.7	-18.7	-14.6	-15.2	-14.5
Own-Source Revenues (% of GDP)	50.1	45.4	42.3	39.7	41.6	43.0	45.3	42.5
Public Spending (% of GDP)	50.4	51.1	47.2	43.4	47.3	47.7	48.2	47.7
Budget Balance (% of GDP)	-0.4	-5.7	-4.9	-3.7	-5.6	-5.3	-3.0	-5.3
Public Debt (% of GDP)	29.0	38.3	40.9	46.0	54.0	58.2	57.3	63.2

SOURCE: Official data and calculations made by the authors.

FDI was largely absent prior to 2005; after that, it exceeded on average 20 percent of GDP through 2009. There was also a rapid increase in the total amount of loans from less than 20 percent of GDP at the beginning of 2005 to almost 100 percent of GDP in 2008. After the sale of property (real estate and shares), which had become valuable overnight, citizens quickly received money that was predominantly used for current spending. The small and nondiversified Montenegrin economy could not produce the goods required by its citizens; therefore, a majority of goods were imported. From 2005 to 2008, imports of goods and services increased from EUR 1.1 billion to approximately EUR 3 billion. Exports did not follow the increase in imports; hence, the current account deficit in 2008 reached more than EUR 1.5 billion (almost 50 percent of GDP). The growth of bank credit was exceptional reaching 170 percent in 2007 compared to the previous year.

The large inflow of capital and imports, along with the increase in domestic consumption enhanced tax revenues, resulting in a budget surplus equal to 3.2 percent of GDP in 2006 and 6.3 percent in 2007. Growth of the economy, the regular servicing of debts and early repayment of debt to the World Bank, along with the partial write-off of debt to the Paris Club of creditors, led to a decrease in public debt from more than 80 percent of GDP in 2002 to less than 30 percent in 2008.

The years of expansion led to reductions in unemployment and the growth of employment, with substantial increases in salaries. Increases in household incomes and income from the sale of property helped reduce the poverty rate from 11.3 percent in 2006 to 4.9 percent in 2008.

In small open economies the greatest danger for stability lies in exogenous shocks. Montenegro did not fully anticipate the risk of the eurozone crisis, which led to an abrupt reduction of FDI, credit, and spending. When the cycle of prosperity within a short time turned into recession (in 2009) it was hard to accommodate the rigid budgetary expenditures, which caused a considerable fiscal deficit financed through borrowing. Montenegro's distinct characteristics were also reflected in the fact that exports were concentrated on only a few commodities (aluminum and steel) whereas imports were significantly diversified since many goods cannot be produced in a competitive manner.

The dramatic drop in aluminum and steel prices seriously impacted those industries and their workers. At the same time, due to the negative experience with the loss of bank savings in the past, citizens started to withdraw deposits. In only several months, over 20 percent of all deposits were withdrawn from Montenegro's banks. In this context, the government used the budget surpluses to help support the banking sector and the depressed industries. In the gloomy annual meetings of the IMF and World Bank in 2008, we were looking for any idea that could help prop up our failing system back home. We introduced a temporary blanket guarantee for all bank deposits and used emergency mechanisms to keep the banking system afloat. The cushion in the wake of the crisis was the large stock of household deposits and low public debt (about 30 percent of GDP) thanks to the previous policies.

The eurozone crisis showed the vulnerability of Montenegro's economy. As a consequence of the spread of illiquidity to the Montenegrin economy, and the accumulation of a large amount of nonperforming loans, new credit was practically suspended. Those few credits that were offered to prudent borrowers were subject to significantly higher interest rates and shorter terms. The debt was partially paid off, followed by the absence of new credit. The high level of nonperforming loans was addressed through specially organized factoring companies that played the role of bad banks who took over the loans. This process was supported by the foreign subsidiaries of Montenegrin banks.

The reduction of credit activities and relocation of nonperforming loans from bank accounts led to the reduction of total credits from 96 percent GDP in 2008 to less than 70 percent at the end of 2014. In the same period, the public debt grew from less than 30 percent of GDP to about 60 percent, which compensated for the drastic drop in the money supply. Thus, the combination of all credits and borrowings of the state at the end of 2014 was almost at the same level as in the precrisis year at the end of 2008 (i.e., 125 percent).

Labor and Pension Reforms

While the consequences of the eurozone crisis are still present and reflected through on-going fiscal adjustment, the reaction to the crisis also involved the implementation of structural reforms and improvement of the business environment. Structural reforms were

implemented primarily in the area of pension insurance where the system of intergenerational solidarity was reformed so that the age limit for retirement was increased to 67 for both genders, with a transitional period. Milorad Katnić had to work it out within the government as it took efforts to convince colleagues from the labor ministry and our coalition partners.

The conditions for early retirement became more rigorous and the adjustment formula was changed. The reform of the social protection system resulted from the need to increase the activity of the working-age population and to make a transfer from the gray zone to the formal sector through incentive measures.

Reforms of labor legislation were implemented with the aim of overcoming problems related to inflexibility of the labor market. Amended regulations introduced more flexible models of employment. Contracts became more widely used for regulating relations between employers and workers, and dismissal procedures became simpler and cheaper.

Business Regulation

Montenegro has made considerable progress in improving the business climate. This is indicated by relevant international ratings. According to the World Bank's *Doing Business* report, Montenegro, which was ranked 92nd during the federation with Serbia in 2005, improved its position by 46 places after the restoration of independence. In the 2015 report, Montenegro ranked 46th. The worst-ranked indicators in the report include the execution of contracts, obtaining construction permits, and registration of real estate, which implies that in these areas we need to make additional reforms in order to make our business environment one of the most attractive in the world.

The improvement of the business environment and the institutional infrastructure implied facing the ingrained problems of the Montenegrin society. Corruption is certainly one of the most perceived problems. According to Transparency International's Corruption Perception Index (CPI), since 2007 Montenegro has improved its position but is still far from satisfactory (Table 2).

The opening of Montenegro's economy with constant improvement of the business environment was also recognized through the increase of economic freedoms. According to the Heritage

TABLE 2
MONTENEGRO IN INTERNATIONAL REPORTS

International Indicator	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Doing Business (World Bank, revised rankings)	70	84	77	65	56	56	51	44	47	46
Corruption Perception Index (Transparency International)	—	84	85	69	69	66	75	67	76	61
Index of Economic Freedom (Heritage Foundation)	—	—	—	94	68	76	72	70	68	66
Economic Freedom of the World (Fraser Institute, report year)	—	60	58	81	66	37	28	49	38	62
Global Competitiveness Index (World Economic Forum)	87	82	65	62	49	60	72	67	67	70
Freedom in the World (Freedom House)	F	PF	PF	PF	F	F	F	F	F	PF
(F = free; PF = partially free)	(S&M)									

NOTE: S&M = Serbia and Montenegro.

Foundation's *Index of Economic Freedom* report, since 2009, Montenegro has improved its ranking from 94th to 66th, and according to the Fraser Institute's *Economic Freedom of the World* report in the same period, Montenegro's ranking has improved from 81st to 62nd.

In the *Global Competitiveness Report* of the World Economic Forum (WEF) for 2015–2016, Montenegro ranks 70th, which is a drop of 3 places compared to the previous year's report and a decline of 21 places relative to 2010. In order to improve competitiveness, according to the WEF report, Montenegro needs to improve its institutional environment by adopting the rule of law and establishing an independent judiciary, and also needs to strengthen macroeconomic stability and improve the physical infrastructure.

The Transition of Values

The vital support to economic transformation comes from the transition of values as it is supposed to be the true purpose of the process (Lukšić 2015). In fact, international processes have that role, helping to define new rules of the game. In this sphere, it is necessary to work hard. This is supported by opinion surveys. According to the opinion of respondents in a survey of the local NGO CEDEM (Centre for Democracy and Human Rights) in 2004, already in the early transition, freedom of speech, independence of the media and judiciary, as well as political association, were widely accepted—by 55.6 percent, 66.4 percent, 70.6 percent, and 45.7 percent of respondents, respectively (CEDEM 2004).

However, this does not mean that values such as egalitarianism were replaced by values like entrepreneurship and self-responsibility. According to the same survey, on the question of evaluating the necessity of great differences in incomes for the purpose of stepping up economic development, 24 percent of respondents “strongly agreed” or “agreed,” while 63 percent “disagreed” or “strongly disagreed.” Moreover, 84.5 percent of respondents “strongly agreed” or “agreed” that differences in incomes are too big, and 87 percent of respondents “strongly agreed” or “agreed” that the state is responsible for reducing income inequality. Moreover, according to the survey, 88 percent of respondents were certain that the government should ensure jobs for all who are willing to work, while almost 95 percent were certain the government should ensure a minimum

living standard. Although 64.7 percent of respondents “strongly agree” or “agree” that the state should protect private entrepreneurs and capital investments, as well as prevent strikes, 93.1 percent “strongly agree” or “agree” that the state should reduce differences and protect the poor and vulnerable. Such results indicate that Montenegro still harbors collectivistic values.

When we compare these findings with the results of the survey on preferences of young people carried out in 2013 by Ipsos Strategic Marketing in Montenegro for the Joint United Nations Youth Program, we understand how much the role of education in a broader sense is important. According to the survey, the aspirations of young people are the usual ones—completion of schooling, employment, and forming a family. Generally speaking, the citizens of Montenegro and the youth advocate social activism and believe it could contribute to development. They believe that only by actively looking for a job can one reach a desired result and that among people there is solidarity and readiness to help others. Still, the findings also show that there are doubts about the future (Ipsos Strategic Marketing 2013).

Regarding expectations about how to deal with poverty, four of five high school students in Montenegro plan to attend college after high school, while 70 percent of the general population and 80 percent of young people believe that further schooling can successfully prepare young people for the labor market. The survey also shows that almost 70 percent of the population (including youth) would rather accept a job outside their profession for a better salary than a job in their own profession for a smaller salary (Ipsos Strategic Marketing 2013).

Nearly 60 percent of the respondents also think that the unemployed do not get adequate support from public institutions and organizations, which indicates that there is a broad paternalistic attitude as far as the economic role of the state is concerned. Only 20 percent of the respondents plan to start their own businesses, which indicates an attitude of fear and lack of entrepreneurship, as well as timidity with regard to the value of self-responsibility. Particularly disconcerting is the fact that more than half of the youth wish to work in the public rather than the private sector. The main reason is job security. Finally, less than 10 percent of young people claim they prefer working in the private sector because of the salary (Ipsos Strategic Marketing 2013).

Conclusion

Every transition process has its specifics. Montenegro's transition is no different, with further reforms needed in several areas. However, one has to be aware that there is transition fatigue. More needs to be done in the fields of fiscal consolidation, structural reforms, and improving the business environment. The inflexible labor market not only hampers economic growth, but it also motivates people to operate in the gray economy. The banking sector has started to recover from the crisis but interest rates remain high. The public health system is still far from consolidated as it takes some important but very hard political decisions. Political reforms should continue as more transparency is needed, and the judicial system and rule of law have to strengthen as preconditions for the solidification of market reforms. Demographic change becomes a big challenge (Katnić and Krsmanović 2012), as does youth unemployment despite efforts like the new apprenticeship law passed in 2012.

However, the process of integration into the EU and NATO provides the necessary anchor. Further democratic consolidation and limits to government power by cultivating the rule of law are essential in establishing a robust political and economic system. In the long run, the biggest challenge lies in education. Instilling free-market liberal values in our children and ensuring a free market in ideas are critical for future success in transforming Montenegro into a mature market economy.

A lot has been done. What was one of the poorest republics of the former Yugoslavia today runs third, in terms of per capita income, after Slovenia and Croatia. However, we are obviously only halfway through, and one of the lessons learned is that, until consolidated, the economic liberalism agenda is still much too fragile. It is important to understand that the transition process is by no means a one-way street and that the job is done only through technocratic changes. Full transition requires the consistent pursuit of the rule of law and economic freedom.

Unfortunately, the Montenegrin society does not seem too willing to push for all the liberal/libertarian reforms. The political elite is generally left leaning and finds it hard to restore the economic liberalism agenda from the early 2000s; deep structural reforms are never popular. Politicians sometimes want to achieve the impossible by facing conflicting goals given a usually time-pressing election

horizon. The need for fiscal consolidation stands in sharp contrast with the wish to quickly develop infrastructure. The political establishment is not much helped with the current state of Europe—both political and economic. In addition, the process of European integration, given the state structure it often stimulates, may be too much of a burden for Montenegro’s small economy. We would have wished to call this article “Transition in Montenegro: The Making of an Entrepreneurial Society,” but that transition will take more time and effort.

References

- Andrijašević, Ž., and Rastoder, Š. (2006) *Istorija Crne Gore od Najstarijih Vremena do 2003 (History of Montenegro from Ancient Times until 2003)*. Podgorica: Centre for Emigrants from Montenegro.
- CEDEM (Centre for Democracy and Human Rights) (2004) *Value Orientations and Ethical Distance in Montenegro*. Podgorica: CEDEM. See www.cedem.me/en.
- Darmanović, S. (2003) “Montenegro: The Dilemmas of a Small Republic.” *Journal of Democracy* 14 (1): 145–53.
- Fraser Institute (various years) *Economic Freedom of the World*. Vancouver, B.C.: Fraser Institute, in conjunction with the Cato Institute and other think tanks. Available at www.freetheworld.com.
- Freedom House (various years) *Freedom in the World*. Available at <https://freedomhouse.org/report/freedom-world>.
- Heritage Foundation (various years) *Index of Economic Freedom*. Washington: Heritage Foundation and the *Wall Street Journal*. Available at www.heritage.org/index.
- Ipsos Strategic Marketing (2013) *KAP Survey: Knowledge, Attitudes, and Practices Survey on Youth Employment and Participation*. Belgrade: Ipsos Strategic Marketing. Available at www.un.org.me.
- ISSP (Institute for Strategic Studies and Prognoses) (2004) *Montenegro Economic Trends*, Vol. 18. Podgorica: IPPS.
- Ivanović, P., and Kuchta-Helbling, C. (2003) “Siva Ekonomija kao Barijera za Razvoj Biznisa” (“Gray Economy as a Business Barrier”). *Journal of Entrepreneurial Economy* 2 (October). Podgorica: Faculty of Economics, University of Montenegro.

- Katnić, M., and Krsmanović, A. (2012) “The Economic Consequences of Demographic Trends.” *Journal of Entrepreneurial Economy* 18 (September): 100–13.
- Lukšić, I. (2015) “On the Lost Purpose of Transition.” Available at <http://ssrn.com/abstract=2756398>.
- Transparency International (various years) *Corruption Perceptions Index*. Berlin: Transparency International.
- Vukotić, V. (2002) “Ekonomске Reforme: Manje Vidljiva Strana.” (“Economic Reforms: Less Visible Side”). Paper prepared for Miločerski Ekonomski Forum. Miločer: Association of Economists of Montenegro and Association of Economists of Serbia.
- _____ (2003) “Montenegro: Microstate.” Paper prepared for the International Conference on Economic Policy for Viable Microstate. Podgorica: Institute for Strategic Studies and Prognoses.
- _____ (2005) “Konceptijske Osnove Novog Ekonomskog Sistema u Crnoj Gori.” (“Concept Basis of the New Economic System in Montenegro”). Podgorica: Institute for Strategic Studies and Prognoses.
- Vukotić, V., and Pejovich, S. (2002) “Tranzicija i Institucije: Šta Dalje?” (“Transition and Institutions: What Next?”). Belgrade: Institute of Social Sciences.
- World Bank (various years) *Doing Business*. Washington: World Bank. Available at www.doingbusiness.org/reports/global-reports/doing-business-2016.
- World Economic Forum (2016) *The Global Competitiveness Report 2015–16*. Davos: World Economic Forum. Available at <http://reports.weforum.org/global-competitiveness-report-2015-2016>.

INTEREST ON RESERVES AND THE FED'S BALANCE SHEET

John B. Taylor

The Federal Reserve's balance sheet has expanded dramatically after three rounds of quantitative easing (QE). Consequently, the monetary base (reserves plus currency) has gone from less than \$800 billion before the financial crisis to nearly \$4 trillion today. Because reserves are a very large part of the Fed's balance sheet, I will start with the balance sheet and then consider the issue of the Fed paying interest on those reserves.

Changes in the Fed's Balance Sheet

The best way to understand what has happened to the Fed's balance sheet in recent years is to look at the actual balance sheet—the consolidated statement of assets and liabilities of all Federal Reserve Banks. Table 1 gives two snap shots of the Fed's balance sheet, one taken in 2016 and the other in 2006.

Table 1 focuses on the Fed's major assets and liabilities, lumping everything else into “other liabilities” and “other assets” categories. The two points in time—the week ending May 11, 2016, and the corresponding week ending May 10, 2006—give before and after

Cato Journal, Vol. 36, No. 3 (Fall 2016). Copyright © Cato Institute. All rights reserved.

John B. Taylor is the Mary and Robert Raymond Professor of Economics at Stanford University and George P. Shultz Senior Fellow in Economics at Stanford's Hoover Institution. From 2001 to 2005, he served as Under Secretary of Treasury for International Affairs. This article is based on his testimony before the Subcommittee on Monetary Policy and Trade, Committee on Financial Services, U.S. House of Representatives, May 17, 2016.

TABLE 1
FED'S BALANCE SHEET (BILLIONS OF DOLLARS)

May 11, 2016			
Assets		Liabilities	
Securities Held Outright	4,234	Federal Reserve Notes	1,407
Other	244	Reserve Balances	2,410
		Other	621
Total Assets	4,478	Total Liabilities	4,438
May 10, 2016			
Assets		Liabilities	
Securities Held Outright	760	Federal Reserve Notes	758
Other	82	Reserve Balances	14
		Other	41
Total Assets	842	Total Liabilities	813

SOURCE: Federal Reserve Statistical Release H.4.1, May 12, 2016, May 11, 2006, selected items.

pictures of the major changes in size and composition of the Fed's balance sheet.

It is striking how the size of the balance sheet—measured by total assets—has expanded during the last decade: from \$842 billion to \$4,478 billion. There are two major reasons for the increase. First, currency (Federal Reserve Notes) increased from \$758 billion to \$1,407 billion, an average annual growth rate of about 6 percent. There is nothing very unusual about this increase in currency; the annual growth rate was in this range in prior decades.

The second reason is much more unusual: securities held outright by the Fed jumped from \$760 billion to \$4,234 billion as the Fed engaged in three bouts of large-scale purchases of Treasury securities and mortgage-backed securities—actions commonly called “unconventional” monetary policy or QE.¹

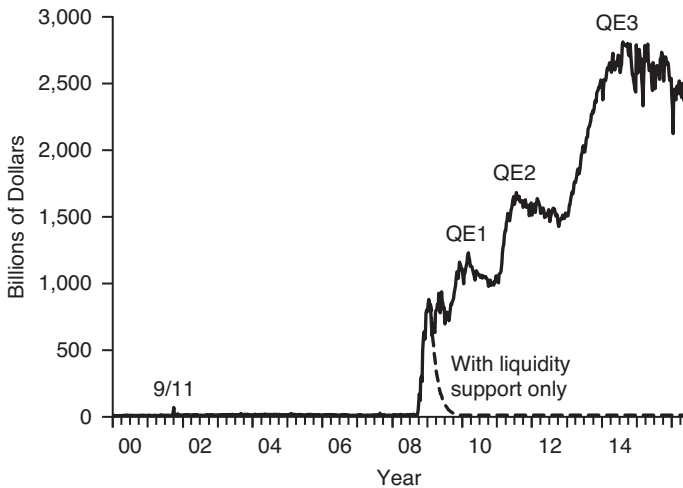
¹Another term is “credit easing” because the securities purchases or loans are in part aimed at easing credit conditions in certain sectors, such as housing. I have also used the term “mondustral policy” because such sector-specific policies are a combination of monetary policy and industrial policy (Taylor 2009).

To get the funds to purchase these securities, which increased by much more than the increase in currency, the Fed credited banks with deposits on itself, and for this reason reserve balances—the deposits that banks hold at the Fed—have exploded from only \$14 billion to \$2,410 billion as shown in Table 1. This large increase in reserve balances is very important because it is on these reserve balances that the Fed is paying interest today. Figure 1 provides some important details about the increase in bank reserves held at the Fed and illustrates how unusual that growth has been.

Reserve balances rose sharply at the times of QE1, QE2, and QE3 as the Fed ramped up its purchases of securities and financed them by creating reserve balances for the banks at the Fed, effectively borrowing the funds from banks. Reserve balances tend to drift down after each of these surges as currency creation continues its upward march and reduces the Fed’s need to create reserve balances.

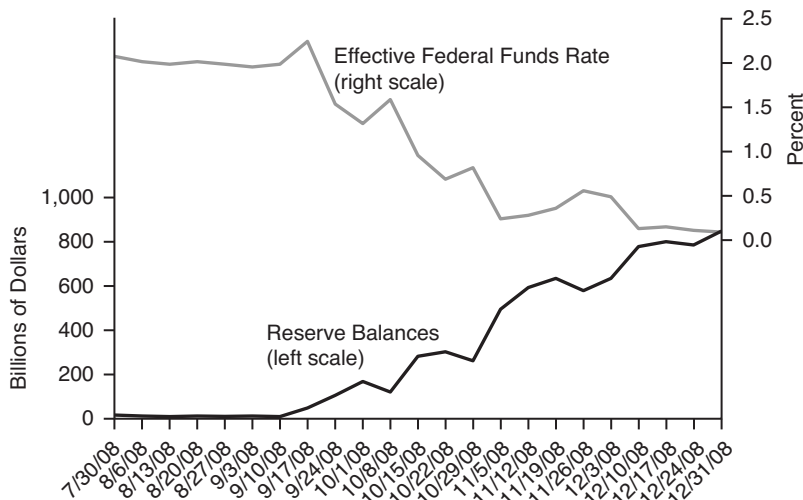
The increase in reserve balances began before the onset of quantitative easing when the Fed set up liquidity facilities to provide

FIGURE 1
RESERVE BALANCES AT THE FED



SOURCE: Federal Reserve Statistical Release, H.4.1, and author.

FIGURE 2
RESERVE BALANCES AND THE
EFFECTIVE FEDERAL FUNDS RATE, 2008



SOURCE: Federal Reserve Statistical Releases, H.4.1 and H.15.

lender of last resort loans during the panic in September 2008.² However, the need for that liquidity support was temporary, and it dissipated soon after the panic, as illustrated in Figure 1 by the dashed line for reserve balances “with liquidity support only.” That line represents a path for reserves that could have occurred if none of the bouts of QE had taken place. Clearly QE is the cause of the large amount of existing reserves.

Interest on Reserves

Such a large increase in the supply of reserves with no increase in the demand for reserves has clear implications for market interest rates: the increase in supply would be expected to drive down the federal funds rate, which is the rate banks charge each other for the overnight use of the reserves. In fact, this is exactly what happened, as shown in Figure 2 for the weeks in fall of 2008.

²The story that the policy of increasing reserves by large amounts started when the Fed’s interest rate target hit zero is incorrect. The explosion of reserves started on September 17, 2008, when the federal funds rate target was 2 percent. The Fed’s interest rate target declined from 2 percent to near 0 percent over the following months.

As the supply of reserves increased, the federal funds rate was driven down. This decline in the interest rate preceded the later decisions of the Federal Open Market Committee (FOMC) to lower the federal funds target during this period. Of course, with the supply of reserves now many times greater than demand, the market interest rate would remain near zero, unless the Fed took some other action, and this is where the policy of paying interest on reserves enters the picture. In order to raise the short-term interest rate when the supply of reserves is many times greater than demand, the Fed has to pay an interest rate on reserves to the banks that is close to the Fed's objective for the short-term interest rate. That way the banks will bid up the federal funds rate (and other short-term interest rates) as they see a profit opportunity in the difference between the federal funds rate and the interest rate on reserves. The federal funds rate will thereby move up close to the interest rate paid on reserves.

Recent events illustrate how this is supposed to work: when the Fed decided to raise the short-term interest rate by 0.25 percentage points at the FOMC meeting in December 2015, it did so by raising the interest rate it pays on reserves (required and excess) by 0.25 percentage points effective December 17, 2015. The effective daily federal funds rate promptly moved up from 0.15 percent on December 16 to 0.37 percent on December 17. Looking at monthly averages, the rate moved from 0.12 percent in November 2015 to 0.37 in April 2016. This change is consistent with the Fed's "Policy Normalization Principles and Plans" released in September 2014 stating that "During normalization, the Federal Reserve intends to move the federal funds rate into the target range set by the FOMC primarily by adjusting the interest rate it pays on excess reserve balances."³

The Financial Services Regulatory Relief Act of 2006 authorized the Fed to pay interest on required reserves (the IORR rate) and also the interest rate on excess reserves (IOER rate) as determined by the

³The federal funds rate deviated significantly below the interest rate on reserves in late 2008, and, for this and other reasons, the Fed developed backup procedures including overnight reverse repurchase agreements to help it control the federal funds rate. However, it appears thus far that the increase in interest on reserves may have been enough to move the federal funds rate as the Fed intended. In its "Policy Normalization Principles and Plans," the FOMC (2014) said that it "will use an overnight reverse repurchase agreement facility only to the extent necessary [to help control the federal funds rate] and will phase it out when it is no longer needed to help control the federal funds rate."

Board of Governors of the Fed. The original effective date was October 1, 2011, but that date was changed to October 1, 2008, by the Emergency Economic Stabilization Act of 2008, so it was available to be used for the purposes described here.

The Road Ahead: Normalization

For many years, including the period of good economic performance during the Great Moderation of the 1980s and 1990s in the United States, the Fed did not pay interest on reserves. The interest rate was determined by the supply and demand for reserves. Thus, there was a direct connection between reserves supplied by the Fed and the interest rate. The short-term interest rate was market determined once the Fed set the amount of reserves.⁴

In contrast, under the current procedures, the short-term interest rate is not market determined. Rather, it is administered by the Fed as it makes its decision about what interest rate it will pay on reserves. The interest rate on reserves can be moved around by the Fed largely independently of the supply of reserves or the size of the balance sheet. The Fed could decide to purchase securities or make loans to a certain sector and finance these by increasing reserve balances, while not moving the interest rate at all or moving it in a countervailing direction.

For the reasons explained above, such a disconnect is unavoidable during the current period of “normalization,” as the Fed calls it. The normalization period is essentially a transition period between the discretionary era of zero interest rates with quantitative easing and a normal period when the interest rate is determined in a more rule-like fashion as it was during the 1980s and 1990s.

Getting back to a normal balance sheet will require that the Fed reduce its securities holdings substantially, unless it waits the long time period required for currency growth to create a normalization

⁴Milton Friedman (1960) recommended the payment of interest on required reserves. According to the Fed’s statement “Interest on Required Balances and Excess Balances,” the interest rate on required reserves is “intended to eliminate effectively the implicit tax that reserve requirements used to impose on depository institutions,” which is in keeping with Friedman’s rationale. In contrast, the interest rate on excess reserves, according to the Fed, “gives the Federal Reserve an additional tool for the conduct of monetary policy.”

in which case the transition period will be so long it will seem permanent. In its September 2014 “Policy Normalization Principles and Plans,” the FOMC said it “intends to reduce the Federal Reserve’s securities holdings in a gradual and predictable manner,” a statement which is an apparent reaction to the taper tantrum of the previous year when the Fed was much less clear about its exit strategy. This is an improvement over previous vague statements, such as that the Fed will keep “the size of the Federal Reserve’s balance sheet at a high level for some time,” as stated in the FOMC Minutes from the January 27–28, 2009 meeting, but it could clarify the exit strategy more specifically. For example, the Fed could indicate that it will sell securities in the open market at a pace determined by the increases in the federal funds rate. I have suggested that the balance sheet be back to normal when the federal funds rate hits 2 percent. In any case, after normalization, when the Fed is back to a normal interest rate policy, the interest rate should be determined by the demand and supply of reserves in the money market—in other words, by market forces.

The Road Ahead: After Normalization

As a long-term policy, a disconnect between the short-term interest rate and the supply of reserves, the money supply, or even the size of the balance sheet is a mistake. It enables the Fed to be a multipurpose institution—helping one sector or another, taking on credit allocation, assuming fiscal policy roles the Constitution assigned to Congress—rather than the limited purpose institution it was designed to be.

If the United States is to have a selective credit policy with inherent credit risks, it is more appropriate for the Treasury or some other agency to take on the job with the approval of Congress with the purposes stated and debated. For the Fed to take on these responsibilities raises questions about its independence and its operations, as it may be called on to do such things as provide discretionary assistance to financial firms or to bolster the housing market or even the student loan market. The success of monetary policy during the Great Moderation period of long expansions and mild recessions was not due to a lot of discretion, but to following more predictable policies and guidelines.

The disconnect would be conducive to more bouts of QE. There is a great deal of uncertainty and disagreement about how effective QE has been, and it may have been counterproductive. I studied the impact of the mortgage-backed securities purchase program, which was part of QE1, in research with Johannes StroebeI. We found that the purchases were largely ineffective in changing mortgage interest rate spreads once credit and prepayment risks were taken into account (Taylor and StroebeI 2012). Others have found announcement effects of QE on long-term interest rates, but such studies cannot trace out reversals following the announcements (Gagnon et al. 2011; Krishnamurthy and Vissing-Jorgensen 2011). Finally, others have found that the effects are not lasting, unless QE signals future short-term interest rate policy and thus long-term rates through the expectations model of the term structure (Thornton 2014; Bauer and Rudebusch 2013).

A simple comparison of 1-year versus 10-year U.S. Treasury spreads does not show any impact: the spread was 1.3 percent from 2003 to 2008 before QE and 2.4 percent from 2009 to 2013 during QE, so other factors must be controlled for. At the least, there seems to be a wide consensus that the effect of QE has diminished over time. And there are other problems. QE is inherently discretionary rather than rule-like and much research indicates that this feature detracts from good economic performance (Taylor 2014). An administered rate can also distort price discovery in markets and prevent money markets from functioning normally (McKinnon 2013). There are also international ramifications as central banks tend to follow other's policies creating international impacts and currency fluctuations that can be destabilizing (Taylor 2016).

Given all these considerations, it is promising that the FOMC (2014) says in its "Policy Normalization Principles and Plans" that it "intends that the Federal Reserve will, in the longer run, hold no more securities than necessary to implement monetary policy efficiently and effectively, and that it will hold primarily Treasury securities, thereby minimizing the effect of Federal Reserve holdings on the allocation of credit across sectors of the economy." Nevertheless, more specificity about the meaning of "efficiently," "effectively," and "primarily" is warranted. In my view, a statement that in the longer run the Fed will pay interest only on required reserves and that the federal funds rate will be determined by the supply and demand for

reserves would help clarify the nature of monetary policy in the normal state following normalization.

Conclusion

This review of the Fed's balance sheet shows that the current high level of reserves is a legacy of QE in the years from 2009 to 2014. Given that the supply of reserves is now many times greater than the demand for reserves, the Fed now has no alternative other than to pay interest on reserves as it carries out its normalization process.

However, as a long-term matter, the size and composition of the balance sheet should be consistent with the interest rate being market determined rather than administratively determined by the Fed as it sets the interest rate on reserves. It is true—as the FOMC says—that paying interest on excess reserves gives the Fed an additional tool. However, this tool enables the Fed to be more like a discretionary multipurpose institution rather than the rule-like limited purpose institution that has delivered good policy in the past and that can deliver good policy in the future.

The transition, or normalization, period during which monetary policy returns to a more normal state should be as short as possible, and, in my view, shorter than currently implied by the Fed's "Policy Normalization Principles and Plans."

References

- Bauer, M. D., and Rudebusch, G. D. (2013) "The Signaling Channel for Federal Reserve Bond Purchases." Federal Reserve Bank of San Francisco, Working Paper No. 2011–21. Available at www.frbsf.org/publications/economics/papers/2011/wp11-21bk.pdf.
- FOMC (Federal Open Market Committee) (2014) "Policy Normalization Principles and Plans," as adopted effective September 16, 2014. Available at www.federalreserve.gov/monetarypolicy/files/FOMC_PolicyNormalization.pdf.
- Friedman, M. (1960) *A Program for Monetary Stability*. New York: Fordham University Press.
- Gagnon, J.; Raskin, M.; Remache, J.; Sack, B. (2011) "The Financial Market Effects of the Federal Reserve's Large-Scale Asset Purchases." *International Journal of Central Banking* 7 (1): 3–44.

- Krishnamurthy, A., and Vissing-Jorgensen, A. (2011) “The Effects of Quantitative Easing on Long-Term Interest Rates.” *Brookings Papers on Economic Activity* (Fall): 215–65.
- McKinnon, R. (2013) “The Near-Zero Interest Rate Trap.” *Wall Street Journal* (July 29).
- Taylor, J. B. (2009) “The Need to Return to a Monetary Framework.” *Business Economics* 44 (2): 63–72.
- _____ (2014) “Re-Normalize, Don’t New-Normalize Monetary Policy.” Monetary Authority of Singapore *Macroeconomic Review* 13 (October): 86–90.
- _____ (2016) “The Federal Reserve in a Globalized World Economy.” In M. Bordo and M. Wynne (eds.), *The Federal Reserve’s Role in the Global Economy: A Historical Perspective*. New York: Cambridge University Press.
- Taylor, J. B., and Stroebel, J. C. (2012) “Estimated Impact of the Federal Reserve’s Mortgage-Backed Securities Purchase Program.” *International Journal of Central Banking* 8 (2): 1–42.
- Thornton, D. L. (2014) “QE: Is There a Portfolio Balance Effect?” Federal Reserve Bank of St. Louis *Review* 96 (1): 55–72.

BOOK REVIEWS

Markets without Limits: Moral Virtues and Commercial Interests

Jason Brennan and Peter M. Jaworski
New York: Routledge, 2016, 239 pp.

Are there some things that should be beyond the market, that is, which should not be permitted to be bought and sold? Jason Brennan and Peter Jaworski think, “No, for the most part.”

Brennan and Jaworski, both philosophy professors at Georgetown University’s McDonough School of Business, dedicate *Markets without Limits* “to the authors, supporters, and readers of the *Business Ethics Journal Review*.” *BEJR* editors Chris MacDonald and Alexei Marcoux paid \$275 for this dedication. Yet, far from a frivolous exercise, this is an example of both the practical seriousness the authors bring to their premise and the good humor with which they go about making their case. The acknowledgments list a couple of dozen other well-wishers who paid for their names to be included, and who, as the authors put it, “greatly assisted us by putting their money where our mouths are.”

First, I should offer a clarification on the title. Brennan and Jaworski’s arguments in favor of the moral nature of market exchange, while very broad in terms of what may be legitimately bought and sold, do not necessarily advocate a market without any limits whatsoever. This is not, as some of their anti-market critics would have it, a tome of “market fundamentalism” that promotes absolute laissez-faire as the only, or even the most, moral economic

system. As the authors point out in the text—and have tried to remind hostile interlocutors since the book's publication—the heart of their argument is, “If you may do it for free, you may do it for money,” not that all goods and services *must* be sold without any restrictions or regulation.

The idea of responding to critics is very much at the heart of this project. In the opening chapter, the authors describe the flood of books in recent years that have attacked markets as inherently immoral or selectively corrupting in dangerous ways. The first three years of the current decade alone saw the publication of *Why Some Things Should Not Be for Sale: The Moral Limits of Markets* (2010) by Stanford's Debra Satz and *What Money Can't Buy: The Moral Limits of Markets* (2012) by Michael Sandel of Harvard. Sandel's attacks on markets, in particular, have received dramatic accolades, including a *Newsweek* contributor describing him as “possibly the world's most relevant living philosopher.”

The critics attacking market transactions, then, are not a club of marginal cranks, but represent some of the nation's most praised and frequently published public voices. Brennan and Jaworski refer to them collectively as the “anti-commodification theorists” and posit themselves as “the critics' critics.” They review several of these theorists' major objections, including concerns about exploitation, misallocation, and corruption, and proceed to examine those thematic critiques in individual chapters through examples of specific controversial products and services.

The authors do a good job of separating incidental objections from the fundamental moral questions at the heart of the anti-commodification debate. In a section titled “Business Ethics vs. What Can Be for Sale,” we are asked to consider objections to a couple of scenarios. Some critics of the fast food chain Chick-fil-A, for example, object to the company's political opposition to marriage equality. Other critics have been appalled at reports of abusive working conditions at facilities operated by Apple's Chinese contractor FoxConn. If we deem these companies' products to be worthy of a boycott, that would represent a kind of limit on the market for chicken and smartphones, but not the kind of limit that concerns the authors. Their primary question is whether there are things that categorically cannot be legitimately bought and sold. They are not concerned, in this book at least, with the objectionable actions or business practices of particular companies. To date, the anti-market

critics have not yet argued that chicken sandwiches and iPhones are inherently immoral items to sell.

That distinction, along with the understanding that Brennan and Jaworski are not necessarily arguing for unregulated markets, sets the stage for considering the anti-commodification theorists' real objections. And while ultimately they conclude that all of those objections can be answered and refuted, they do take them seriously. For example, Brennan and Jaworski engage in fascinating discussions on whether public betting on the likelihood of future terrorist attacks should be legal, as well as on more well-trod debate topics like legalizing sex work and the moral status of surrogate motherhood.

By conceding that a highly regulated market can legitimately address some people's concerns about unethical market practices, the authors are able to answer a wide range of objections to allegedly problematic market exchanges. A highly regulated market is still a market, after all, and thus market exchanges are allowed if the regulations are met. The question for them is, "When are market exchanges permitted?" not "What are the correct regulations for a permitted market exchange?" For every accusation that a certain kind of transaction might generate a negative outcome, they suggest a market structure or legal safeguard to alleviate the concern. If you're worried that the interests of the baby in a surrogacy agreement won't be adequately represented, require an independent legal advocate for the baby. If you're worried about the ethics of arranging surrogacy through a paid broker, forbid brokers. If you're worried that prospective mothers won't be adequately compensated, institute a minimum compensation package. A market would still exist under these, or even far greater, restrictions.

There are obvious practical limits to this approach, however, both from the free market and the anti-commodification perspectives. Brennan and Jaworski suggest, via a clever analogy to consumer electronics, that certain markets simply need to be "dialed in" correctly in order to be considered morally permissible; that is, they need to have their terms of payment, mode of exchange, prices, or other conditions legally bounded. Under a certain set of regulated conditions, even the most problematic market transaction can theoretically be made acceptable. But those restrictions, while they might answer the objections of certain critics, obviously can create problems of their own by restricting entry, raising prices, increasing inefficiency, and incentivizing black markets.

On the flipside, the legal acrobatics necessary to alleviate all of the anti-commodifiers' concerns may, in effect, render certain exchanges illegal under all but a handful of mostly hypothetical conditions. Why bother, such a critic might ask, creating an elaborate legal and regulatory structure to supervise a transaction with such a narrow range of morally acceptable applications? Even George Mason University's Ilya Somin, no anti-capitalist gadfly, has raised this objection to Brennan and Jaworski's argument that paying someone for his vote should be legally permissible. Somin writes, "It may well be impossible in practice to separate out the (relatively rare) cases of defensible vote-buying from the much more common ones where people are paid for their vote for the purpose of advancing the goals of some narrow interest group or voting the straight party line." The authors claim that "most of the repugnant markets could be 'fixed' rather easily" with sufficiently clever market design, but I'd like to hear from some of the anti-commodification theorists themselves on whether they consider the fixes to be quite so easy.

The authors move most decisively from unlikely hypotheticals to practical public policy when they consider the market for human kidneys. They understandably bring a more forceful case to arguing for the permissibility of markets in life-saving organs than for ones in line-standing services and household chores. They counter objections to organ sales, as they did with paid surrogacy services, with specific policy proposals. They also recognize that many opponents will remain unconvinced even when their announced objections are addressed. The authors describe this state of implacable resistance as "moral dumbfounding."

A large portion of the general public's objections are in this category—a nonrational sense of disgust at the idea of buying and selling human body parts. Contrary to some widely read theorists like Leon Kass, who coined the phrase "the wisdom of repugnance," Brennan and Jaworski suggest that visceral reactions of disgust to certain ideas are not a meaningful guide to moral judgement. They argue that semiotic objections to kidney selling—the idea that, if nothing else, it signals an inappropriate lack of reverence toward the human body—should be discarded.

Therein lies the challenge, of course. To the extent that anti-commodification arguments are really about nonrational dislike of market processes, the practical applications for this insight seem limited. In the United States, allowing a national market for kidney sales

could save thousands of lives a year. If even that hasn't been able to budge public opinion, it's difficult to see how policy advocates will be able to persuade their squeamish fellow citizens to simply "get over" their aversions. That's not Brennan and Jaworski's job, of course, but it does somewhat deflate the enjoyment of their well-reasoned arguments to realize that the task at hand is nothing less than "change American culture." I'll get right on that.

Markets without Limits probably won't, as the authors hope, "put philosophers out of the business of talking about the moral limits of markets" (especially since that seems to be one of the few commercially popular topics in academic philosophy), but hopefully it will help generate better conversations among academics and students. Clarifying the debate on the morality of markets *per se* versus the business practices of particular market actors will, on its own, go a long way towards helping capitalism's critics better frame their objections and capitalism's defenders better respond to them. Brennan and Jaworski's book both clarifies the debate over the limits of markets and shows how honest intellectuals can forthrightly and effectively respond to critics of markets.

Richard Morrison
Competitive Enterprise Institute

Zoning Rules! The Economics of Land Use Regulation

William A. Fischel

Cambridge, Mass: Lincoln Institute of Land Policy, 2015, 432 pp.

Something has gone terribly wrong in America's cities in the last few decades. Real estate construction has fallen behind demand in cities like New York, San Francisco, Austin, and Miami. Rents are rising, and both current and prospective residents are having a hard time keeping up. The good news is that we know why this is happening. Mainstream economists agree that burdensome restrictions on building new housing in prosperous cities hurt economic growth, exacerbate inequality, and stifle entrepreneurship. The bad news is that it's going to be a huge challenge to fix. This is the lesson that *Zoning Rules!* teaches us.

The book is the magnum opus of Dartmouth economist William Fischel, the follow-up to his acclaimed 1987 book *The Economics of Zoning Laws*. In the nearly 30 years between these tomes, Fischel

spent his career exploring how land use rules came to be and what role they would play in the economy and society. Through careful history, a thorough understanding of property law, and sound economics, *Zoning Rules!* weaves a story of the rise of the exclusionary zoning laws that have come to strangle housing development in major metropolitan areas.

The early years of zoning laws were relatively harmless and mundane compared to what would come later. Urban planners and the officials they appointed generally made pragmatic decisions that reflected the opinions of their constituents. When excluded from one town, developers could find another nearby. Minority factions had trouble blocking development if they lacked political power. It was the era of “good housekeeping” zoning. Fischel defines “good housekeeping” as basic separation of activities, such as noxious industry from residential areas, without attempts to control or micromanage growth. The period saw a boom in suburban building that would taper off over time.

Fischel’s theory is that something changed in the 1970s. A confluence of events precipitated the rise of exclusionary laws in the suburbs of the nation’s cities. Employment decentralized from areas near ports with the advent of trucking, the interstate system, and the shipping container. Freed from center cities, industry fled to places where land was cheap and plentiful. The period saw housing grow to take up a greater share of household wealth. Combined with the civil rights movement’s reforms, the period saw a serious increase in demand for exclusionary housing restrictions. Yet, demand is only part of the story. The supply pressures of exclusionary laws were serious. Unlike labor or capital, moving land from one municipality to another, while not impossible because of processes like municipal annexation, is a serious challenge. With ability to relocate to places with more hospitable legal institutions, weak zoning laws play a key role in the land use decisionmaking process.

The period also saw the federalization of the environmental movement combine with the expansion of legal standing of development opponents and the advent of multilayered review by state government. Different departments and government bodies were given power to oversee various aspects of building. Development was bureaucratized, and with it, the transaction costs of the building process increased. The power dynamics of local government

underwent a fundamental change. What was once a regime of “good housekeeping” laws was corrupted.

More than anything, *Zoning Rules!* is a textbook of land use economics. Fischel is no pundit. Indeed, his grand theory isn’t presented until the middle of a chapter in the middle of the book. Like its predecessor, *Zoning Rules!* is written to both explore and explain. It begins with a chapter laying out basic concepts of land economics, the basics of supply and demand for housing. The second chapter explains how zoning laws work and how they are administered. He takes particular issue with the rise of so-called “conservation easements,” agreements for developers to provide part of their land as green space in perpetuity in exchange for regulatory concessions. Such easements make adapting to changing local conditions in the unknown future more difficult, and locking up ever more land for greenery makes little sense, Fischel argues.

Fischel’s third chapter provides a thorough examination of how various court rulings from the 1920s onward have shaped the legal landscape of land use regulation, starting with the famed case *Euclid v. Ambler*. In *Euclid*, the Supreme Court found zoning laws to be a legitimate use of the police power so long as an action was in promotion of health, safety, morals, or general welfare. Fischel makes clear that zoning was and is a popular, organic institution that arose to address real citizen concerns about locating noxious, often-industrial uses near homes. When courts in New Jersey deemed zoning unlawful, the state constitution was changed with overwhelming citizen support. “We have zoning because a large majority of voters almost everywhere want it,” notes the professor.

The legal history of land use regulation has seen many attempts to attack exclusionary aspects of zoning. In the mid-1970s, cases such as *Warth v. Seldin* and *Arlington Heights v. Metropolitan Housing Corp.* precluded claims under the Equal Protection Clause of the 14th Amendment. The chapter goes on to explore the details of the legal evolution of the Takings Clause. The chapter gives an overall impression that the Supreme Court’s land use rulings are generally pragmatic. The courts have generally avoided rulings that would give the judiciary a central role in the land use dispute adjudication process. Rulings have given cities wide latitude to make land use decisions as they see fit, so long as the rules are not particularly egregious, such as attempts to institute racial zoning or take physical property.

Fischel then lays out his theory of how zoning transformed from a generally benign, popular, functional system of “good housekeeping” regulation to one captured by special interests. Of particular note is the rise of new layers of project review that give a political voice to those not directly affected by a particular development. Today, too many people have a say in what gets built where. Minority factions have the power to drag out the approval process or disapprove projects outright. What were once local matters between a developer, a municipality, and potential neighbors have seen the advent of layers of review by environmental agencies, transportation agencies, community groups, and others. The balance of power in such negotiations has shifted away from developers and toward special interests. Unlike in the era before the 1970s, today the permitting process can take years, with ample opportunities for special interests to demand concessions for their cooperation. Developers must either accept the demands, negotiate, or risk the fate of their project. Today, large developments often include a suite of community amenities from the beginning of the design process—typically things like public parks or plaza spaces, artwork, pedestrian or bicycle paths, transportation improvements, or support for community groups—in an effort to stem calls for even greater concessions.

Fischel has spent his career arguing that housing prices incorporate all positives and negatives about living in a particular place. The price of housing is not simply the price of the house, but the aggregate of the positives and negatives of living in that particular place. In a sense, housing prices approximate the bundle of club goods available to those who live there. Prices are increased by amenities like good schools, nearby parks, access to jobs, well-run city services, low traffic, and other such factors. Prices are decreased by disamenities like crime, proximate blighted properties, weak government services, high taxes, and other things that might dissuade a prospective resident.

Moreover, ease of building can be an amenity, and difficulty of building can be a disamenity. According to Fischel, the relative challenge of building will be incorporated in real estate prices because homeowners value their right to build on their property or be able to sell to a party who would do so. In this sense, the ease of building is just like any other amenity or disamenity that can increase or decrease housing prices.

A major component in the value of a property is the implied value of changing its use. Current land use restrictions are incorporated into property values, albeit indirectly. When rules disallow most redevelopment, such as by setting maximum occupancy limits, minimum lot sizes, or restrictions on what activities may take place on the property, the property's value is lower than it otherwise would be. The rules are counterbalanced by the fact that they increase property values of homes across the area by limiting competition for prospective buyers. Overall, homeowners benefit from decreased competition in the local housing market—even if an individual homeowner loses the opportunity for greater personal benefits. Existing homeowners have every reason to organize in favor of growth-limiting laws, such as greater lot sizes, more parking, and fewer multifamily homes. The process is exacerbated by the fact that, in most places in the United States, school attendance is tied to geography.

Homeowners form the most vocal faction in local government, argues Fischel, and that's why growth controls have overgrown their original purpose. The link between school assignments and real estate throws gasoline on this fire. The benign, "good housekeeping" zoning of the past has given way to the era where cartels of homeowners wield market power to depress competition and capture rents. The age of the "homevoter" is upon us. Fischel explained this theory of local government captured by activist, cartelized homeowners in his 2002 book, *The Homevoter Hypothesis. Zoning Rules!* follows up on the past book with reflections on legal developments in the decade after. For instance, the first 15 years of the 21st century included the famed Takings case *Kelo v. New London*, when the Supreme Court relaxed the requirements for a "public use" taking, and the backlash it created.

Fischel deserves credit for being honest about circumstances where he changed his mind on issues and for his humility about how challenging reform will be. Nearly three decades of history stand between *The Economics of Zoning Laws* and *Zoning Rules!* Fischel spent part of that time as a zoning practitioner, sitting on the zoning board of Hanover, New Hampshire. While he correctly understood the perverse political economy of land use from the outset, Fischel makes clear that he was overly optimistic that action by the judiciary branch would stem the homevoter tide. Unlike many scholars, he understands that his desired reforms would be unpopular for a reason, and that homevoter interests truly are a

great challenge to overcome. Toward the end of the book Fischel admits that even he, a leading zoning scholar, sitting on the board of a town of only 10,000 people, struggled to bring any substantial reform to the town's zoning laws.

While generally a sound analysis, Fischel's argument does have opponents. He advocates for a piecemeal approach to zoning, allowing a great deal of leeway for developers to negotiate the regulation of a parcel with both the municipality and potential neighbors. Some, most famously Yale professor David Schleicher in his article "City Unplanning," have argued for a more holistic approach to the land use regulation process. Schleicher argues that a strong city plan, combined with insulation of zoning institutions from political pressure and special interests could be enough to stem the shortage of housing in desirable cities. These plans seem fragile to Fischel. Urban planning has proven throughout its history to be prone to interest group capture. He argued, in *Regulation* magazine and elsewhere, that any new, stronger planning institutions may, too, fall victim and could make matters worse.

There are also questions about one of Fischel's fundamental insights. Fischel pins the anti-development impulse among homeowners on the uninsured risk that development could cause property values to fall. His theory is strong for most locations, but might not be sufficient everywhere. In desirable places, homeowners stand to make windfall profits from denser zoning when they sell their properties to developers, yet, instead, they advocate for decreased density and other disamenities. This is the manifestation of an exclusionary impulse to signal to prospective residents who is welcome in a particular place and who is not. Neighbors are attempting to curate the kind of people with whom they may be forced to interact. In many cases the rich may attempt to exclude the poor. Many college towns attempt to use building occupancy maximums to sequester students to particular areas by preventing the subdivision of single family homes into group houses. Fears of gentrification can lead to attempts to prevent redevelopment for the wealthy to prevent displacement of poor neighborhood residents. The list goes on. Importantly, these rules shift costs of zoning laws from developers and neighbors to disenfranchised potential neighbors. Fischel's theory, while valuable, does not do enough to show how this impulse to curate the makeup of one's neighborhood plays into the political economy of land use.

Like *The Economics of Zoning Laws*, *Zoning Rules!* is one of the most important books on land use economics in a generation. The book came out at an auspicious time, as many are beginning to turn against zoning laws. In 2015, intellectuals on the left and right, academics and policymakers alike, have realized the damage high housing costs in prosperous places is doing. Fischel's theory of homeowner capture of local politics elegantly explains how the nation arrived at this unhealthy equilibrium.

People may disagree with the author over whether zoning is the best way to handle the problems that arise from coordinating land use decisions, but it is clear that his theory is based on sound economic history and legal analysis. *Zoning Rules!* is not a book of punditry, out to push a theory. It is a book that seeks to explain how land use decisionmaking works. At that, it succeeds better than any of its peers. If you want to read a single book about land use regulation, *Zoning Rules!* should be that book.

Nick Zaiac
Mercatus Center
George Mason University

The Unquiet Frontier: Rising Rivals, Vulnerable Allies, and the Crisis of American Power

Jakub J. Grygiel and A. Wess Mitchell

Princeton, N.J.: Princeton University Press, 2016, 240 pp.

In *The Unquiet Frontier*, Jakub Grygiel and A. Wess Mitchell have articulated a provocative justification for a revitalized strategy of containment focused on China, Iran, and Russia. That strategy is based on what they call the “rimland imperative”—the notion that U.S. security and prosperity are vitally dependent upon supporting allies against the encroachment by “revisionist” states (those intent upon overturning the established geopolitical order) on the periphery of Eurasia, stretching “from the Baltic Sea to the Black Sea in Europe, through the Levant and Persian Gulf to the Indian Ocean and up through the littoral Asia to the Sea of Japan.”

In many ways, *The Unquiet Frontier* is unduly alarmist. The contention that “the U.S.-led global alliance network could unravel in coming years” is certainly hyperbolic. Over the past decade, a number of prominent analysts have voiced trenchant arguments in

favor of U.S. retrenchment. But that perspective has failed to exert much (if any) influence over U.S. foreign policy. In fact, the wars in Iraq and Afghanistan have prompted the Obama administration to place renewed emphasis on maintaining strong alliances—largely as a means of distributing the burden for addressing threats to international security. In his FY2017 budget proposal, President Obama recommended increasing funding for the European Reassurance Initiative by nearly 350 percent to \$3.4 billion. Reports of the death of the American alliance system have thus been greatly exaggerated.

Nevertheless, Grygiel and Mitchell provide a thorough and interesting explication of the general benefits of alliances. They can deter revisionist states by increasing the expected costs and reducing the expected gains of aggression—and in so doing temper the ambitions of U.S. rivals. In the event that deterrence fails, alliances also extend the reach of the U.S. military by securing access to overseas bases. And by reassuring smaller states, U.S. alliances can prevent them from pursuing independent security initiatives that could be destabilizing—most notably, developing their own nuclear weapons. Given those benefits, Grygiel and Mitchell view alliances as essential for preventing “the emergence of a power or combination of powers within the Eurasian landmass that could invade or economically dominate the United States.”

In promoting a strategy of forward containment, Grygiel and Mitchell lean heavily on classical geostrategic theory. They frequently invoke the arguments of Halford Mackinder and Nicholas Spykman stressing the importance of the global rimland. Moreover, their advocacy of the vital importance of “frontier allies” is cut from the same cloth as the perimeter defense concept inherent in U.S. containment doctrine from the early Cold War period. Any failure to resist revisionist probing, which Grygiel and Mitchell define as any “low-intensity and low-risk test aimed at gauging the opposing state’s power and will to maintain security and influence over a region,” is purported to encourage further aggression, which permits the growth of menacing Eurasian rivals to the United States.

Even if one acknowledges the general benefits of alliances, the contention that small states on the periphery of the Eurasian rimland are vital to U.S. security is unpersuasive. It is difficult to see how picking off small frontier states, through either aggression or accommodation, would substantially augment the power of a revisionist state.

In fact, such states often prove to be liabilities rather than assets. The suggestion that Estonia and Taiwan are vital to U.S. security and prosperity thus borders on the absurd.

Grygiel and Mitchell do acknowledge, however, that the United States cannot possibly defend every far-flung outpost in the world. Since “to defend everything is to defend nothing,” they argue that “America’s strategic goal should not be to defend some abstract global architecture or global principle but to defend specific states against specific threats.” This acknowledgment renders the advocacy of a perimeter defense strategy even more questionable. The inherent limits of American power would seem to suggest that a “strong-point” containment strategy, which analysts such as George Kennan gradually gravitated toward in the early Cold War period, would be more pragmatic. Rather than trying to contain Russia, Iran, and China along a 10,000-mile rimland, the United States could focus on deterring aggression against key modern industrial states—and perhaps less developed states in key strategic locations—whose absorption would substantially augment the material power of revisionist rivals. In other words, the United States should focus on doing what it has been doing: preserving a strong, independent Western Europe; deterring aggression against either Japan or South Korea; and working with Turkey and Saudi Arabia to maintain access to the straits of the Bosphorus, Dardanelles, and Hormuz.

A more limited containment strategy might also be preferable since it would be less likely to antagonize China, Iran, and Russia. That represents the largest hole in *The Unquiet Frontier*: Grygiel and Mitchell devote almost no attention to the security dilemma—the idea that actions one state takes to enhance its own security can perversely engender greater insecurity in other states. Yet recurring rounds of NATO expansion have clearly engendered fears of isolation and encirclement in Russia. Given those fears, Russia’s intervention in Ukraine should not necessarily be interpreted as a revisionist probe; in many ways, it constitutes a desperate attempt to preserve the status quo by preventing Ukraine from being absorbed into western political and military alliances. It is therefore imperative to consider whether maintaining military alliances on the doorsteps of countries like China, Iran, and Russia might do more harm than good by confirming their suspicions that the United States is intent upon encircling them.

Ultimately, *The Unquiet Frontier* is symptomatic of a pervasive fear that a new Cold War may be dawning. Yet the book makes one

wonder whether adopting a rimland containment strategy would likely turn those fears into a self-fulfilling prophecy. As George Kennan, the so-called father of containment, wrote in 1947, “It is an undeniable privilege of every man to prove himself right in the thesis that the world is his enemy; for if he reiterates it frequently enough and makes it the background of his conduct he is bound eventually to be right.”

Brad Stapleton
Cato Institute

Side Effects and Complications: The Economic Consequences of Health-Care Reform

Casey B. Mulligan

Chicago: University of Chicago Press, 2015, 352 pp.

It is a cruel fact of history, but for seven decades and counting, the U.S. government has joined the market for health insurance in unholy matrimony with the market for labor.

It’s cruel to workers. Separation from a job, for whatever reason, means, at a minimum, disruption of one’s health insurance coverage, and often disruption of one’s access to care. Rather than encourage coverage that would stay with workers after they retired, for example, the shotgun marriage of these two markets casts millions of workers out of their health plans the moment they reached retirement age—many of them with suddenly uninsurable preexisting conditions.

It’s cruel to taxpayers, because its impact on retirees fueled the creation of the incredibly expensive and wasteful Medicare program.

It’s cruel to economists, who, if they seek to understand and improve the functioning of one market, must become experts on two.

It’s cruel to policymakers, in that it fosters a misleading picture of trends in worker compensation, along with a network of tripwires and unintended consequences that stymie sensible reform.

And, finally, it’s doubly cruel to workers, because it allows policymakers to hide the cost of insensible reforms in forgone wages—as Congress quite consciously did under the Affordable Care Act of 2010, not to mention previous and equally dubious “affordable-care acts.”

This shotgun marriage could not have survived without some beneficiaries, notably: the health sector, on which it bestows large

implicit subsidies; large employers, to whom it grants a competitive advantage over smaller competitors; and policymakers, on whom it bestows greater importance and a reason to legislate.

And legislate they have. Having conferred an enormous tax preference on a lousy insurance product, Congress is continually struggling to fill in cracks in the health sector that Congress itself created or widened. Medicare, the largest purchaser of medical care in the world, exists to fill just *one* such gap.

Congress also regulates everything from the content to the pricing to the timing of employer-sponsored health insurance. Where not preempted by federal law, states have regulated even further. With the exceptions of Medicare and Medicaid, however, no insensible reform surpasses the Affordable Care Act (ACA), known colloquially as ObamaCare, at bringing together all of the above-mentioned cruelties.

The ACA “builds upon” the employer-based system. That is to say, it leaves in place an insensible tax exclusion for employer-sponsored insurance and creates health insurance subsidies for taxpayers with low or moderate incomes who do not have access to employer-sponsored coverage. Those with very low incomes can, in willing states, enroll in an expanded Medicaid program. Those with incomes between one and four times the federal poverty level (\$24,300 to \$97,200 for a family of four) may receive health insurance subsidies for nominally private coverage purchased through a pseudo-market called an “Exchange.”

These seemingly simple changes set off a chain reaction of perverse incentives and cascading cruelties. To the extent workers would benefit more from the new subsidies than from the old tax exclusion, the ACA creates incentives for employers not to offer coverage and for workers to gravitate to such employers. Since the new subsidies increase as income falls, they create incentives for workers to work less or not at all. Supporters eagerly note the new subsidies can mitigate “job lock” by freeing workers to take a different or lower-paying job because they no longer have to work for the health benefits. It is a curious selling point that Congress replaced the incentives it created to work suboptimally with incentives not to work at all.

To mitigate some of *those* perverse incentives, the ACA penalizes employers with more than 50 employees who fail to offer a minimum level of coverage—which, in turn, creates further perverse incentives for small employers to remain below the 50-employee threshold that

triggers penalties and incentives for large employers to reorganize and/or lay off workers so they fall below it.

Robbing Peter to subsidize Paul always creates incentives for both parties to work less. But the ACA achieves such redistribution in a manner so convoluted, with so many hidden subsidies and penalties, layered atop so much preexisting complexity at this intersection of the health care and labor markets that it might be a Herculean task just to quantify the magnitudes of those incentives, much less to predict how workers and employers will respond. Thankfully, Casey Mulligan has waded into this morass so you won't have to.

A labor economist at the University of Chicago, Professor Mulligan approaches these questions with all the precision of an economist and more. He also brings the dispassion of a true economist, treating the creation of penalties and the withdrawal of subsidies as economically identical—a lesson conservatives would do well to learn.

Motivating Mulligan to think systematically about the ACA's complexity is its enormous potential impact on labor markets and economic performance. He concludes that, when one includes the impact of the ACA's implicit taxes, the law's impact on economic output is "vastly more important than, say, the interest rate on federal funds."

You might think that all economists who attempt to quantify the ACA's effect on labor markets would take its implicit taxes into account. But you'd be wrong. Mulligan notes several studies that fail to do so. He pays particular attention to a 2010 study by economists David Cutler and Neeraj Sood that claimed the ACA would boost employment by up to 400,000 jobs, an estimate later endorsed by nearly 300 economists. In addition to not incorporating the effect of the law's implicit taxes, Cutler and Sood ignored the potential administrative costs employers would face. In practice, those costs alone have left some employers wanting to repeal the employer mandate or even the entire law. "It's not because they don't want to offer coverage," explains the author of one employer survey. "It's because proving that they offer coverage is so much work."

On average, Mulligan finds, the ACA adds 6 percentage points to the average implicit marginal tax rate workers face throughout the economy. That may not sound like much, but it creates a larger disincentive to work than any other piece of legislation Congress

enacted over the prior 70 years. In some cases, the ACA subjects “Paul” to implicit marginal tax rates that exceed 100 percent. Subsidies phase out so rapidly as income changes that many workers will find that working more reduces their income while working less increases it.

All workers will bear the cost of these provisions, Mulligan shows, even those whom these provisions do not touch directly. Even in firms that comply with the ACA’s employer mandate, that mandate will cause wages for high-skilled workers to fall by 1.3 percent and will depress wages for low-income workers by 3 percent.

Broadly speaking, “The ACA will have the nation working fewer hours, and working those hours less productively, so that its non-health spending will be twice diminished: once to pay for more health care and a second time because the economy is smaller and less productive. . . . I predict that the ACA’s impacts—that is, the difference between the economy with the ACA and a hypothetical and otherwise similar economy without the ACA—will include about 3 percent less employment, 3 percent fewer aggregate work hours, 2 percent less GDP, and 2 percent less labor income.” Mulligan further projects some “5 million workers plus roughly 5 million dependents will work part-time schedules as a consequence of the ACA,” there will be 19 million fewer uninsured, and the number of people with employer-sponsored insurance will fall by 13 million.

Mulligan cautions he is merely offering projections, calculated before it was possible to collect actual data on the ACA’s impact on labor markets. He offers them as a benchmark against which we can compare actual experience under the law. One can sense his eagerness—excitement, even—to learn what he got wrong.

With a statute so large and complex, of course, error is practically inevitable. Mulligan claims the IRS may use liens to collect unpaid individual-mandate penalties. In fact, the ACA specifically prohibits it. (The error belongs more to the Congressional Research Service, Mulligan’s source for this claim.)

Avoiding error becomes impossible when the executive branch continually and quietly rewrites such a complex statute on the fly. The ACA’s premium subsidies impose an implicit tax on work. Mulligan quantifies that implicit tax on such a granular level that he incorporates what happens when workers receive more premium subsidy than the law allows and then must repay a statutorily

determined portion of the unauthorized portion of their subsidy at tax time.

Yet Mulligan does not incorporate a twist on that reconciliation process that the IRS developed with neither fanfare nor statutory support. If subsidy recipients later turn out to be totally ineligible for a subsidy because their income was too low (i.e., below the poverty level), Mulligan assumes they will repay the entire subsidy, as the ACA requires. The IRS, however, has deemed such taxpayers will not have to repay a cent. In effect, since subsidies rise as *actual* (as opposed to projected) income falls, the *largest* subsidies therefore go to taxpayers who have incomes below the poverty level—even though they are statutorily ineligible. This administrative rewrite of the ACA both expands the ACA's implicit tax on work (by creating a disincentive to earn more than the poverty level) and imposes an implicit tax on honesty (by penalizing those who accurately project their income will fall below the poverty level). These implicit taxes must have at least some effect on labor markets. If the change escaped Mulligan's notice, perhaps it was because authors of the regulation were not eager to draw attention to their handiwork.

How are Mulligan's projections holding up so far? Some ACA supporters claim that, aside from a reduction in the number of uninsured, there is no evidence the ACA is having the effects Mulligan predicts. The responsible ones note that it is difficult to isolate the ACA's effects, given that it was enacted at the nadir of the Great Recession, that anticipation and implementation of its provisions coincided with the recovery, and that administrative and congressional action have delayed implementation of many of its taxes on labor (the employer mandate, the Cadillac tax). There is ample evidence that, at least beneath the aggregate figures, employers and workers are responding to the ACA's implicit taxes on labor (see, e.g., Michael F. Cannon, "Obamacare Is Destroying Jobs—and Here's the Evidence," *Forbes.com*, February 4, 2016). Only time and careful economic analysis will tell.

As one of its architects infamously admitted, the ACA never could have become law if voters understood what it actually does or saw all the taxes it imposes. *Side Effects and Complications* brings transparency to a law whose authors designed it to be opaque. It is a one-of-its-kind inquiry into *all* of the ACA's effects on jobs,

incomes, and health insurance coverage. One hopes it will not enjoy that distinction for long. Future analyses will have to take into account not only Mulligan's projections, but more important his methodology.

Michael F. Cannon
Cato Institute

The Law of the Land: A Grand Tour of Our Constitutional Republic

Akhil Reed Amar

New York: Basic Books, 2015, 376 pp.

Akhil Reed Amar's *The Law of The Land: A Grand Tour of Our Constitutional Republic* seeks to take the reader on a "grand tour" of the various regions of the American Republic and define their contribution to constitutionalism in general. The object was to explicate the "différance," as Derrida might say, between Amar's identified 12 distinct cultural regions and to tie that uniqueness into the present tapestry of our constitutional fabric.

This book is a bit of a mixture. On the one hand, it does a fine job exploring constitutional history of various clauses (such as the Second Amendment qua Wisconsin constitutionalism) and individuals (for example, Justice Robert Jackson). Yet, on the other, it doesn't really provide a regionalized "tour" of the constitutional republic. Instead, it attempts to shoehorn a book on significant moments and aspects of constitutional history into a book on regional constitutionalism.

The author somewhat readily admits this flaw in the conclusion, giving a mea culpa for focusing on 12 constitutional instances rather than performing a 50-state survey. But the flaw is a bit deeper. Consider the chapter on Justice Hugo Black.

Amar's premise is that certain constitutional figures or moments are emblematic of a region's contribution to our constitutional fabric. But his discussion of Hugo Back is almost entirely divorced from Alabama and the larger region of the Deep South. Aside from some references to Justice Black's being from Alabama and ruling on cases that came out of Alabama, there is nothing to tie the Deep South to Justice Black's legacy of textual originalism and total incorporation of the Bill of Rights to the states. Indeed, the chapter—while a

fascinating discussion of Justice Black's legacy—does not illustrate anything in particular about Deep South constitutionalism.

Amar's discussion of Kansas's contribution to American constitutionalism through the lens of *Brown v. Board of Education* suffers from the same flaw. While Amar stresses the town of "Topeka" several times in the chapter, he does not at any point describe how Topeka is unique in its contribution to the legacy of American apartheid—and Amar readily admits that segregation is not unique to Kansas or even the Outer Southern Region (how Amar describes the belt around the Deep South that contains Kansas). "Topeka" serves as more of a talismanic incantation than a fixed point in analyzing regional constitutionalism. Thus, the reader is left thinking little about Kansas or the Outer South as a unique region. The most unique part of the chapter is when Amar gives his personal justification of *Brown* as an originalist case based on some interpretative gymnastics around the Titles of Nobility Clause.

To be sure, this flaw is not seen everywhere in the book. Amar makes a much more intricate biographical discussion of Abraham Lincoln and how notions of constitutionalism in Illinois and the Old Northwest shaped Lincoln's understanding of the Republic's constitution and federal supremacy in particular. Moreover, the argument is made much more convincing through an analysis of Lincoln's inaugural address. Nevertheless, readers should take the book with a grain of salt: it is a solid piece of work on important aspects of constitutional history, but the focus on regionalism is a bit misplaced in some chapters.

One historical point in Amar's conclusion in particular bears particular relation to recent events—the death of Justice Antonin Scalia. While the book was written well before the death of the late justice, its brief historical discussion of circuit-riding seems to bear on the present controversy over Senate consent to Supreme Court nominees.

Amar points out that, until the late 1800s, the seats on the Supreme Court were thought of as the Southern seat, the Northern seat, and so on—and for good reason. Justices, as they are now, were assigned a particular circuit, yet, unlike now, they had to visit and hear cases in that circuit. Back then, transportation was so cumbersome as to require the appointment of a justice who *could* ride a particular circuit, necessitating a regional division of justices rather than a political or ideological one per se, regional interests notwithstanding.

The qualifications for justices, at least informally, once included special familiarity with the law of the circuit and presence within the circuit (to facilitate riding). Nowadays, justices are not selected based on their regionalism but are selected based on their judicial philosophy, which inevitably is politicized during nominations.

Indeed, the current controversy over Justice Scalia's seat readily exemplifies this shift from regionalism to ideology. The concern over filling the seat is not that a particular region will not be represented but that the conservative viewpoint on the court will be repressed after nearly 40 years of conservative control of the Court. The Scalia battle is over the nationalization of the judiciary and the shift to a focus on pure ideology and balance, rather than on the nomination process being shaped by practical and regional considerations.

All in all, Amar's book is a good read. It underdelivers on the promise of a grand tour of our constitutional republic, but it is a fine survey of key points of constitutional history. It also provides an historical window into modern conflicts, such as the battle over Scalia's seat.

Randal J. Meyer
Cato Institute

WEALTH, POVERTY
AND POLITICS

REVISED AND ENLARGED EDITION

THOMAS SOWELL

Wealth, Poverty and Politics is a new approach to understanding age-old issues about economic disparities among nations and within nations. These disparities are examined in the light of history, economics, geography, demography and culture.

Wealth, Poverty and Politics is also a challenge to much that is being said today about income distribution and wealth concentration— a challenge to the underlying assumptions and to the ambiguous words and misleading statistics in which those assumptions are embedded, often even by leading economists. This includes statistics about the much-discussed “top one percent.”

This revised and enlarged edition should be especially valuable to those who teach, and who want to confront their students with more than one way of looking at issues that are too important to be settled by whatever the prevailing orthodoxy happens to be.

A true gem in terms of exposing the demagoguery and sheer ignorance of politicians and intellectuals in their claims about wealth and poverty... Dr. Sowell's new book tosses a monkey wrench into most of the things said about income by politicians, intellectuals and assorted hustlers, plus it's a fun read. (Professor Walter E. Williams, George Mason University)

At a time when many politicians, academics and media commentators are focusing on income inequality, Thomas Sowell's *Wealth, Poverty and Politics: An International Perspective* offers a refreshing and stimulating view. (Professor John B. Taylor, Stanford University)

Sowell... draws from this well of research to do what he has done so well for so long: question basic assumptions behind public policy and follow the facts where they lead him. (Jason Riley, *Wall Street Journal*)

FOR COMPLIMENTARY EXAMINATION COPIES OR DESK COPIES, PLEASE CONTACT perseusacademic.com.

The Journal of
LAW & ECONOMICS

Insider Trading in Supervised Industries
David M. Reeb, Yuzhao Zhang, and Wanyi Zhao

*The Distribution of Legal Traditions around the World:
A Contribution to the Legal-Origins Theory*
Daniel Oto-Peralías and Diego Romero-Ávila

Valuation of Patented Product Features
Greg M. Allenby, Jeff Braxell, John R. Howell, and Peter E. Rossi

The Role and Growth of New-Car Leasing: Theory and Evidence
Justin P. Johnson, Henry S. Schneider, and Michael Waldman

Trial and Settlement: A Study of High-Low Agreements
J.J. Prescott, Kathryn E. Spier, and Albert Yoon

Litigation Risk and Agency Costs: Evidence from Nevada Corporate Law
Dain C. Donelson and Christopher G. Yost

Rational Inattention and Energy Efficiency
James M. Sullivan

*Institutional Quality, Culture, and Norms of Cooperation:
Evidence from Behavioral Field Experiments*
Alessandra Casari, Giovanna d'Adda, and Pauline Grosjean

Published for The Booth School of Business of the University of Chicago
and The University of Chicago Law School

THE UNIVERSITY OF CHICAGO PRESS

*Widely cited
economic
analysis of
regulation,
legislation, law,
and finance*

*The Journal of Law and
Economics*

*Editors: Dennis W. Carlton, John P. Gould,
Richard Holden, Anup Malani, and Sam
Peltzman*

*Society: The University of Chicago Booth School
of Business and the University of Chicago Law
School*

Visit journals.uchicago.edu/jle to
sample content, subscribe, and contribute.

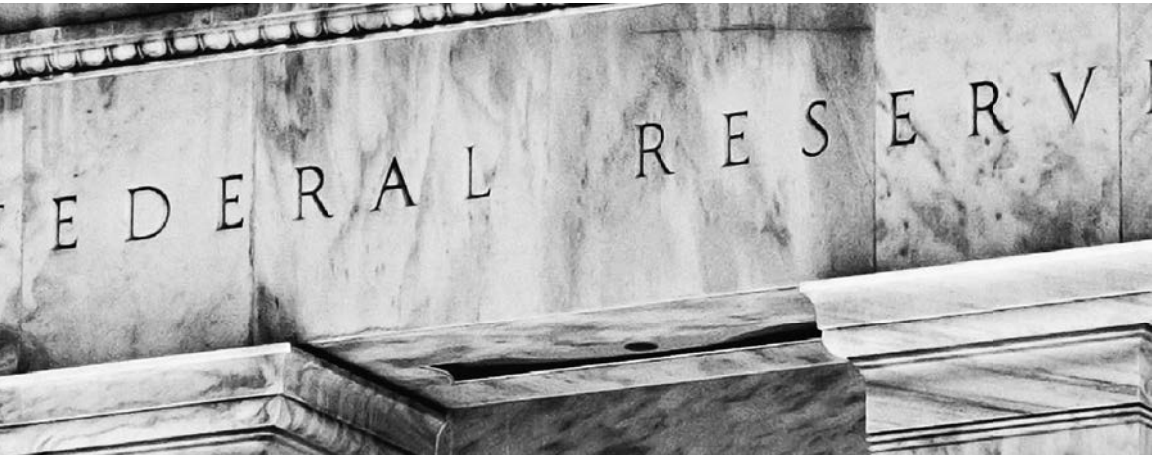


THE UNIVERSITY OF CHICAGO PRESS JOURNALS

THE CATO INSTITUTE'S CENTER FOR MONETARY AND FINANCIAL ALTERNATIVES
INVITES YOU TO THE 34TH ANNUAL MONETARY CONFERENCE

Central Banks and Financial Turmoil

THURSDAY, NOVEMBER 17 • WASHINGTON, D.C.



Cato's 34th Annual Monetary Conference will bring together leading experts to address the risks inherent in the unconventional monetary policies of the world's leading central banks and the steps that need to be taken to restore long-run economic growth. The lack of any monetary rule to guide policy decisions has created great uncertainty and increased financial volatility. Zero or negative interest rates and quantitative easing have created severe distortions in asset markets by increasing risk taking and politicizing credit allocation while failing to bring about robust economic growth.

FEATURED SPEAKERS

THOMAS J. SARGENT - *Nobel Laureate in Economics, and Professor of Economics at New York University*

THOMAS M. HOENIG - *Vice Chairman, Federal Deposit Insurance Corporation*

HON. PHIL GRAMM - *Former Chairman, Senate Banking Committee*

H. ROBERT HELLER - *Former Member, Federal Reserve Board of Governors*

JAMES GRANT - *Editor, Grant's Interest Rate Observer*

J. CHRISTOPHER GIANCARLO - *Commissioner, Commodity Futures Trading Commission*

STEVE HANKE - *Professor of Applied Economics, Johns Hopkins University*

MARK SPITZNAGEL - *President and CIO, Universa Investments, LP*

JOIN LEADING EXPERTS AS THEY DISCUSS

- Central Banking and Market Volatility. • Monetary Mischief and the "Debt Trap".
- The Problem of a Centrally Planned Monetary System. • Rethinking the Monetary Transmission Mechanism.



TO REGISTER AND FOR FURTHER INFORMATION,
VISIT WWW.CATO.ORG/MONETARY.

Cato Institute

Founded in 1977, the Cato Institute is a public policy research foundation dedicated to broadening the parameters of policy debate to allow consideration of more options that are consistent with the principles of limited government, individual liberty, and peace. To that end, the Institute strives to achieve greater involvement of the intelligent, concerned lay public in questions of policy and the proper role of government.

The Institute is named for Cato's Letters, libertarian pamphlets that were widely read in the American Colonies in the early 18th century and played a major role in laying the philosophical foundation for the American Revolution.

Despite the achievement of the nation's Founders, today virtually no aspect of life is free from government encroachment. A pervasive intolerance for individual rights is shown by government's arbitrary intrusions into private economic transactions and its disregard for civil liberties. And while freedom around the globe has notably increased in the past several decades, many countries have moved in the opposite direction, and most governments still do not respect or safeguard the wide range of civil and economic liberties.

To address those issues, the *Cato Institute* undertakes an extensive publications program on the complete spectrum of policy issues. Books, monographs, and shorter studies are commissioned to examine the federal budget, Social Security, regulation, military spending, international trade, and myriad other issues. Major policy conferences are held throughout the year, from which papers are published thrice yearly in the *Cato Journal*. The Institute also publishes the quarterly magazine *Regulation*.

In order to maintain its independence, the Cato Institute accepts no government funding. Contributions are received from foundations, corporations, and individuals, and other revenue is generated from the sale of publications. The Institute is a nonprofit, tax-exempt, educational foundation under Section 501(c)3 of the Internal Revenue Code.

CATO INSTITUTE
1000 Massachusetts Ave., N.W.
Washington, D.C. 20001

ENLIGHTENED REGULATION OF E-COMMERCE

The private sector should lead. Governments and regulators should avoid undue restrictions; support a predictable, consistent, and simple legal environment; and respect the “bottom-up” nature of the technology and its development in a global marketplace. . . . I look forward to working with my fellow CFTC commissioners, U.S. lawmakers, and other financial services regulators here and abroad to develop a “do no harm” framework from which to launch a new era of innovation in distributed ledger technology for the good of our markets and the people they serve.

—**J. Christopher Giancarlo**

“Regulators and the Blockchain: First, Do No Harm,”
Speech delivered to the Depository Trust and Clearing
Corporation Blockchain Symposium, March 29, 2016.

The **CATO JOURNAL**

VOLUME 36 NUMBER 3 • PUBLISHED BY THE CATO INSTITUTE • FALL, 2016