ARTICLES

Roger Pilon
American Constitutional Theory and History: Implications for European Constitutionalism

Deirdre N. McCloskey
Bauer on Culture and the Great Enrichment

Ian Vásquez
Peter Bauer Was Not a Cultural Pessimist: Comment on McCloskey

Alberto Mingardi
P. T. Bauer and the Myth of Primitive Accumulation

Paul E. Godek
Determining State Preferences for the Electoral College: 1788–2016

Martin Hutchinson and Kevin Dowd
The Apotheosis of the Rentier: How Napoleonic War Finance Kick-Started the Industrial Revolution

Edward J. Timmons and Catherine Konieczny
Untangling Hair Braider Deregulation in Virginia

Charles W. Calomiris
Restoring the Rule of Law in Financial Regulation

James A. Dorn
Leland B. Yeager: Market Grandmaster
BOOK REVIEWS

American Default: The Untold Story of FDR, the Supreme Court, and the Battle over Gold
Sebastian Edwards

Reviewed by Gerald P. O’Driscoll Jr. 737

The Virtue of Nationalism
Yoram Hazony

Reviewed by Alberto Mingardi 741

Hans Gersbach

Reviewed by Peter Bernholz 747

Capitalism without Capital: The Rise of the Intangible Economy
Jonathan Haskel and Stian Westlake

Reviewed by Thomas A. Hemphill 750

The Bitcoin Standard: The Decentralized Alternative to Central Banking
Saifedean Ammous

Reviewed by Tyler Whirty 759
AMERICAN CONSTITUTIONAL THEORY AND HISTORY: IMPLICATIONS FOR EUROPEAN CONSTITUTIONALISM

Roger Pilon

It is perhaps not impertinent to suggest that American constitutional theory and history, owing to the longevity of the document that is their subject, hold lessons for constitutionalism everywhere, but especially for European constitutionalism—the more recent and ever evolving treaties that serve as a “Constitutional Charter” for the European Union. An American constitutionalist looking east today, seeing everything from Brexit to Grexit plus the reactions in European capitals, must be struck by the tension in the EU between exclusion and inclusion in its many forms, including individualism and collectivism. Those themes underpin my discussion here. The issues surrounding them are universal. They are at the heart of the human condition.

In America we wrestled with them at our founding over 200 years ago, again in the aftermath of our Civil War, and yet again with the
advent of Progressivism, which culminated in our New Deal constitutional revolution. And we are still wrestling with them. Because America was founded on philosophical principles—First Principles, coming from the Enlightenment—it is particularly appropriate that we look at that experience to shed such light as we can on this more recent European constitutional experience.

But my more immediate concern is this: In liberal democracies today—nations constituted in the classical liberal tradition—we see the same basic problem, albeit with significant variations. It is that the growth of government, responding mainly to popular demand, has raised seemingly intractable moral and practical problems. First, increasing intrusions on individual liberty; and second, the unwillingness of people to pay for all the public goods and services they are demanding. Therefore, governments borrow. And that has led to massive public debt that saddles our children and grandchildren, to bankruptcy, and to the failure of governments to keep the commitments they have made.

In Italy, we need only look east, to the birthplace of democracy. But Greece is not alone in this. Nor are we in America immune. Cities like Detroit have gone bankrupt. So too, just recently, has the American territory of Puerto Rico. The state of Illinois has a credit rating today just above junk status, and Connecticut and New Jersey, among other states, are not far behind. At the national level, America’s debt today exceeds $20 trillion—that’s trillion—more than double what it was only a decade ago. And our unfunded liability vastly exceeds that (Cogan 2018).

What has this to do with constitutionalism? A great deal. Constitutions are written, after all, to discipline not only the governments they authorize but the people themselves. The point was famously stated by James Madison ([1788] 1961), the principal author of the U.S. Constitution. “In framing a government which is to be administered by men over men, the great difficulty lies in this,” he wrote: “you must first enable the government to control the governed; and in the next place oblige it to control itself. A dependence on the people is, no doubt, the primary control on the government,” Madison concluded, “but experience has taught mankind the necessity of auxiliary precautions.”

The principal such precaution, of course, is a well-written constitution. But no constitution is self-executing. It is people who
Constitutionalism

ultimately execute constitutions. In the end, therefore, the issue is cultural, a point I will come back to.

America’s Founders were deeply concerned with the problem of undisciplined, unlimited government. After all, they had just fought a war to rid themselves of distant, overbearing government. In drafting the Constitution, therefore, they were not about to impose that kind of government on themselves. In fact, during the ratification debates in the states, there were two main camps—the Anti-Federalists, who thought the proposed Constitution gave the government too much power, and the Federalists, who responded by pointing to the many ways the proposed Constitution would guard against that risk. The Federalists eventually won, of course, but the point I want to secure is that there was not a socialist in the group! There were limited government people, the Federalists, and even more limited government people, the Anti-Federalists.

So under a Constitution that has not changed that much, how did we go from limited to effectively unlimited government? The answer lies in the fundamental shift in the climate of ideas that began with Progressivism at the end of the 19th century, which the New Deal Supreme Court institutionalized in the 1930s. To illustrate that, I will first look closely at America’s founding documents: the Declaration of Independence, signed in 1776; the Constitution, ratified in 1788; the Bill of Rights, ratified in 1791; and the Civil War Amendments, ratified between 1865 and 1870, which corrected flaws in the original Constitution. Together, those documents constitute a legal framework for individual liberty under limited government, however inconsistent with those principles our actual history may have been.

I will then show how progressives rejected the libertarian and limited government principles of America’s Founders and how they eventually turned the Constitution on its head, not by amending it but through political pressure brought to bear on the Supreme Court. The problems that have ensued include the ones just noted: less liberty and increasing debt. But perhaps of even greater importance, for eight decades now the Supreme Court has struggled to square its post-New Deal decisions with the text and theory of the Constitution. That amounts to nothing less than a crisis of constitutional legitimacy.

And again, the basic reason for that crisis is the fundamental shift in outlook. Many Americans today no longer think of government as
earlier generations did. Whereas the Founders saw government as a “necessary evil,” to be restrained at every turn, many today think that the purpose of government is to provide them with vast goods and services, as decided by democratic majorities.

The Importance of Theory

I come, then, to the first important point I want to flag. You cannot understand the U.S. Constitution unless you understand the moral and political theory that stands behind it. And that was outlined not in the Constitution but in the Declaration of Independence (Sandefur 2015). The Constitution was written in a context, as were the later Civil War Amendments, and that context was one of natural law, Anglo-American common law, and even elements of Roman Law, all of which are captured succinctly in those famous words of the Declaration that I will quote in a moment. Indeed, President Abraham Lincoln’s famous Gettysburg Address, written in the throes of a brutal Civil War, begins with these words: “Fourscore and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.” Lincoln was reaching back to the Declaration, not to the Constitution.

Yet no less than my good friend and Italy’s gift to American constitutionalism, the late Justice Antonin Scalia, all but dismissed the Declaration as “philosophizing,” contrasting it with the Constitution’s “operative provisions” (Scalia 1997: 134). And his conservative colleague when the two served on the nation’s second highest court, the late Judge Robert Bork, wrote that “the ringing phrases [of the Declaration] are hardly useful, indeed may be pernicious, if taken, as they commonly are, as a guide to action, governmental or private” (Bork 1996: 57). Is it any wonder that there is constitutional confusion in America today when the document that is essential to understanding it plays little or no part in that understanding?

Let me now flesh out the argument by focusing on the underlying moral, political, and legal principles at stake, after which I will offer just a few reflections on how those principles might illuminate issues in the European context. Again, I want to show how the shift from limited to effectively unlimited government took place in America, despite very few constitutional changes. I should note, however, that it will be some time before I get to the Constitution. If a proper
understanding of the Constitution requires a proper understanding of the theory behind it, and if that theory is found implicitly in the Declaration, then that should be our initial focus, and will be for some time. That will take us into some of the deeper reaches of moral and political theory, the aim being to better understand the Constitution itself—and especially the broad principles that underpin it.

The first thing to notice about the American constitutional experience is how relatively different its beginnings were from those of many other nations. Constitution making and remaking often take place in the context of a stormy history stretching back centuries, even millennia. By contrast, America was a new nation. We came into being at a precise point in time, with the signing of the Declaration of Independence. To be sure, American patriots had to win our independence on the battlefield. And before that we had a colonial history of roughly 150 years. But America was created not by a discrete people but by diverse immigrants with unique histories all their own.

A second, crucial feature distinguishing America’s constitutional experience is that it unfolded during the intense intellectual ferment of the Enlightenment, including the Scottish Enlightenment, with its focus on the individual, individual liberty, and political legitimacy, all of which reflected the sense of “a new beginning.” Indeed, the motto on the Great Seal of the United States captures well the spirit of America’s origins: Novus ordo seclorum, “a new order of the ages.”

The Declaration of Independence

Let us turn, then, to that new order, as outlined in the Declaration. Penned near the start of our struggle for independence, the Declaration in form is a political document. But were it merely that, it would not have so endured in our national consciousness. Nor would it have inspired countless millions around the world ever since, leading many to leave their homelands to begin life anew under its promise, including millions from Italy who now enrich America. It has so inspired because, fundamentally, it is a profound moral statement. Offered from “a decent Respect to the Opinions of Mankind” and invoking “the Laws of Nature and of Nature’s God,” it was written not only to declare but to justify our independence. And it did so not simply by listing the king’s “long Train of Abuses and
Usurpations,” which constitute the greater part of the document, but by first setting forth the moral and political vision that rendered those acts unjust.

And so we come to those famous words that flowed from Thomas Jefferson’s pen in 1776, words that capture fundamental principles concerning the human condition:

> We hold these Truths to be self-evident, that all Men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the Pursuit of Happiness—That to secure these Rights, Governments are instituted among Men, deriving their just Powers from the Consent of the Governed.

The first thing to notice about that passage is that its propositions are asserted as “truths,” not mere opinions. The Founders were not moral relativists. They were confident in their claims. And why not? Their truths were said to be “self-evident,” grounded in universal reason, accessible by all mankind—and the evidence supports that.

Notice too the structure of the passage: There are two parts—and the order is crucial. The moral vision comes first, defined by equal rights. The political and legal vision comes second, defined by powers, as derived from the moral vision. And right there is the second major point I want to flag: Unlike today, where politics, grounded in will, so often determines what rights we have, for early Americans, morality, grounded in reason, determined our rights. The Founders were concerned fundamentally with moral and political legitimacy. Rights first, government second, as the means for securing our rights (Barnett 2016, Pilon 1999).

Given that order of things, the Founders were engaged in “state-of-nature theory,” a rudimentary form of which can be found in the writings of Seneca (see Corwin 1955: 15). A fuller discussion came much later in the work of Thomas Hobbes (1651) and, especially, John Locke (1690)—often said to be the philosophical father of America.

State-of-nature theory is a thought experiment. The idea is to show how, without violating any rights, a legitimate government with legitimate powers might arise from a world with no government. Thus, the first step is to show, from pure reason, what rights we would have in such a world.
For that, as the Declaration implies, we turn to the natural law tradition—more precisely, the natural rights strain coming from the Reformation and the Enlightenment. Simply put, natural law stands for the idea that there is a “higher law” of right and wrong, grounded in reason, from which to derive the positive law, and against which to criticize that law at any point in time. There is nothing suspect about that idea, as modern moral skeptics argue. We appeal to natural law when the positive or actual law is thought to be morally wrong. In America, the abolitionists, the suffragists, and the civil rights marchers all invoked our natural rights in their struggles to overturn unjust law.

The origins of this law are in antiquity. Many of its particulars are in Roman Law, especially the law of property and contract. Over some 500 years in England, prior to the American Revolution, this law was refined and reduced to positive law by common-law judges consulting reason, custom, and what they knew of Roman Law as they adjudicated cases brought before them by ordinary individuals (Corwin 1955: 26; Leoni 1961). And John Locke drew largely on that body of common-law rights as he crafted a theory of natural rights, much as Jefferson drew on Locke when he drafted the Declaration.

To correct a common misunderstanding, these are the rights we hold against each other, and would hold in a state of nature. Later, once we create a government, they will serve as rights we hold against that government, and likely be included in a bill of rights.

To discover and justify these rights in detail, as I and others have done (Pilon 1979; Epstein 2003), we would need to delve into the complex issues of moral epistemology and legal casuistry, and this is not the occasion for that. Suffice it to say that, when that foundational work is done, the conclusion one reaches is the same one America’s Founders reached through reason and experience—namely, that our basic right is the right to be free from the unjustified interference of others, and all other rights are derived from that basic right, as the facts may warrant. What results approximates largely the judge-made common law of property, torts, contracts, and remedies, a law that defines our private relationships, as it did in early America both before and long after the Revolution. It is a law that says, in essence, that each of us is free to pursue happiness, by his own subjective values, either alone or in association with others, provided we respect the equal objective rights of others to do the same. In short, it is a live-and-let-live law of liberty.
And I can summarize it with three simple rules, so simple that even a child can understand them.

- Rule 1: Don’t take what belongs to someone else. That is the whole world of property, broadly conceived as Locke did—our property in our “Lives, Liberties, and Estates.”
- Rule 2: Keep your promises. That is the whole world of contracts and associations.
- Rule 3: If you have wrongly violated rules 1 or 2, give back what you have wrongly taken or wrongly withheld. That is the whole world of remedies.

There is a fourth rule, however, but it is optional: Do some good. You’re free not to be a Good Samaritan, but you should be one if you are a decent human being and the cost to you is modest. Unlike much continental law, Anglo-American law never compelled strangers to come to the aid of others (Ratcliffe 1966). It did not because individual liberty is its main object. And it saw that there is no virtue in forced beneficence. We are free to criticize those who don’t come to the aid of others, and we should, even as we defend their right not to.

Why have I mentioned this fourth, voluntary rule? Again, it is because, when we start from a theoretical state of nature, we need to know what rights we do and do not have for government to enforce once we bring government into the picture. And the Good Samaritan is the modern welfare state writ small. If there is no right to be rescued, there is no correlative obligation for government to enforce. Recognizing that raises important questions about the very legitimacy of the welfare state.¹

Leaving the State of Nature and the Problem of Political Legitimacy

To get to the Constitution, however, we need now to take the last step in the argument. We need to derive a legitimate government

¹That is not to say that, as a practical matter, elements of the welfare state may not be justified as a last resort. Rather, such elements are not brought into being “by right.” Put differently, there is a strong moral presumption against such measures—against forcing people to assist others through taxation or otherwise—and a strong presumption in favor of voluntary private assistance and private charity.
Constitutionalism

with legitimate powers—and that is no easy matter. I have said little about enforcement so far. The Declaration says that government’s purpose is to secure our rights, its just powers derived “from the consent of the governed.” Thus, the Founders invoked the social contract, which grounds political legitimacy in consent.

But there are well-known problems with consent-based social-contract theory as a ground for political legitimacy. The question is how to move legitimately from self-rule to collective rule. Unanimity will achieve legitimacy, of course, but rarely if ever do we get it. Majoritarianism will not solve the problem, because it amounts to tyranny over the minority that has not consented. Nor will the social contract work, except for those in the original position who agree thereafter to be bound by the will of the majority. Nor, finally, will so-called tacit consent work—“you stayed, therefore you’re bound by the majority”—because it puts the minority to a choice between two of its rights, its right to stay where it is and its right not to be ruled by the majority, precisely what the majority must justify on pain of circularity. As for elections, an occasional vote hardly justifies all that follows.

As a practical matter, the social contract argument may be the best we can do, but recognizing its infirmities leads to a compelling conclusion—and to the third basic point I want to flag, namely, that there is an air of illegitimacy that surrounds government as such. Government is not like a private association that we can join or leave at will. It is a forced association. Its very definition entails force. And once we recognize its essential character, that should compel us, from a concern for legitimacy, to do as much as we can through the private sector where it can be done voluntarily and hence in violation of the rights of no one, and as little as possible through the public sector where individuals will be forced into programs they may want no part of.

In short, as a moral matter, there is a strong presumption against doing things through government. We should turn to government not as a first but only as a last resort, when all else fails.

Still, we can refine this conclusion. We can distinguish three distinct powers in decreasing degrees of legitimacy. The first is the police power—the power, through adjudication or legislation, to more precisely define and enforce our rights. As such, it is bound by the rights we have to be enforced, although it includes the power to provide limited “public goods” like national defense, clean air, and
certain infrastructure—goods described by nonexcludability and nonrivalrous consumption, as economists define them (Cowen 2008).

When we leave the state of nature, we give government that power to exercise on our behalf. But because we had the power in the state of nature—Locke called it the “Executive Power” each of us has to secure his rights—to that extent it is legitimate. Only the anarchist who would prefer to remain in the state of nature can be heard to complain. Fortunately, there are few of those.

Less legitimate is the eminent domain power—the power to condemn and take private property for public use after paying the owner just compensation—because none of us would have such a power in the state of nature. Such legitimacy as this power enjoys, at least in America, is because we gave it to government when we ratified the Constitution’s Fifth Amendment, which includes the Takings Clause; and it is “Pareto optimal,” as economists say, meaning that at least one person is made better off by its use—the public, as shown by its willingness to pay—and no one is made worse off—the owner, provided he is indifferent as to whether he keeps the property or receives the compensation, which he rarely is, unfortunately.

The third great governmental power, ubiquitous today, is the least legitimate. In fact, from a natural rights perspective, it enjoys no legitimacy. It is the redistributive power, and it takes two forms, material and regulatory. Through redistributive taxation, government takes from A and gives to B. Through redistributive regulation, government prohibits A from doing what he would otherwise have a right to do or requires him to do what he would otherwise have a right not to do, all for the benefit of B. Those powers describe the modern redistributive and regulatory state. No one would have them in the state of nature. How then could government get them legitimately, since governments, in the classical liberal tradition, get whatever powers they have from the people, who must first have those powers to yield up to government?

There are three main answers. First, if that redistribution arose through unanimous consent, there would be no problem; but again, rarely if ever does that occur in the public domain. Second, majorities gave governments those powers. That raises the classic problem of the tyranny of the majority, as already mentioned. And third, special interests have learned how to work the system for their benefit,
as public choice economists have long explained. That is the tyranny of the minority—and the main source today of such schemes.

We can conclude this examination of the moral foundations of the classical liberal vision by imagining a continuum, with anarchy or no government at one end—our state of nature—and totalitarianism at the other end, where everything possible is done through government. At the anarchy end, individuals are free to plan and live their lives as they wish, alone or in cooperation with others. They will soon find, however, that there are some things best done collectively, like the provision and enforcement of law, national defense, clean air and water, limited infrastructure, and the like—public goods—and most will consent to the public provision of such goods. But as we move up the continuum toward totalitarianism and try to bring more and more private goods under public provision—education, health care, child care, jobs, housing, ordinary goods and services—people start voting with their feet. The Berlin Wall was not built to keep West German workers out of the workers’ paradise to the east.

The moral, political, and legal vision implicit in the Declaration of Independence is closer to the anarchy end of that continuum. America’s Founders envisioned a land in which people were free to live as they wished, respecting the equal rights of others to do the same, with government there to secure those rights and do the few other things it was authorized to do.

That basic moral vision is perfectly universalizable. How to secure it through the rule of law is another matter. Certain basic legal principles are themselves universalizable and are common to most legal systems, but whether a nation has a parliamentary system as in much of Europe, or a republican form of government as in America, or some other arrangement is not a matter of natural law. Let us now see how the Founders framed a constitution to secure the Declaration’s moral vision.

The Constitution

After we declared independence, and during our struggle for it, we lived under our first constitution, the Articles of Confederation. As its name implies, it was a loose agreement among the 13 states, authorizing a national government that hardly warranted the name.

The classic arguments are by Buchanan and Tullock (1962).
Three main problems lay ahead. Surrounded on three sides by great European powers, our national defense was painfully inadequate. Second, states were erecting tariffs and other barriers to free interstate trade. And finally, our war debts remained unpaid. After 11 years, the Framers met in Philadelphia to draft a new Constitution.

The main problem they faced was how to strike a balance. They needed to give the new government enough power to address those problems and accomplish its broad aims, yet not so much power as to risk our liberties. Those aims were set forth in the Constitution’s Preamble:

We the People of the United States, in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common defence, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America.

Notice: states aside, regarding the proposed new government, we are right back in the state of nature, about to “ordain and establish” a constitution to authorize it and bring it into being. All power rests initially with “we the people.” We bring the constitution and the government that follows into being through ratification. We give it its powers, such as we do. The government does not give us our rights. We already have our rights, natural rights, the exercise of which creates and empowers this government.

How, then, does Madison strike the balance between power and liberty in service of those aims? First, through federalism: Power was divided between the federal and state governments, with most power left with the states, especially the general police power—the basic power of government to secure our rights, as just discussed. The powers we delegated to the federal government concerned national issues like defense, free interstate commerce, rules for intellectual property, a national currency, and the like.

Second, following Montesquieu, Madison separated powers among the three branches of the federal government, with each branch defined functionally. Pitting power against power, he provided for a bicameral legislature, with each chamber constituted differently; a unitary executive to enforce national legislation and conduct foreign affairs; and an independent judiciary with the
Constitutionalism

implicit power to review legislative and executive actions for their constitutionality—a novel institution at that time, and a crucial one as time went on.

Third, although the Constitution left most of the rules for elections with the states, it provided for periodic elections to fill the offices set forth in the document, thus leaving ultimate power with the people.

But while each of those provisions and others struck a balance between power and liberty, the main restraint on overweening government took the name of the doctrine of enumerated powers. And I can state it no more simply than this: if you want to limit power, don’t give it in the first place. We see that doctrine in the very first sentence of the Constitution, after the Preamble: “All legislative Powers herein granted shall be vested in a Congress . . . .” By implication, not all powers were “herein granted.” Look at Article I, section 8, and you will see that Congress has only 18 powers or ends that the people have authorized. And the last documentary evidence from the founding period, the Tenth Amendment, states that doctrine explicitly: “The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people.” In other words, the Constitution creates a government of delegated, enumerated, and thus limited powers. If a power is not found in the document, it belongs to the states—or to the people, never having been given to either government.

As noted earlier, when the Constitution was sent out to the states for ratification, it met stiff resistance as Anti-Federalists thought it gave too much power to the national government. Only after the Federalists agreed to add a bill of rights was it finally ratified. During the first Congress in 1789, Madison drafted 12 amendments, 10 of which were ratified in 1791 as the Bill of Rights. That document sets forth rights that are good against the federal government, such as freedom of religion, speech, press, and assembly, the right to keep and bear arms, to be secure against unreasonable searches and seizures, to due process of law, to compensation if private property is taken for public use, to trial by jury, and more.

But it is important to note that the Bill of Rights was, as Justice Scalia (2017: 161) said, an “afterthought.” Unlike with many European constitutions, which begin with a long list of rights, many aspirational, the Framers saw the Constitution’s structural provisions as their main protection against overweening government (National
Lawyers Convention 2017). And on that score, it is crucial to mention the Ninth Amendment, which reads: “The enumeration in the Constitution, of certain rights, shall not be construed to deny or disparage others retained by the people.”

The history behind that amendment is instructive. During the ratification debates, there were two main objections to adding a bill of rights. First, it would be unnecessary. “Why declare that things shall not be done,” asked Alexander Hamilton ([1788] 1961), “which there is no power to do?” Notice that he was alluding to the enumerated powers doctrine as the main protection for our liberties: where there is no power, there is a right.

And second, it would be impossible to enumerate all of our rights, yet, by ordinary principles of legal construction, the failure to do so would be construed as implying that only those rights that were enumerated were meant to be protected. To guard against that, the Ninth Amendment was written. It reads, again, “The enumeration in the Constitution, of certain rights, shall not be construed to deny or disparage others retained by the people.” Notice: “retained by the people.” You can’t retain what you don’t first have to be retained. The allusion is to our natural rights, which we retained when we left the state of nature, save for those we gave up to government to exercise on our behalf, like the right to enforce our rights.

For a proper understanding of the Constitution, the importance of the Ninth Amendment, which speaks of retained rights, and the Tenth Amendment, which speaks of delegated powers, cannot be overstated (Pilon 1991: 1). Taken together, as the last documentary evidence from the founding period, they recapitulate the vision of the Declaration. We all have rights, enumerated and unenumerated alike, to pursue happiness by our own lights, to plan and live our lives as we wish, provided we respect the rights of others to do the same; and federal and state governments are there to secure those rights through the limited powers we have given them toward that end. There, in a nutshell, is the American vision, reduced from natural to positive law.

But apart from our failure too often to abide by that vision, there was a structural problem with the original design. There were too few checks on the states, where most power was left. And the reason

---

3Expressio unius est exclusio alterius.
Constitutionalism

was slavery. To achieve unity among the states, the Framers made their Faustian bargain. They knew that slavery was inconsistent with their founding principles. They hoped it would wither away in time. It did not. It took a brutal civil war to end slavery and the Civil War Amendments to “complete” the Constitution by incorporating at last the grand principles of the Declaration, especially equality before the law (Reinstein 1993).

The Thirteenth Amendment, ratified in 1865, rendered slavery unconstitutional. The Fifteenth Amendment, ratified in 1870, protected the right to vote from being denied on account of race. And the Fourteenth Amendment, ratified in 1868, defined federal and state citizenship and provided federal remedies against a state’s violating the rights of its own citizens.4

Unfortunately, only five years after the Fourteenth Amendment was ratified, a deeply divided 5–4 Supreme Court eviscerated the principal font of substantive rights under the amendment, the Privileges or Immunities Clause.5 Thereafter the Court would try to do under the less substantive Due Process Clause what was meant to be done under privileges or immunities, and the misreading of the Fourteenth Amendment has continued to this day. Among other things, the upshot was Jim Crow racial segregation in the South, which lasted until the middle of the 20th century.

Progressivism

We turn now to the great ideological watershed, the rise of Progressivism at the end of the 19th century. Coming from the elite universities of the Northeast, progressives rejected the Founders’ libertarian and limited government vision (Pestritto and Atto 2008). They were social engineers, planners enamored of the new social sciences. Insensitive when not hostile to the power of markets to order human affairs justly and efficiently, they sought to address what they saw as social problems through redistributive regulatory legislation. They looked to Europe for inspiration: Bismarck’s social security scheme, for example, and British utilitarianism, which in ethics had replaced natural rights theory. The idea was that policy, law, and

4Prior to that time, the Bill of Rights applied only against the federal government. Barron v. Baltimore, 32 U.S. 243 (1833).
5Slaughterhouse Cases, 83 U.S. 36 (1873).
judgment were to be justified not by whether they protected our natural and moral rights but by whether they produced the greatest good for the greatest number—often by giving rights to some, taken from others.

A particularly egregious example of that rationale concerned a sweetheart suit brought against a Virginia statute that authorized the sterilization of people thought to be of insufficient intelligence. Part of the bogus “eugenics” movement, the law was designed to improve the human gene pool. Writing for a divided Supreme Court in 1927, the sainted Justice Oliver Wendell Holmes upheld the statute, ending his short opinion with the ringing words, “Three generations of imbeciles are enough.” There followed some 70,000 sterilizations across the nation.

Some of what the progressives did was long overdue, like promoting municipal health and safety measures and attacking corruption. Yet they also sowed the seeds for later corruption, especially through regulatory schemes ripe for special interest capture, replacing markets with cartels (Epstein 2006). And their record on racial matters was abysmal (Sowell 2016).

During the early decades of the 20th century, progressives directed their political activism mostly at the state level, but they often failed as the courts upheld constitutional principles securing individual liberty and free markets. With the election of Franklin Roosevelt in 1932, however, progressive activism shifted to the federal level. Still, during the president’s first term the Supreme Court continued mostly to uphold limits on federal power, finding several of Roosevelt’s programs unconstitutional.

With the landslide election of 1936, however, things came to a head. Early in 1937, Roosevelt unveiled his infamous Court-packing scheme, his threat to pack the Court with six new members. Uproar followed. Not even an overwhelmingly Democratic Congress would go along with the plan. Nevertheless, the Court got the message. The famous “switch in time that saved nine” justices followed. The Court began rewriting the Constitution, in effect, not through amendment by the people, the proper way, but by reading the document as it hadn’t been read for 150 years—as authorizing effectively unlimited government (Leuchtenburg 1995).

---

The Court did that rewrite in three basic steps. First, in 1937 it eviscerated the very centerpiece of the Constitution, the doctrine of enumerated powers. Then in 1938 it bifurcated the Bill of Rights and gave us a bifurcated theory of judicial review. Finally, in 1943 it jettisoned the nondelegation doctrine. Let me describe those steps a bit more fully so you can see the importance of recognizing and adhering to the theory that stands behind and informs a constitution.

The evisceration of the doctrine of enumerated powers involved three clauses in Article I, section 8, where Congress’s 18 legislative powers are enumerated: the General Welfare Clause, the Commerce Clause, and the Necessary and Proper Clause. All were written to be shields against government. The New Deal Court turned them into swords of government through which the modern redistributive and regulatory state has arisen.

The first of Congress’s enumerated powers, where the General Welfare Clause is found, authorizes Congress, in relevant part, to tax to provide for the “general Welfare of the United States.” As Madison wrote in *Federalist* No. 41, that qualifying language was simply a general heading under which Congress’s 17 other powers or ends were subsumed, for which Congress may tax, but only if they serve the general welfare of the United States, not particular or local welfare.

Instead, the New Deal Court read the clause as an independent power authorizing Congress to tax for whatever it thought might serve the “general welfare.” That reading could not be right, however, because it would enable Congress to tax for virtually any end, thus rendering Congress’s other powers superfluous, as Madison, Jefferson, and many others noted when the issue arose early in our history. Indeed, it would turn the Constitution on its head by allowing Congress effectively unlimited power. Such is the result from ignoring the document’s underlying theory of limited government.

Similar issues arose that year with the Commerce Clause, which in relevant part authorized Congress to regulate interstate commerce. Recall that, under the Articles of Confederation, states had begun erecting tariffs and other protectionist measures, and that was leading to the breakdown of free trade among the states. Thus, the Framers gave Congress the power to regulate—or make regular—commerce among the states, largely by negating state actions that

---

impeded free trade, but also through affirmative actions that might facilitate that end (Barnett 2001).

Over several decisions, however, beginning in 1937, the New Deal Court read the Commerce Clause as authorizing Congress to regulate anything that “affected” interstate commerce, which of course is virtually everything. Thus, in 1942 the Court held that, to keep the price of wheat high for farmers, Congress could limit the amount of wheat a farmer could grow, even though the excess wheat in question in the case never entered commerce, much less interstate commerce, but was consumed on the farm by the farmer and his cattle. The Court held that the excess wheat he consumed himself was wheat he would otherwise have bought on the market, so “in the aggregate” such actions “affected” interstate commerce. Such were the economic theories of the Roosevelt administration.

The last of Congress’s 18 enumerated powers authorizes it “to make all laws which shall be necessary and proper for carrying into execution the foregoing powers.” Thus, the clause affords Congress instrumental powers—the means for executing its other powers or pursuing its other enumerated ends. “Necessary” and “proper” are words of limitation, of course: Not any means Congress desires will do. Yet the New Deal and subsequent Courts, until very recently, have hardly policed those limitations (Blumstein 2012: 86).

Turning now to the second step, despite the demise in 1937 of the doctrine of enumerated powers, one could still invoke one’s rights against Congress’s expanded powers. So to address that “problem,” the New Deal Court added a famous footnote to a 1938 opinion. In it, the Court distinguished two kinds of rights: “fundamental,” like speech, voting, and, later, certain personal rights; and “nonfundamental,” like property rights and rights we exercise in “ordinary commercial relations.” If a law implicated fundamental rights, the Court would apply “strict scrutiny” and the law would likely be found unconstitutional. By contrast, if nonfundamental rights were at

---

10For a recent exception, see NFIB v. Sebelius, 567 U.S. 519 (2012).
12To satisfy the strict scrutiny test, the government must have a “compelling interest,” and the means it employs must be “narrowly tailored” to serve that interest.
issue, the Court would apply the so-called rational basis test, which held that if there were *some* reason for the law, if you could *conceive* of one, the law would be upheld. Thus was economic liberty reduced to a second-class status. None of this is found in the Constitution, of course. The Court invented it from whole cloth to make the world safe for the New Deal programs (Pilon 2003).

Finally, in 1943 the Court jettisoned the nondelegation doctrine,\(^{13}\) which arises from the very first word of the Constitution: “*All* legislative Powers herein granted shall be vested in a Congress . . . .” Not some; all. As government grew, especially during the New Deal, Congress began delegating ever more of its legislative power to the executive branch agencies it was creating to carry out its programs. Some 450 such agencies exist in Washington today. Nobody knows the exact number.

That is where most of the law Americans live under today is written, in the form of regulations, rules, guidance, and more, all issued to implement the broad statutes Congress passes. Not only is this “law” written, executed, and adjudicated by unelected, non-responsible agency bureaucrats—raising serious separation-of-powers questions—but the Court has developed doctrines under which it defers to agencies’ interpretations of statutes, thus largely abandoning its duty to oversee the political branches. Governed largely today under administrative law promulgated by the modern executive state, we are far removed from the limited, accountable government envisioned by the Founders and Framers (Hamburger 2014, 2017).

This completes my overview of American constitutional theory and history. From it, as I mentioned early on, the main lesson to be drawn is that culture matters. The Founders and Framers were animated by individual liberty under limited government. When the post–Civil War Framers revised our original federalism, they did it the right way, by amending the Constitution to make it consistent with its underlying moral and political principles. The New Deal politicians, having less regard for the Constitution and its underlying principles, rejected that course, choosing instead to browbeat the Court into effectively rewriting the Constitution, undermining its moral and political principles in the process.

\(^{13}\)National Broadcasting Co. v. United States, 319 U.S. 190 (1943).
But don’t take my word for it. Here is Franklin Roosevelt (1935), writing to the chairman of the House Ways and Means Committee: “I hope your committee will not permit doubts as to constitutionality, however reasonable, to block the suggested legislation.” And here is Rexford Tugwell (1968: 20), one of the principal architects of the New Deal, reflecting on his handiwork some 30 years later: “To the extent that these [New Deal policies] developed, they were tortured interpretations of a document intended to prevent them.” They knew exactly what they were doing. They were turning the Constitution on its head.

Thus, the problem today is not, as so many America progressives think, too little government. It is too much government, intruding on our liberties and driving us ever deeper into debt. And it isn’t as if our Founders did not understand that. As Jefferson famously wrote, “The natural progress of things is for liberty to yield, and government to gain ground” (Boyd 1956: 208–10). The remedy for that “progress” is a good constitution, but it must be followed. And that takes good people at every stage—including, ultimately, the people themselves.

A Few Implications for European Constitutionalism

So what lessons might we draw from the American experience for European constitutionalism? Recall my mentioning earlier of being struck by the tension in the EU between exclusion and inclusion in its many forms, including individualism and collectivism. As we have seen, that same tension runs through America’s constitutional history as well. To address deficiencies in the Articles of Confederation, the original Constitution moved toward greater inclusion to form “a more perfect Union.” But the resulting federalism did not get the balance right either. It left too much power with the states, enabling the southern states to continue enacting slavery. So the Civil War Amendments increased the inclusion, correctly. The adjusted federalism gave more power to the federal government, enabling it to block states from oppressing their own citizens—a higher power checking a subsidiary power.

But that balance, reflecting the nation’s underlying principles, was upended again by the far more inclusive New Deal constitutional revolution. Giving vastly more power to the federal government, contrary to the nation’s limited government principles, this change swept
ever more Americans into public programs, leading many to want out. They wanted to be excluded from the socialization of life, as reflected by the rise of the conservative and libertarian movements in the second half of the 20th century.

Are there parallels with post-War developments in Europe? To this sometime-student of European affairs, there seem to be; but the inclusion that began with the 1951 Treaty of Paris and continued through the many treaties since makes it difficult if not impossible to speak of three distinct periods, as in America, much less point to a “golden mean” in this evolution akin to America’s post–Civil War settlement. In recent years, however, the impetus toward exclusion, in many forms, is unmistakable, Brexit being only the most prominent example, the ongoing refugee resettlement crisis being another.

Federalism within nations is a delicate balance. Federalism among sovereign nations, which is what the EU amounts to, is far more difficult, especially when cultural differences loom large. And on that score, here is a paradox. Europeans have always been more comfortable than Americans with collectivization in the form of the welfare state, certainly within their respective nations (Rhodes 2018). But with collectivization among nations, cultural differences—rich and poor being only one axis—can easily exacerbate the cooperation that is required if collectivization is to work at all, much less with any measure of efficiency. The evidence suggests that the EU has gone too far in that direction. At the same time, the evidence is equally clear that the failure to make EU border security an EU responsibility, leaving it instead to individual members, has raised serious problems, too (Rohac 2016).

In America, border security became a federal government function once the Constitution was ratified. Within our borders, however, to keep states honest, the Founders instituted competitive federalism, whereby states compete for the allegiance of citizens; and it has largely worked as states with high taxes and excessive regulations lose firms and people to states with low taxes and reasonable regulations. People vote with their feet, much as in the Schengen Area. But the federal income tax plus the direct election of senators, both enacted as constitutional amendments in 1913 and both promoted by progressives, unleashed cooperative federalism whereby federal and state officials collude, using federal funds and enacting federal regulations, to undercut state autonomy and the discipline that competitive federalism was meant to secure (Greve 2012; Buckley 2014).
Earlier I said that you cannot understand the American Constitution unless you understand the theory behind it. Well, what is the theory behind the treaties that compose the EU Constitution? Peace through trade and cooperation, yes—given Europe’s long history of wars. But beyond that, what? We have seen how a radical shift in the climate of ideas in America, especially in the direction of collectivism, has led, as many lonely voices predicted, to a reaction that today reflects a deeply divided nation, unable to restrain its appetite for “free” goods and services, even in the face of crushing debt. The divisions surfacing recently in Europe are no accident. People and peoples yearn to breathe free—in an earlier understanding of that idea. The balance needed to ensure that freedom may be difficult to find. But to discover it, as we celebrate Italy’s Constitution today and reflect on Italy’s place within the larger European Community, we could do no better than to repair to the First Principles that are the very foundation of civilized nations.

References


Bauer on Culture and the Great Enrichment

Deirdre N. McCloskey

I first commented on P. T. Bauer 30 years ago, at a point I optimistically thought was his “resurrection” as a voice in the fraught field of development economics. I wrote in 1987 that his “story follows William James’s three stages in the rhetoric of academic disputes: at first what Bauer says is plainly false; then it is trivially obvious; and finally it is so true that we, not he, invented it” (McCloskey 1987: 253). By now the joke has come true. Some of us have forgotten, but many now know, that Bauer invented in the 1950s and 1960s, reiterating later, what has become trivially obvious from the experience of China and India, Ireland and Botswana—namely, that leaving people alone, while assigning the government to the few if important tasks that do not obstruct opportunity, is the path to wealth. What does not work is “socialism with Chinese characteristics,” that is, political tyranny and unprofitable governmental enterprises. The economic liberalism of the competition for business among Chinese xians is what worked, leaving people alone to innovate, just as Bauer would have said (Coase and Wang 2013, and the later works of S. N. S. Cheung).
Bauer: A Classical Liberal

As Bauer remarked about his book of 1954 at its reissue in 1963, “the discussion of price stabilization and of the operation of the marketing boards [anticipating the political scientist Robert Bates (1989)] aroused much controversy at the time, but the analysis and the conclusions are no longer disputed” (Bauer [1954] 1963: xviii–xix). Likewise his views about the corrupting effects of foreign aid, anticipating those of the political scientist Edward Banfield (1963), and more recently the economist William Easterly (2001), as Easterly has admitted, is now no longer much disputed.

Yet it seemed to us lefties of the early 1960s to be plainly false. Surely the way to wealth in Ghana, we thought in our admiration for Kwame Nkrumah, is giving massive aid to the Ghanaian government, out of, say, Norwegian taxes. Anything less would be cruelly selfish. Shame on you conservatives and classical liberals who doubt. But in 1954 when he was criticizing what Easterly nearly 50 years later called, appropriately, the “capital fundamentalism” of the World Bank and other foreign aiders (Easterly 2001), Bauer wrote: “the comparative lack of local technical and administrative skills aggravates the effects of the scarcity of equipment. . . . For this reason indiscriminate import of capital, or even substantial capital accumulation in the hands of public organizations, alone would not necessarily improve the situation” ([1954] 1963: 13).

Bauer’s great advantage was that unlike many economists he understood “price theory,” as we called it in the good old days at the University of Chicago (e.g., McCloskey 1985). That is, he understood the way an economy works through scarcity, entry, and supply and demand curves. Back in 1848 the field of economics, or rather “political economy” as it called itself then, had a reasonable grip on such matters, which guided liberals such as Mill and Bastiat and Cobden. The grip was strengthened by the marginal revolution in the economics of the 1870s.

But the 1870s was also the era in which theories of American protectionism and British New Liberalism and the German Historical School among other anti-economic movements started to take hold outside the price-theoretic and British/Austro-Hungarian core of the field. By 1975, Bauer noted with irritation,

Some economists holding senior academic positions confuse free goods and scarce resources [e.g., in thinking that

588
development spending is a net addition to national income, regardless of its opportunity cost], ignore the dependence of supply and demand on price [e.g., speaking of the numerical “structure” of jobs or exports or the balance of payments without regard to their sensitivities to price elasticities of supply and demand], or neglect patent empirical evidence pertinent to their arguments [e.g., evidence of entry at the pull of profit] [Bauer 1975: 287].

Bauer was therefore not misled, as so many economists are, by the litany of “imperfections” in the market, of which I have recently counted fully 110 imagined since 1848—monopoly, externalities, inadequate aggregate demand, irrational consumers, informational asymmetries, and on and on, recently bearing fruit in many a Nobel Memorial Prize (McCloskey 2018a). Not one of them—startlingly in what purports to be a serious empirical science—has been shown to be a substantial obstacle to economic progress, except on the blackboard. All are used to recommend corrective governmental action by saintly geniuses able to predict and therefore to engineer the future without flaw for the good of us all. As Comte, the master of such thinking, put it in 1830, Savoir pour prévoir, afin de pouvoir, “Know in order to predict, to be able to act with power” in the state. Meanwhile the highly “imperfect” economy, chiefly by ignoring the statist advice of the increasing number of illiberal economists, yielded a Great Enrichment from 1800 and especially from 1848 to the present of 3,000 percent more goods and services for the poorest among us, uniquely in economic history.

The behavioral economist Richard Thaler is the best example of an un-Bauerish and illiberal approach to price theory and practice that human frailty is likely to yield (McCloskey 2018b). He combines the 110 imperfections of the market with the 257 cognitive biases that the psychologists have discerned.\(^1\) He concludes that without governmental help we cannot be trusted to walk across the street, and certainly not to make any serious economic decisions, considering that the imperfection-crippled market will not offer us useful protections from our idiocy. Therefore we need to be nudged to safety, like a two year old grabbed by his mother before he carries out his intention to run in front of

\(^1\)The number 257 comes from the Wikipedia entry “Cognitive Biases.”
a tram. The conclusion by most economists of the past century has been that we are little children, or idiotic adults, and need to be economically engineered, by those very economists. We naïve statistists in the 1960s called it “fine tuning.” Now “nudging.” In other cases, socialism and fascism.

Like the agricultural economist and Nobelist at the University of Chicago, Theodore Schultz, and a few other brave souls writing in the 1950s and 1960s, Bauer didn’t think that people in poor countries were little children or in other ways idiotic (Schultz 1964). For example, Bauer did not believe the racist assumption, widely if sometimes unselfconsciously held in the 1950s and 1960s, that Africans or Indians or whoever could not possibly achieve the Great Enrichment of 3,000 percent, available only to Europeans sporting melanin-challenged skin. We young students of economic development in the 1960s, at any rate if we were not studying at the London School of Economics or Chicago with Peter Bauer or B. S. Yamey or Theodore Schultz, were taught that such folk would never grow rich, that they were caught in a low-level trap of the sort that Professors Myrdal and Nurkse exposited. After all, the Indians were mostly Hindus, or at best Muslims, and many of them in the south of India were dark fellows who could not possibly develop a world-supplying computer-service industry. The same held for the Chinese, those hopeless Confucians or Communists, and in any case, you will note, yellow, who could not possibly develop a world-supplying electrical-machinery industry.

Bauer, in other words, was a classical liberal at the height of a statism of the left or the right or the middle.

A Social Scientist as Well as a Price Theorist

But he was also a social scientist of his age, I think, in taking cultural obstacles as more powerful than subsequent experience suggests. Thus too Edward Banfield in his dismal, classic study of the “amoral familism” of Italy south of Rome (1958), followed by Robert Putnam and coauthors in their dismal, classical study of the same place (1993), felt that the South was hopelessly trammeled by its culture, and would always be. No reining in of bad governmental policy could solve the puzzle of culture inherited from the past. The Italian party of separation, once called Lega Nord per l’Indipendenza della Padania (“North League for the Independence of Padania,” that is, of the Po Valley), believes the same to this day; and there has long
been respectable opinion in the study of Italian history that the uni-
ification of the peninsula in the Risorgimento was a mistake (Mack

Against the Banfield-Putnam pessimism about the South, though,
it can reasonably be suggested—and I would in fact suggest it on the
basis of recent experience such as China’s and India’s unavailable to
such observers—that a more thorough liberalization of the peninsula
would change the picture radically, despite the culture. After all,
when Italians moved to New York or London from the Mezzogiorno
in great masses, they did very well within a couple of generations.
Italian Americans had by the 1970s the third highest rate of univer-
sity graduation by ethnic group, third only to Jews and Irish—both of
whom in their turn had been despised as incorrigible in their inher-
ited culture. The optimistic case can be suggested, too, against
Bauer’s similar pessimism about poor countries more widely: when
Indians and Chinese moved to places in which they were permitted
to have a go, they also flourished, yet did not abandon their culture.
As Bauer in another mood had said.

Consider a radical liberal policy for Italy. If the Mezzogiorno
broke off from Italy, or was rudely broken off, and in particular
broken off from the massive subsidies it now receives annually from
Rome for its Rome-approved vote, and sat on its own bottom, a true-
believing classical liberal would expect it to prosper mightily. Sicily is
not inconveniently located for sea trade, for example, and its sons and
daughters in America have done exceptionally well.

After all, what is in effect foreign aid from the North, Bauer had
said in 1977 about “technological” free lunches to be given to India,
has the problem that it, like power, tends to corrupt. “When those
who have to pay for the technology [or in the Mezzogiorno’s case to
pay for an autostrada to nowhere] spend their own resources, they
are far more likely to purchase it [viz., the technology or the
autostrada] selectively and in accordance with considerations of costs
and feasible alternatives” (Bauer 1977: 154). The corruption from
free money is elementary price theory, denied by many economists
who, as Bauer elsewhere remarked in one of his stiletto footnotes,
believe that “acceptance of nonsense may be necessary for participa-
tion in political decisions” (Bauer 1975: 312n29).

Yet early and late Bauer emphasized the obstacles that culture
posed to the Great Enrichment of poor countries. He complains of
John Hicks’s economistic theory of economic history that “neither
religion nor any other belief is mentioned as influencing either conduct or social institutions” (Bauer 1971: 166). Bauer believed that caste in India and witchcraft in Africa posed major obstacles. Of India he wrote in 1961 for example, in a surprisingly conventional way, about “the contemplative, non-experimental, uncurious, and fatalistic outlook of large sectors of the Indian population, especially the rural population and certain sectors of the intelligentsia” (Bauer 1961: 26). True, in accord with price theory he hastened to add that “it should not be inferred . . . that the propositions of economics are irrelevant to India . . . . The Hindu peasant will not kill a cow, but he will sell his output where he can get the highest price” (ibid., 28). Yet at about the same time, by contrast, the Indian professor of English literature Nirad Chaudhuri (1959: 178) pointed out that Christian England was actually less profit-oriented in its prayer for daily bread than was the daily Hindu prayer to Durga, the Mother Goddess: “Give me longevity, fame, good fortune, O Goddess, give me sons, wealth, and all things desirable.”

The businessman and public intellectual Gurchuran Das notes that the second stage of a worthy Hindu life is that of the householder: “The dharma texts recognize the value of the second stage, which was the indispensable material basis of civilization” (Das 2009: xxxiv). Among the successive goals for a flourishing life in Hinduism is “a second goal . . . artha, ‘material well-being,’ which makes sense, for how can one be happy in conditions of extreme deprivation?” (ibid., xxxviii). How indeed?

Most social scientists in the 1950s and beyond looking at Holy India—and, I am saying, Bauer, too, despite his well-reasoned attacks on such a conventional view—saw only vicious circles of poverty (see Bauer 1965). During the 40 years after independence such a rhetoric of a Gandhi-cum–London-School-of-Economics socialism held the “Hindu rate of growth” to 3.2 percent per year, implying a miserable 1 percent a year per person as the population grew. Nehru wrote with satisfaction in 1962 that “the West also brings an antidote to the evils of cut-throat civilization—the principle

---

2 Also see “Money and the Englishman” (Chaudhuri 1959: chap. 5). Chaudhuri made his first trip to England after the Second World War.

3 See also Das (2000), which he regards as his sympathetic treatment of the “householder” stage.
of socialism. . . . This is not so unlike the old Brahmin idea of service” (quoted in Lal 2006: 166).4

At last, however, such anti-commercial rhetoric derived from European thought of the 1930s and “the old Brahmin idea of service” faded. A profiting and bettering rhetoric took root in India, partially upending the “License Raj,” as the Indians described the 44 years after Independence (Adhia 2010, 2013). A third of a million Indians subscribe to the fortnightly Indian magazine Business Today, founded in 1992, which contains breathless articles praising enterprise. And so India commenced, after liberal economists took charge in 1991, to increase the production of goods and services at annual rates shockingly higher than in the days of five-year plans and corrupt regulation and socialist governments led by students of Harold Laski. By 2008 Indian national income was growing at fully 7 percent a year per person. Birth rates fell, as they do when people get better off and therefore have access to birth-control devices.

After 1991 and Singh’s liberal allies, though, most of the culture didn’t change, and probably won’t change much in future. Economic growth, as the Japanese have long shown, does not entail becoming identical to Europeans. Unlike the British, the Indians in 2030 will probably still give offerings to Lakshmi and the son of Gauri, as they did in 1947 and 1991. Unlike the Germans, they will still play cricket, rather well. And in 2050, after merely two generations at the rates of growth possible for economies launching on the Great Enrichment by adopting liberal economic policies, average income will have risen by a factor of fully 16 over what it was in 2008. The level will then be well over what it was in the United States in, say, 2003. Even by 2050 in much of their talk and action the Indians will not have the slightest temptation to become like Chicagoans or Parisians, any more than the once appallingly poor southern Italians have taken on an American style of driving or a British style of food, though they are now by international standards rich. The Italians even of the Mezzogiorno did, however, adopt in part a northwestern European rhetoric about the economy, as the Indians have largely now. They entered the modern world, and the modern word, of a bourgeois civilization, and were made the better for it, materially and spiritually.

4One is reminded of the old and vulgar joke in which the farmer says, “When I hear the word ‘service,’ I wonder who is getting screwed.”
And most assuredly, I say again, their rich cousins in the United States, or United Kingdom, or Australia had done so.

Bauer insisted on cultural pessimism. In a review of the dismal effect of Marxism on theories of economic development, he declared, contrary to what “both Marxists and non-Marxists often believe,” that “men are obviously not equally endowed by nature in physical, intellectual or economic capacities” (Bauer 1975: 304). Yet such a truth considered individual by individual does not imply that groups are so radically unequal in, say, economic capacities. At the least they have plenty of outliers in their number with entrepreneurial tastes. Growth can occur, if given a liberal chance.

The worry about culture and the optimism about price theory create a persistent tension in Bauer’s work that one does not see for example in the more cheerfully optimistic work of his American ally I have mentioned, Theodore Schultz. For example, on the last page of Bauer’s *Indian Economic Policy and Development*, he says again, as he had said throughout the book, that “criticism of Indian economic planning . . . should not be mistaken for a plea for governmental inactivity in economic and social life” (Bauer 1961: 141). He says it, I suppose, to fend off the accusation of anarchism that has come so easily to the lips of indignant statists since the Great War.

Admittedly, then he immediately takes it back: “what is required in India is essentially a redirection of the activities of government, away from policies restricting the energies and opportunities of its subjects, and away from acts of emulation of the pattern of the Soviet world” (ibid.). The choice, he says in the last sentence in the book, was between “the development of the opportunities of the people” and “the establishment of a socialist society.” Yes. It is no accident that the optimistic part of Bauer’s advice started to be heeded only after the fall of the Soviet Union, a few years after my premature announcement of his “resurrection. The ideological veil over socialism’s inefficiencies and injustices put up after the Great War fell to the ground. Bauer was risen from the dead.

And Bauer was fiercely opposed to the notions of vicious circles of poverty such as Myrdal and Samuelson believed. In his 1975 essay his target becomes clearer, and the apparent tension I am pointing to is partially resolved. On the one hand he deprecates “the suggestion that the economic capacities of people are substantially equal, and differences reflect political manipulation or
exploitation” (Bauer 1975: 311). But he is claiming that the egalitarianism he does not favor is used to justify, he thinks, socialist and protectionist excesses. But he also makes the point, as I have emphasized, that overseas Indians do just fine (Bauer 1961: 28), which suggests that culture can’t be it. And he views as disastrous Indian policies such as minimum wages (ibid., 92–93) and central planning (chaps. 2–6, which is to say most of the book). “The large reserves of human energy and talent,” so evident to us now in the growth of India after 1991, were “inhibited by the restrictive forces of custom,” to be sure, but “enhanced [that is, made worse] at present [in 1961] by the restrictive effects of government policies”—the License Raj and the attempts to apply social democracy straight away.

That is, Bauer was not quite as much of an egalitarian, optimist, and thoroughgoing liberal as was, say, Adam Smith. Smith (1755) believed, “Little else is requisite to carry a state to the highest degree of opulence from the lowest barbarism, but peace, easy taxes, and a tolerable administration of justice; all the rest being brought about by the natural course of things” (quoted in Stewart 1812: IV, 25).

The Tension between Bauer and Hicks

It is quite typical of the imperfectionist and statist habits of economics since 1848—being in this quite unBauerish—that the political scientist Barry Weingast in quoting the famous sentence by early Smith adds the magic word for statists, “infrastructure”: “If peace, easy taxes and a tolerable administration of justice represent the market-supporting infrastructure necessary to sustain markets, just how does this infrastructure come about?”5 Pointing to the Lectures on Jurisprudence (1762–63), assembled from notes by Smith’s students and finally printed in 1896, Weingast replies, “Markets without legal infrastructure work poorly at best and fail to develop in the absence of contract enforcement, secure property rights, and the division of labor. No so-called invisible hand has transformed modern sub-Saharan Africa or South Asia into rich, developed countries.” Yet it is doing so now.

5Barry Weingast webpage on the Smith Project at https://web.stanford.edu/group/mcnollgast/cgi-bin/wordpress/adam-smith-project.
Weingast has argued on many other occasions that the visible hand of government supplies what growth needs. True, nonpredation by the very state is necessary, the “peace, easy taxes, and tolerable administration of justice” Smith spoke of. All of them are activities of government whose lack will indeed crush individual ingenuity. But to make out of this a claim that government must supply at first an “infrastructure” is to give to government an active role contrary to “all the rest being brought about by the natural course of things.” True, Bauer (1961: 12) emphasized in the brief introduction to Indian Economic Policy and Development that “economic development [does] not emerge directly from the operation of market forces.” Yes, law is necessary. But China for centuries had peace, easy taxes, and a tolerable administration of justice, but without the liberal regime allowing ordinary people to have a go that Smith was recommending.

John Hicks, whom Bauer criticized sharply in his review in 1971 of Hicks’ A Theory of Economic History (1969), believed that economic history “has a recognizable trend” (p. 7 of Hicks). Many economic and other historians assume it does have such a trend, of steady, gradual improvement. For instance, the group of excellent economic historians contributing to the Maddison Project do so, at any rate implicitly. They see English economic history of the past millennium as culminating, slowly, slowly, in the Industrial Revolution. But in fact the Great Enrichment, the follow-on to an industrial revolution not notably different from earlier efflorescences, as Jack Goldstone calls them, was an astonishing discontinuity, long, long after English law reigned, within occasional periods of peace among the quarrelsome British (Goldstone 2002). A Rise of the Market spread over centuries is a false explicandum. As archaeologists are beginning to discover from the earliest remains, we Homo sapiens exhibit markets. At least since the Middle Stone Age, an era receding in time with each new discovery, humans have imported shells for decorations and obsidian for spear points.

Again, Bauer agrees with Hicks, and with many others, that what Hicks calls a “custom and command” or “revenue” (for the lords) economy gave way in, say, the 17th century to a

---

\(^6\) I discuss his statist claims about the formation and protection of property in McCloskey (2017).
“mercantile” economy, and also agrees with Hicks that we at last seem to be returning to command and revenue, if not custom (Bauer 1971: 166–67). The chart of rise and fall resembles, with some adjustments in timing, Karl Polanyi’s (1944) “double movement,” from tradition to market to socialism. None of it, however, fits with the best historical research since Polanyi wrote. Close study of medieval peasants, for example, finds them, as Bauer would have expected in some moods and Theodore Schultz in all moods, acting rationally within their constraints (McCloskey 1976).

What was special, and discontinuous, was the abrupt rise of liberalism in the 18th century. In my trilogy on the “Bourgeois Era” (McCloskey 2006, 2010, 2016), I have chronicled the rise and its astonishing consequences. It inspired widening masses of people, formerly indistinguishable from the most tradition-bound peasants of India and China, to have a go. Innovation exploded after 1800 in places like Britain or the United States and then more widely: mechanical reapers, railways, steel ships, electricity, forward markets, steam presses, universities, the germ theory, automobiles, autobahns, airplanes, ball points, containerization, the pill, the computer, the internet. One can try to make the ingenuity endogenous to the economy by claiming that liberalism itself arose in part from the success of the early mercantile economy around 1700. It seems doubtful, considering that mercantile economies existed for centuries from Tlatelolco in Mexico to Osaka in Japan, and as Bauer observes from Phoenicia and Carthage to modern times, without any sort of liberalism springing up.

In other words, thanks to the rise of liberalism the optimism of Hicks about economic growth, which struck Bauer in his pessimistic moods as quite absurd, proved in the end to be correct. Hicks (1969: 157) spoke of “a couple of generations” as sufficing. He had signed on, it would seem, to the “analytical egalitarianism” that Sandra Peart and David Levy (2008) have traced to the 18th century social theorists, especially Smith. Bauer expressed vexation with such a hypothesis, here and at many places in his writings: “Hicks does not even so much as hint at possible differences in faculties, attitudes, mores and institutions anywhere in the world, in the past or in the present” (Bauer 1971: 171; his vexation led to a rare slip in his late-learned and usually amazing mastery of English, the redundancy of “even so much as”). Hicks had pointed to protectionist policies in
poor countries as the main obstacle to growth, to which Bauer responds indignantly that “it is surely naïve to suppose that their abandonment would invariably bring about early and substantial material progress” (ibid., 172). He announces an alternative hypothesis: “Recognition of the relevance of economic policies should not obscure the limits set by parameters usually regarded as non-economic.” In 1977 he disparaged again the “idea or assumption that individuals, groups and societies are approximately equal as potential economic performers” (Bauer 1977: 144).

Let us test it. The world’s laboratory for protectionism has been Latin America after Juan Peron in politics and Raul Prebisch in economic theory. Latin America has had plenty of apparent noneconomic limits, such as native traditions and swollen militaries. But when the experiments in freer trade were tried, they regularly brought about early and substantial material progress (Reid 2017). And the biggest experiment has been in China and India, with similar results. Hicks’s two generations do not look at all improbable set beside the history of the Asian Tigers, or the Celtic one, and above all China and India growing since 1978 and 1991 at 7 to 10 percent per year per capita. At 7 percent per year, of course, income quadruples in a generation of 20 years, and increases by a factor of 16 in merely two such generations.

Conclusion

No one would deny that deep ignorance as much as charming customs can obstruct the choices that Bauer put in the midst of his account of growth. But ignorance and custom are not always permanent. They can change, sometimes with startling speed, in which case the conditions that Bauer thought so sluggish can become suddenly favorable. And choice—the profit motive that even a mere consumer exercises when she is free—can overwhelm the ignorance and custom. That’s the dynamic extension of “price theory,” the Austrian-Hungarian dynamics of discovery by free people.

It is an odd feature, in other words, of Bauer’s courageous advocacy for ordinary people having a go, free from the arrogance of governmental planning and tariffs and industrial policy, that—sometimes—he was pessimistic on the cultural score.
References


Peter Bauer Was Not a Cultural Pessimist: Comment on McCloskey

Ian Vásquez

Deirdre McCloskey is an optimist, and rightfully so, when it comes to the development potential of poor countries. As she notes, the impressive records of India, China, and other countries in recent decades in achieving sustained, rapid progress justify that outlook. They also confirm the role of market liberalism in escaping mass poverty. She thus understandably admires Peter Bauer for his courageous advocacy of classical liberal policies at a time when that meant going against the grain in development economics.

McCloskey is bothered, however, by what she considers an oddity or a tension in Bauer’s work that at once recognizes the ability of poor people to create wealth when they are given the freedom to do so, and that also recognizes that some widespread cultural factors in some societies can adversely affect progress. She states that “Bauer insisted on cultural pessimism.” She compares his views to those of Banfield (1958, 1963) and Putnam et al. (1993) about southern Italy, which leads one to conclude, in McCloskey’s words that, “No reigning in of bad government policy could solve the puzzle of culture inherited from the past.”

Bauer did believe that noneconomic factors such as culture play an important role in development, but McCloskey pushes her point too far. He did not consider culture to be as powerful or permanent
an obstacle to progress as she depicts. His views were rather more complex and, I would argue, made him an optimist on development. To be fair, McCloskey also qualifies her criticism by citing instances where Bauer notes how the poor have responded rationally to incentives such as price signals to better their situation. She takes this as evidence of an inconsistency in his mood.

In a career as long and prolific as Bauer’s, it would be surprising if his views on some issues did not change or evolve. Bauer indeed admitted to having held mistaken opinions—for example, the belief in the primary importance of physical resources such as capital—especially early in his career. But he did not consider his views on the role of culture to be mistaken. On the contrary, he felt that the disregard of noneconomic factors was a major shortcoming of development economics and was impressed again and again with their impact on economic performance and the functioning of society more generally.

One of the reasons McCloskey may take issue with Bauer on his cultural views is that culture is a term that is often ill-defined or that may mean different things to different people depending on how the word is used. As I shall explain, Bauer’s conception is probably nearer to McCloskey’s understanding of cultural attitudes and changes than is at first apparent. Another reason for the difference of opinion is that McCloskey writes from the perspective of 2018, whereas Bauer began writing in the 1940s as a pioneer in a field that would largely disagree with him and even disparage him. This gave him both a different perspective on development than that of McCloskey and a different objective in his writings.

Bauer as an Applied Economist

Bauer was an applied economist who regularly warned against abstraction and favored direct observation in the study of economic development. He was skeptical of national income and investment statistics in poor countries because their compilation was methodologically flawed, and they missed significant economic activity. It is difficult to find national growth rates, for example, in his writings on development. Instead, he discusses broad transformations such as the move from subsistence to exchange or the fall in fertility rates, and he cites specific, major examples of progress. This was evident from his work in the 1940s and early
1950s in British Malaya, the Dutch East Indies, and West Africa. There he documented vast wealth creation on the part of the poor (Bauer 1948, 1954).

It was thus evident that the ordinary people of the ldes [less developed countries] were not necessarily torpid, rigidly constrained by custom and habit, economically timid, inherently myopic, or generally deficient in enterprise. In the space of a decade or two, the illiterate peasantry of South-East Asia and West Africa planted millions of acres to produce new cash crops; and rubber, cocoa and kola trees, for example, take five years to become productive. In all, this represented large volumes of direct investment made possible by voluntary changes in the conduct, attitudes, and motivations of numerous individuals, in many cases involving the sacrifice of leisure and the modification of personal relationships. Yet Malays, Indonesians and Africans were precisely among those who were depicted (as they still sometimes are) as incapable either of taking a long view or of creating capital, and as being hobbled by custom and habit [Bauer 1984: 5].

Bauer was not a cultural pessimist. It was a constant in his career to point out widespread examples of formally uneducated poor people contributing notably to economic progress. He cited scholars’ accounts of rapid economic change in the less developed world since the end of the 19th century to counter the prevailing view depicting poor countries as necessarily stagnant. He did so also to counter the “development orthodoxy” that assumed only state-led development could achieve high rates of growth.

Thus, when Bauer warned against lofty expectations of growth, he did so typically as part of a critique of forced industrialization and other dirigiste schemes. He and Basil Yamey lamented what they considered an ahistorical “belief, or at least the frequent statements, that most less developed countries can reach the level of material attainment of the West in a few years or decades” (Bauer and Yamey 1972: 681). Even with 21st century hindsight, their apprehension about such a pace of growth does not seem unjustified. The high growth that has occurred in many developing countries—which has resulted from market reforms rather than dirigisme—still has not brought many of those countries to Western levels, and in those that have become developed, the process took a couple of generations.
In that sense, McCloskey’s concern for what Bauer considered unrealistic rates of growth seems misplaced.

Noneconomic Factors and Cultural Change

It is true, as McCloskey claims, that Bauer drew attention to the bearing that culture had on development and that he often cited attitudes and customs that adversely affected material advance. But he viewed culture as one of many factors affecting development. “Economic achievement,” he noted, “depends on personal, cultural, social, and political factors, that is, people’s own faculties, motivations, and mores, their institutions, and the policies of their rulers. In short, economic achievement depends on the conduct of people and that of their governments” (Bauer 1991: 42).

McCloskey observes that “ignorance and custom are not always permanent.” She refers to “the conditions that Bauer thought so sluggish [which] can suddenly become favorable” and asserts that “choice . . . can overwhelm the ignorance and custom.” Yet Bauer would not disagree that culture can change, and that aspects of it can change rather rapidly. He provided numerous examples of such salutary changes that resulted from increases in market exchange. Indeed, he favored such conditions for cultural change precisely because they were more apt to succeed and because they avoided the coercion inherent in development planning. In the following passage, for example, he explains why—contrary to the prevailing orthodoxy—he approved of continued agricultural production in developing countries:

There are various reasons why in many poor countries a large measure of continued reliance on agriculture, notably on agricultural production for sale, is likely to represent the most effective deployment of resources for the promotion of higher living standards. One reason is the familiar argument in favor of comparative costs. Another, less familiar, reason is that production of cash crops is less of a break with traditional methods of production than subsidized or enforced industrialization. Agriculture has been the principal occupation in most of these countries for centuries or millennia. Thus in the production of cash crops the difficulties of the adjustment of attitudes and institutions in the course of the transition from subsistence production to an exchange or money economy are
not compounded by the need to have to acquire at the same
time knowledge of entirely new methods and techniques of
production. After some time spent on the cultivation of cash
crops, people find it easier to get used to the ways, attitudes
and institutions appropriate to a money economy. This greater
familiarity with the money economy facilitates effective indus-
trialization. In these conditions of transition from subsistence
to a money economy, conditions widely prevalent in poor
countries, production of cash crops and effective industrializa-
tion are thus complementary through time. The unfavorable
contrast often drawn between agriculture and manufacturing,
to the detriment of the former, is an example of a time-less,
unhistorical approach to economic development, an approach
which is inappropriate to the historical development of soci-
eties [Bauer 1972: 144–45].

Elsewhere he would note:

If people voluntarily give up their beliefs and modes of con-
duct to secure material advantages, this implies that they
value more what they receive than what they have given up.
The change, therefore, will not produce a resentful people.
Throughout the underdeveloped world, there are many
examples of large-scale voluntary adjustment to new oppor-
tunities [Bauer 1972: 202].

Bauer never tired of citing the strong, positive impact of trade on
progress. External contacts, he noted, “also engender a new outlook
towards material possessions and the means of obtaining them. And
perhaps most important, they undermine customs, attitudes and
values which obstruct material advance” (Bauer 1972: 301). Bauer
often referred to the ready development of economic attitudes favor-
able to the market and modernization. The kind of cultural features
and changes that Bauer describes to support development seem con-
sistent with the liberal values McCloskey advocates as supportive of
economic progress in a diversity of cultural settings.

Since Bauer never specified exact time spans nor growth rates
for his own expectations of poor countries’ development, it is difficult
to determine how far off his views on the potential pace of develop-
ing countries’ progress were from those of McCloskey. Bauer was
skeptical about anybody’s ability to engage in such predictions, so
he avoided doing so. He did, however, mention the long centuries
of institutional, technical and cultural development that Western Europe went through prior to its modern era of economic growth as a contrast to the developing world where such secular development had not taken place. Those features of his writing may explain in part why McCloskey interprets his views as pessimistic.

Another reason that might explain and to some degree justify McCloskey’s judgment on Bauer was his view about Indian culture. Bauer believed that Indian attitudes and customs inimical to growth, such as the caste system and the religious encouragement of begging, were unique to India and especially pronounced. He considered the need to change those cultural features “more difficult, necessary and urgent than elsewhere, notably more so than in the earlier history of the developed countries” (Bauer 1961: 29). Here, McCloskey is on more solid ground, and of course we have now observed high growth rates, especially in this century (and after Bauer’s death), and the erosion of the caste system. Yet even in this case, Bauer cited evidence of India’s capacity for economic advance and the abandonment of the caste system by Indians who emigrated to Malaya, suggesting the susceptibility to cultural change when exposed to increased economic opportunities.

Group Differences and the Complex Role of Culture

Bauer’s view on the role of culture in development was complex. He observed differences in qualities and performance among groups and individuals, the importance of entrepreneurship within culturally distinct societies, and the necessarily uneven spread of economic advance. The interplay of these differences affected economic and cultural change within and among different cultural groups. The most alert individuals and groups respond first to new economic opportunities such as those presented by outside trade. This produces both what some have called inequality and an incentive for many more to create wealth. As Bauer (1972: 196) explained, “Differences in incomes are more often a condition of material progress and evidence of its occurrence rather than an obstacle to

1 See, for example, Aiyar (2015).
2 Bauer (1972: 298) referred to these as “conspicuous differences between individuals and groups in economic aptitudes, such as industry, enterprise, curiosity and ability to perceive and exploit economic opportunity.”
its realisation. In Asia, as elsewhere, the prospects of material rewards have encouraged millions of people to work hard, save, experiment and invest, often in distant countries.”

Just as Bauer did not see an inherent problem with inequality of wealth or income, he was not bothered by differences in group performance. As a corrective to Bauer’s insistence on group differences, McCloskey implores that “Growth can occur, if given a liberal chance.” Bauer would fully agree. He believed that a government that adequately performs its limited, essential functions “usually contributes substantially to the voluntary transformation of mores and modes of conduct harmful to material progress” (Bauer 1972: 203). That in itself does not guarantee that different cultures or societies will produce the same outcomes under a liberal regime, but, again, since Bauer did not specify expected rates of growth or levels of development, we cannot be sure how much his view about the prospects of development really differs from that of McCloskey. I suspect that they are not as far apart as McCloskey paints them to be.

The complexity of Bauer’s beliefs about the impact of culture could also be seen in his critique of J. R. Hicks. Bauer’s main complaint about Hicks is that general theories of history are fanciful, and that Hicks’s version was so poorly constructed as to be “an unhelpful travesty of reality” (Bauer 1991: 175). With regard to the prospects of poor countries, Bauer was mainly bothered by Hicks’s account of the growth process rather than by predictions about growth rates.

Hicks envisages that the relationships between specifically economic activities and such factors as people’s attitudes and social institutions are largely one-way. He regards the causal relationship as running almost entirely from the former to the latter, even when the reverse relationship would seem to be far more significant [Bauer 1991: 185].

Elsewhere, Bauer also noted the critical interplay between non-economic factors and economic policies in determining economic outcomes. “The neglect of cultural and political factors,” he warned, “necessarily involves disregard of the reciprocal interaction between the familiar variables of economic analysis and these determinants of economic performance and progress” (Bauer 2000: 13). To Bauer,
institutions, culture, political regimes and policies all influenced each other in complex ways. He did not formulate a theory about that relationship (no one has), but he was way ahead of his time, and still is, in recognizing those reciprocal influences.

Surely, Bauer was right to bring attention to that complexity. It is all well and good for a country to adopt a liberal regime, but nothing guarantees that it will do so, or that once it does implement liberal policies, they will endure. Culture, prevailing institutions and other factors will play prominent roles in the adoption of economic policies and in turn will be influenced by those policies and the growth that they encourage or discourage. Liberalism has benefitted the world’s poor, as Bauer would have expected, but other developments during this era of globalization also vindicate Bauer’s views: the impossibility of predicting poor countries’ growth trajectories; erratic swings in policy and political regimes in many developing countries; the rise of populism in rich and poor countries; and the large variance in growth rates among developing countries.

Bauer does seem more pessimistic as cited by McCloskey when he says that it would be naïve to think that the abandonment of protectionist policies in developing countries “would invariably bring about early and substantial material progress.” Given Bauer’s lifelong advocacy of free trade and its benefits, that statement may seem surprising. But the key word here is invariability. Bauer (1972: 302) correctly cautioned that “External contacts by themselves are of course not sufficient to ensure progress if other factors are missing.” He cited cases from Latin America and North Africa to bolster his point. More recent cases abound and would include Mexico, which has had an open economy since the 1990s yet has managed to achieve only mediocre growth rates. Parts of its economy are quite modern and large parts are not. Ideology, politics, flawed institutions, and even custom have conspired to prevent much needed reform in a way that would not surprise Bauer.

I doubt also that McCloskey is surprised. All in all, the differences between McCloskey’s and Bauer’s views on cultural change and on culture’s impact on development may be more of degree than of kind.

Bauer later revised that essay to say, “But it is naïve to believe that more favorable policies would bring an early, worldwide industrial revolution” (my emphasis) (Bauer 1991: 176).
References


For many years, P. T. Bauer (1915–2002) was the “dissenter on development.” Among other dissents, he demonstrated that foreign aid was ineffective. Today that view is held in higher regard than in his day, thanks to numerous contemporary scholars, but Bauer is seldom recognized as its originator.¹ An exception is Nobel Laureate Angus Deaton, who acknowledged Bauer was right on foreign aid and population. In The Great Escape: Health, Wealth, and the Origins of Inequality, Deaton (2011: 273) writes that Bauer was the first to show that “the hydraulic approach to aid is wrong, and fixing poverty is nothing like fixing a broken car.”

¹ Bill Easterly remarked that “it is amazing how much of the research and thinking of my like-minded co-authors and me was anticipated decades ago by Bauer, without us realizing it. A not so obvious example of this is Bauer’s skepticism about investment and capital accumulation as a very important force in economic development, which people like Ross Levine, Lant Pritchett, and I have shared in several papers in the last decade.” This quotation comes from a private email to Ian Vásquez (2007: 208).
Peter Bauer was an eccentric figure. His father was a bookmaker in Budapest. Happily, “one of his clients suggested to him that his industrious son might benefit from a British university education, possibly Cambridge.” Having “no contacts” in that country, he “simply turned up” in Cambridge in March 1934 and “presented himself at half a dozen colleges.” His journey as an economist began when he entered Caius College. He had little English and “found it very difficult to follow the lectures of even ordinary conversation.” He “never read a book on economics or economic history before coming to Cambridge” (The Caian 1985: 33). Forty-nine years later he was raised to the peerage by Margaret Thatcher.

Basil Yamey has remarked that it was “largely by accident that Bauer’s interest turned” to development economics (Yamey 1987: 21). After working in “a London firm prominent in the Malayan rubber industry,” he “used a research fellowship to study that industry, and at the same time was commissioned by the British Colonial Office to prepare a report on rubber smallholdings” in Malaysia (ibid., 22). This research project resulted in his book The Rubber Industry (Bauer 1948).

While economists’ concern for “development” dates back to Adam Smith, “the term ‘economic development’ was rarely used before the 1940s” (Meier 1984: 6). From the beginning, Bauer was the dissenting voice within the field. A few years later the young Bauer was again summoned by the Colonial Office to study trading activities in West Africa with special reference to monopolistic tendencies. The result was a study of unusual length and scope for an empirical work of this kind: West African Trade (Bauer 1954). This work perfectly represents Bauer’s many talents. It is profoundly empirical: it showers the reader with facts. It is argumentatively sharp: Bauer refutes one economic fallacy after the other. It is learned and fully conversant with the history of economic thought. Bauer was a most perceptive reader of Adam Smith’s The Wealth of Nations, in a time when

\[2\] If Bauer today garners only limited appreciation, it may well be because he was a clear writer, who always aimed at a larger audience than technical economists. This strategy had its pay-offs. Yamey reports a rather amusing fact: novelist James Gordon Farrell commented that reading Bauer was most valuable to him in his attempts “to recreate the Far East of forty years ago” in his 1978 novel The Singapore Grip. Yamey notes, “This is a remarkable compliment to be paid by a novelist to an economist” (Yamey 1987: 22–23).
Smithian scholarship was not as ubiquitous as it is today. He was indeed a follower of the great Scot: Smith’s understanding of men as “trading animals” pervades Bauer’s thinking.

After the publication of *West African Trade*, Bauer found himself at the center of a battle surging around the concept of “economic development.” While undoubtedly a pioneer of these investigations, Bauer has long been considered a “fringe” thinker. This has clearly to do with his *politics*. To paraphrase Deirdre McCloskey, Bauer thought economic growth required that people should be allowed to have a go, to be at freedom to follow their intuitions and needs rather than being “nudged” in the direction of this or that particular productive effort by government masters.3

Bauer—and his long time coauthor and friend, Basil Yamey4—were that rare thing: development scholars with no ambition toward social engineering. This didn’t make them popular in a profession monopolized by wannabe central planners.

While Bauer’s views on foreign aid and development grew out of his field research, they retain an importance that goes beyond the scope of development economics. In fact, they account for a forceful refutation of the historically crucial concept of “primitive accumulation,” which was alas commonplace in development economics and informed the concept of the “vicious circle of poverty,” Bauer’s *bête noire*. As Curzon-Price (2002: 82–83) noted, “Although Marxism as a normative, prescriptive policy has failed . . . the positive Marxist assertion that the possession of wealth is the result of exploitation still holds great sway. It is doubtless the most durable of all the fallacies that Lord Bauer spent his long and distinguished career exposing.” This article aims to illustrate that point.

The Myth of Primitive Accumulation

The question of how industrial capitalism came about has been with us for a long time. As you need capital to have factories and

3McCloskey (2015) writes: “The main, and the one proven, social discovery of the 19th century [was that] ordinary men and women do not need to be nudged or planned from above, and when honored and left alone become immensely creative.”

4In interviews with the present author, friends of both testified that the cooperation between Bauer and Yamey was far more developed, constant, and important than their shared authorship of some works shows. It was indeed, for a good part of Bauer’s life, a relationship of intellectual camaraderie bordering on symbiosis.
industrial establishments, the idea of a “primitive” or original accumulation of capital seemed reasonable to many. Since a single capitalist needs somehow to get hold of resources to invest before moving on with his entrepreneurial plan, why shouldn’t the same be true for society at large? Shouldn’t a society be thriftily saving so that it can transfer those savings into capital which can in turn generate development? While there are obvious differences between persons and groups, this idea appeals to us as we know many important processes in life are indeed cumulative, starting with that piling up of notions after notions that we call “learning.”

Adam Smith is sometimes seen as the originator of the idea of “primitive accumulation.” In the introduction to the second book of The Wealth of Nations, Smith ([1776] 1981: 276–78) somewhat casually notes that an “accumulation of stock must, in the nature of things, be previous to the division of labour.” Smith’s point is that “in that rude state of society in which there is no division of labour, in which exchanges are seldom made, and in which every man provides everything for himself, it is not necessary that any stock should be accumulated or stored up beforehand, in order to carry on the business of the society.” But as labor becomes more and more subdivided, everyone needs to have “a stock of goods of different kinds,” stored somewhere, in order to concentrate exclusively on his or her own activities.

This vignette is familiar and commonsensical: when the savage “is hungry, he goes to the forest to hunt; when his coat is worn out, he clothes himself with the skin of the first large animal he kills: and when his hut begins to go to ruin, he repairs it, as well as he can, with the trees and the turf that are nearest it” (Smith [1776] 1981: 277). But when the civilized man who knows how to manage his trade, and little else, demands something, he must be able to rely upon other people’s produce. Coexisting labor is what makes, for Smith, a commercial society and offers the hope of economic growth. The more labor coexists (the more extended the market is), the better it is.

In spite of Wikipedia’s claim that “Adam Smith’s account of primitive-original accumulation depicted a peaceful process, in which some workers laboured more diligently than others and gradually built up wealth, eventually leaving the less diligent workers to accept living wages for their labour,” Smith is hardly an originator of the idea of primitive accumulation. The biblical Parable of the Talents (Matthew 25:14–30), for example, speaks of accumulated wealth to be invested.
But for Karl Marx it was very important to assume that Smith was the originator. Marx ([1890] 2002: 2068) begins his treatment of the subject by pointing out that “original [i.e., primitive] accumulation plays in Political Economy about the same part as original sin in theology.” Primitive accumulation plays the role of a foundational myth justifying that “through painful toil you will eat food from it all the days of your life” (Gen 3:17). In a note to Chapter 7, Marx makes fun of British economist Robert Torrens who, “by a wonderful feat of logical acumen . . . has discovered, in this stone of the savage the origin of capital.” He quotes Torrens noting, “In the first stone which he [the savage] flings at the wild animal he pursues, in the first stick that he seizes to strike down the fruit which hangs above his reach, we see the appropriation of one article for the purpose of aiding in the acquisition of another, and thus discover the origin of capital” (ibid., 2375).\(^5\)

And yet the concept may in fact have come from Marx. For him, primitive accumulation is key to “proving” that the process, therefore, that clears the way for the capitalist system, can be none other than the process which takes away from the labourer the possession of his means of production; a process that transforms, on the one hand, the social means of subsistence and of production into capital, on the other, the immediate producers into wage labourers. The so-called original accumulation, therefore, is nothing else than the historical process of divorcing the producer from the means of production. It appears as “original,” because it constitutes the pre-historic stage of capital and of the mode of production corresponding with it [ibid., 2072–73].

Primitive accumulation is basically robbery on the part of future capitalists (or their great-grandfathers) of workers’ properties. In fact, the theory is a way to attribute to the bourgeoisie a series of events, from the colonies and the protectionist system to trading monopolies to enclosures. For Marx, “The treasures captured outside Europe by undisguised looting, enslavement, and murder, floated back to the mother-country and were there turned into capital” (ibid., 2177–78).

\(^5\)While Marx made of Torrens a subject of mockery, McCloskey (2016: 104) notes that Acheulian hand “axes” were indeed accumulated in massive amounts by *Homo erectus*. 
As Deirdre McCloskey (2010: 156) pointed out, “it is not a good business plan” to steal from poor people. “Stealing from poor people” cannot explain economic growth, whether one steals from the poor living in her own nation or from people living in other nations.

Primitive accumulation plays a powerful role in Marx’s theory. It is perhaps one of its key concepts, as Marxism is first and foremost a theory of history. And how could you really justify the “expropriation of the expropriators,” if the latter didn’t expropriate all that much?

But the consequences of this idea have extended well beyond the boundaries of Marxism. As Alexander Gerschenkron (1957: 33) noted, the vignette of primitive accumulation demands us to believe in “an accumulation of capital continuing over long historical periods—over several centuries—until one day the tocsin of the Industrial Revolution was to summon it to the battlefields of factory construction.”

What kind of thief robs masses of stuff, saves it meticulously and waits for generations, relying on the power of compound interest, until the proper investment opportunity pops up? How farsighted such expropriators would be!

**Primitive Accumulation and the Vicious Circle of Poverty**

No matter how unrealistic, some version of this view had its triumph in development policy. Particularly after decolonization, significant amounts of foreign aid were transferred to poorer countries, with the aim of growing their stock of capital. There is no question that higher capital investment would make workers more productive, and higher productivity would make them more prosperous, if the capital were invested intelligently. But the key issue here is that it was thought that massive transfers were needed in poorer countries to equip them with the “proper” amount of capital required to make growth possible at all.

The often mysterious way in which ideas travel makes it by no means necessary to be an orthodox Marxist to advocate foreign aid. But the concept of primitive accumulation was the backdrop against which the advocacy of foreign aid was staged. I would rely here on prominent socialist commentator, and likely Soviet spy, Harry Magdoff: “For many of us in the field of economic development, this issue of accumulation, from the standpoint of capital or money, or
reserves, has become the most important issue because of this discussion. Because this debate applies to underdeveloped countries and to economic development it is a perfectly proper question to ask” (Magdoff 2013).

The period following World War II was the zenith of the Keynesian consensus. Keynesian economics, by legitimizing ever-growing “state interventionism, provided many developing countries with a theoretical justification for a more active role of the government in the economy and emphasized the accumulation of capital as the foundation of economic growth” (Osiatynsky 1993: 185). This appropriation of Keynes is a bit of a paradox, as Keynesian economics, if considered charitably, aimed chiefly at counteracting cyclical fluctuations, not at preparing the seed corn for economic development in preindustrial countries. Postwar theories of industrialization seem to be indebted to Marxist ideas in more than one respect. The consensus against which Bauer battled was centered around the idea that underdevelopment stimulates “circular and cumulative processes” that consolidate economic backwardness and make it impossible to break out of the “vicious circle of poverty” (Bauer 1965).

Indeed Ragnar Nurkse, a pioneer of development economics, framed the issues with poverty as “Problems of Capital Formation in Underdeveloped Countries” (Nurkse 1953). For Nurkse, the challenge for poorer countries could be reduced to the question of achieving a sufficient rate of capital accumulation. Maurice Dobb (1951: 7) claimed that “we shall not go far wrong if we rear capital accumulation . . . as the crux of the process of economic development.”

There is more in this approach than the commonsensical view that higher levels of capital allow for greater productivity. As Bauer himself remarked, “Insistence on the vicious circle of poverty and

---

6 In a way, in his famous The End of Laissez-Faire, Keynes himself recognizes the merit of free market policies in “earlier” states of development, posing the questions of managing population and the business cycle in the context of a developed, industrial economy (Keynes 1926).

7 This commonsensical view tends to assume that capital is employed in profitable endeavor, somehow presupposing those features of a free market that make for a sensible allocation of resources (though by no means impermeable to mistakes).
on the stagnation of the underdeveloped world has promoted the flow of foreign aid, which is a major object of policy for many people” (Bauer 1965: 48). It makes sense if you assume that lack of capital is so severe that local populations would never bridge the gap, either by thrift or by saving and investing the profits of their trade. On the other hand, the success of such an accumulation depends crucially on employing capital just where it is needed—pouring money can hardly produce happy, entrepreneurial endeavor. The assumption thus shows clear marks of the Marxist idea that industrialization was possible only because of centuries of robbery and exploitation.

The “consensus approach” that Bauer battled against was best summarized by the slogan: “Aid, not trade.” Here lies the idea that fostering development through trade, an approach that was reminiscent of the 18th and 19th century theory of doux commerce, of commerce breeding brotherhood and civilization, is only a hypocritical cover-up of economic exploitation on the part of the rich.

Bauer himself knew and pointed out that the “current orthodoxy” in development economics was heavily indebted to Marxist-Leninist views. So did Karl Brunner, who underlined how “the Leninist extension of the Marxian vision, with its stress on imperialism and the political ‘avant-garde,’ has found a particular echo in the Third World” (Brunner 1978: 7). “Whatever the exact process of their intellectual derivation, these views are widely and frequently expounded by well-known writers not regarded as Marxist or Leninist” (Bauer 1976: 173).

Bauer (1976: 165) traced such influence back to the following four key concepts:

First, that the underdeveloped world is not only desperately poor but stagnant or even retrogressing; this notion is the current version of the doctrine of the ever-increasing misery of the proletariat. Second, that the exploitation of undeveloped by developed countries is a major cause of this poverty . . . . Third, that political independence is meaningless without economic independence; this is an extension of the suggestion that political freedom and representative government are meaningless
under capitalism. Fourth, that comprehensive development planning is indispensable for economic advance . . . and especially for the industrialisation required for material progress.

The first two concepts are obviously related to the idea of the primitive accumulation of capital. Bauer opposed with particularly strong emphasis the idea that “contact with advanced economies is damaging to underdeveloped countries” (Bauer 1957: 65). The view has a Marxist background. In Marxist jargon, “The most important source of primitive accumulation, other than the exploitation of the peasantry, lay overseas in the exploitation of pre-capitalist societies. The process took a number of forms, principally trade, plunder and slavery” (Ure 1975: 29, emphasis added). In this scheme, the supposedly “unequal” exchange of values on the international market between highly productive “First World” countries and far less productive “Third World” countries is a mechanism of wealth transfer from the poor to the rich. It is consistent with that basic intuition which lies at the roots of Marxism: that inequalities signal exploitation.

The core of the argument is well known. Primitive accumulation in England, with enclosures, resulted in the transformation of peasants into future industrial workers: “the agricultural people” were “first forcibly expropriated from the soil, driven from their homes, turned into vagabonds, and then whipped, branded, tortured by laws grotesquely terrible, into the discipline necessary for the wage system” (Marx [1867] 2002: 2129). A similar process continues internationally. As Rosa Luxemburg ([1913] 1951: 365) remarked, “Capitalism in its full maturity also depends in all respects on non-capitalist strata and social organizations existing side by side with it. . . . Capital needs the means of production and the labour power of the whole globe for untrammelled accumulation; it cannot manage without the natural resources and the labour power of all territories.” The capitalist system could therefore not survive in the absence of foreign trade, as foreign entanglements with non-capitalist countries are providing it with the bulk of the resources it needs.

Luxemburg’s views were not unanimously accepted by contemporary Marxists. It ought to be remembered that major attempts to

---

9Historical research suggests enclosures made for population growth and better hygiene—reducing diseases that spread easily, as all the animals of the village were turned out on the commons (see Philpot 1975).
update the theoretical underpinning of Marxism, including the substitution of the concept of the falling rate of profit by a law of increasing surplus, were made when capitalism entered its “monopolistic” phase (Baran and Sweezy 1966). But the new “dependency theory,” intertwined with the Marxist understanding of monopoly capitalism, had nonetheless very similar normative suggestions. For Baran (1957), capitalism caused the misery of the developing world, as international oligopolistic firms impeded the growth of that surplus, which forms locally in poor countries and otherwise would be spent for local development.

These views, as Bauer knew well, had momentous influence and particularly impressed an entire generation of development experts. Bauer distinguished between Marx and his followers: Marx accuses “the metropolitan countries of plundering the colonies but he also regarded them as a progressive force in promoting modernization” (Bauer 1976: 164). It was Lenin and subsequent Marxists’ analysis of imperialism that “extended” Marx’s influence in the developing countries and attributed “the poverty of underdeveloped countries to exploitation by advanced countries” (ibid., 167). For it is not difficult to detect, as Bauer did, “a drop of Leninism” in the General Principle Number Fourteen of the First United Nations Conference on Trade and Development (UNCTAD): “the liquidation of the remnants of colonialism in all its forms is a necessary condition for economic development and the exercises on sovereign rights over natural resources” (Bauer 1975: 303).

The idea that economic development should come with “economic sovereignty” is at the intersection of socialism and nationalism, as they both tend to dismiss as propaganda the possibility of free trade being mutually beneficial. Such a view, though in milder forms, profoundly influenced the post–World War II consensus, which considered “external trade . . . at best ineffective for the economic advantage of less developed countries (LDC), and more often . . . damaging.” It subscribed instead to the belief that “the advance of LDCs depends on ample supplies of capital to provide for infrastructure, for the rapid growth of manufacturing industry, and for the modernization of their economies” (Bauer 1984b: 1). It builds on an understanding that

external factors [were] generally responsible for the poverty of LDCs, an example of the ancient and widely entertained
fallacy that economic activity is a zero-sum game, that the incomes of individuals, especially of the relatively prosperous, are somehow extracted from others, rather than representing a return for services performed. This notion . . . long antedates Marxism, but its influence has been reinforced by Marxist ideology in which property incomes imply exploitation and service industries are regarded as unproductive [Bauer 1975: 299].

Bauer versus the Vicious Circle of Poverty

Bauer’s first investigations in development economics were not “inspired by these topics” (Bauer 1984b: 1) His first works were research “boots on the ground” in Malaysia and West Africa. But in both cases,

even before setting foot in South-East Asia and West Africa I knew that their economies had advanced rapidly (even though they were colonies!) . . . it required no instruction in development economics to know that before 1885 there was not a single rubber tree in Malaya nor a single cocoa tree in British West Africa. By the 1930s rubber, cocoa and other export crops were being produced on millions of acres, the bulk of them cultivated by non-Europeans [Bauer 1984b: 2].

Bauer thus found in Malaysia and West Africa economies that, though “backward” as seen from the vantage point of European economists, were in fact dynamic. “A developed infrastructure was not a precondition for the emergence of the major cash crops of South-East Asia and West Africa,” he noted, recognizing that the historical experience he came to know “was not the result of conscription of people or the forced mobilisation of their resources” (ibid., 4). This is the crux of the argument, as the Marxist idea (and its modern derivations) assumes that such labor was indeed conscripted.

Bauer was inoculated from the Leninist idea that “monopoly” capitalism was the landmark of imperialism by studying competition, monopolization, and the effect of marketing boards in the regions. His conclusion was that the economic landscape was far richer and more complex than it was typically understood to be.

By reading Bauer’s West Africa Trade, which is in many ways his most ambitious empirical work, one immediately sees that Bauer
had a supreme understanding of a principle Vilfredo Pareto (1896) made central in his theorizing—namely, the mutual interdependence of economic (and social) phenomena. Those phenomena are complex and tend to eschew mono-causal explanations. President Harry Truman apparently once asked to be sent a one-armed economist, being tired of economists relentlessly saying “On the one hand, this” and “On the other hand, that.” Peter Bauer had two hands and used them both. Instead of feeling content with speculations upon the immediate, proximate, and deliberate consequences of any policy, he endeavored to understand, in Frédéric Bastiat’s famous phrase characterizing the true economist, “what was not seen.”

For Bauer (1979: 48), would-be central planners tended to practice “economics without prices and costs.” They focused on grand narratives of top-down economic development, neglecting the nuts and bolts of real economic life. His opposition to foreign aid pretending to foster capital accumulation in less-developed countries was built upon a rich understanding of the interconnection of different phenomena.

For one thing, focusing on capital accumulation was a “one-armed” way of looking at growth. “In the planning literature, especially in recent Marxist-Leninist literature, economic advance is usually defined without reference to general living standards but primarily in terms of industrial development. . . . The advocates of large-scale industrialisation, and especially of massive development of heavy industries, hardly ever refer to prices, incomes, cost, demand or standard of living” (Bauer 1976: 172).

It is not that large government-to-government transfers (i.e., foreign aid) produce no beneficiaries. Indeed, foreign aid may enrich specific subgroups of a population, and perhaps it could even trigger developments and growth in a particular field. The mistake is to equate some particular development (i.e., industrialization) with growth of the entire economy, meaning rising standards of living. “Of course, state planning can augment the resources available to particular sectors of the economy, by expanding these at the expense of other activities. But this has nothing to do with the expansion of the economy as a whole. This is obscured in the ubiquitous practice of identifying the output or progress of one sector with that of the economy as a whole; this practice is followed by economists and officials who neglect costs” (Bauer 1977: 149).
Capital fundamentalism, or infrastructure fundamentalism, may obscure more relevant facts. “If all conditions for development other than capital are present, capital will soon be generated locally, or will be available to the government or to private businesses on commercial terms from abroad, the capital to be serviced out of higher tax revenues or from the profits of enterprise. If, however, the conditions for development are not present, then aid—which in these circumstances will be the only source of external capital—will be necessarily unproductive and therefore ineffective” (Bauer 1976: 97).

Neither can we assume that “more capital” is per se a panacea. As Bauer pointed out in West African Trade, “the comparative lack of local technical and administrative skill aggravates the effects of the scarcity of equipment; it is not lack of capital alone which retards development” (Bauer 1954: 13). In other words, without the necessary human capital we can hardly assume machineries (or roads) to increase productivity per se. And how can we know, ex ante, whether those skills and a certain capital will or will not match? As observed by Ben Powell (2014: 130), “The right capital is the capital that best complements the existing capital and labor to produce the greatest value for society in the ultimate production of consumer goods. To find out what capital is best, we need the market’s competitive process to operate.”

This is something well known to anybody who ever visited the South of Italy. In the aftermath of World War II, aid from the North financed businesses’ investment (mostly by northern businesses) in ostensibly productive facilities, some of which were quickly abandoned as this river of transfers drained. Such factories located in regions with little industrial history and unskilled employees were named “cathedrals in the desert”: temples waiting for a faithful people, who were very unlikely to show up. Italy, a country where per capita income has been and still is diverging between North and South for virtually all of its history, is a microcosm of aid failure as sketched by Bauer. In this sense, his lesson can be further generalized: development can’t be forced by transferring resources, within or beyond the boundaries of the nation state.

P. T. Bauer’s bête noire was the so-called vicious circle of poverty. For him, the vicious circle of poverty could not really be a circle, because rich countries did exist and all of them “started poor, with low incomes per head and low levels of accumulated capital, that is with the economic features which now define underdeveloped
countries” (Bauer 1976: 165). Yet countries “have advanced, usually without appreciable outside capital and invariably without external grants which would have been impossible according to the thesis of the vicious circle of poverty and stagnation.”

For Bauer the foremost sin of foreign aid is that it cannot avoid being politicized: “the granting of foreign aid necessarily draws the donor country into the internal politics of the recipient country” (Bauer 1961: 120). In particular, such meddling with grantee-countries grew the anti-market sentiment in those countries, because “what the Third World learns from the West, or about it or about present and past economic relations between the West and Third World countries, comes from or is filtered through opponents of the market” (Bauer 1978: 172). The “devaluation by intellectuals of voluntary exchange” (McCloskey 1987: 250) can’t help economic growth.

In a sense, Bauer pointed to what might be called a “vicious circle of anti-capitalism.” Lack of prosperity in less developed countries is interpreted, in more or less openly Marxian fashion, as the sum of damages inflicted by Western action (trade) or inaction (lack of aid). This sense of guilt of Westerners creates the political demand for foreign aid, which Bauer (1976: 115) noted, “is a process by which poor people in rich countries help rich people in poor countries.” Donor-countries may bet on the importance of fostering capital formation, but they tend to ignore the circumstances of grantee-countries. Foreign aid thus doesn’t drag the poor out of poverty, and thereby the circle starts anew.

For this reason, Bauer gallantly opposed what he regarded as a pernicious form of egalitarianism, even when practiced by the Catholic Church (Bauer 1984a). For this reason, he opposed the corruption of language by the means of dangerous metaphors. “Nation-building,” for example, implies considering people “lifeless bricks, to be moved by some master builder” (Bauer 1984b: 5). Likewise, the use of the expression “Third World” implies that all underdeveloped countries are alike. In fact, “without foreign aid initiated and organized by the West, there would be no Third World or South” (Bauer 1980).

---

10This passage seems reminiscent of Smith’s famous description of the “man of system” who imagines that “he can arrange the different members of a great society with as much ease as the hand arranges the different pieces upon a chessboard” (Smith [1759] 1984: 380–81), with which an avid reader of Smith such as Bauer was certainly familiar.
Besides being a sharp critic of his contemporaries, Bauer advanced his own theory of economic development—though not in a systematic fashion. As Jim Dorn (2002: 357) argued, Bauer sees “the essence of development” as “the expansion of individual choices”:

I regard the extension of the range of choice, that is, an increase in the range of effective alternatives open to people, as the principal objective and criterion of economic development; and I judge a measure principally by its probable effects on the range of alternatives open to individuals [Bauer 1957: 113].

In this context, what truly is key for Bauer is voluntary exchange. His view of economic growth is essential Smithian: “the division of labour is limited by the extent of the market,” the more division of labor the better, the more extended the market the better. Very aptly his last book was titled, after one of its collected essays, From Subsistence to Exchange, because these are the two poles of development in Bauer’s view. “Advance from subsistence production involves trading activities” (Bauer 2000: 8), and the growth of such activities produces, little by little, more knowledge and better coordination, thereby making possible increasing production.

Such a process, such a transition, is fostered by trading at all levels. Bauer appreciated, as few did, that the poor are great traders. If we need to summarize his thought in a few words, the best way to do it would be to point out that he didn’t believe the poor were stupid. He thought we need to understand the specificities of local circumstances before proposing all-encompassing public policies. He thought culture had a great role in development (and anti-market prejudice could stifle it), but he didn’t think that the “South of the world” was populated by irrational economic agents, in desperate need of our enlightened guidance. That was his view, whether he wrote about development, migrations, or so-called overpopulation. He never practiced paternalism, that eternal habit of the intellectuals.

Conclusion

“Through the 1950s and 1960s Peter Bauer’s writings frequently aroused anger, if not apoplexy, but little reasoned criticism. He was ignored, dismissed, but not answered” (Desai 1982: 291). Bauer was
indeed “a hero of the market revolution” (Dorn 2002: 355) when socialism seemed invincible. His ideas were never popular within the economics profession (Vásquez 2007: 208-9) and, though they were vindicated as times went by, they were very rarely recognized as path-breaking.

And yet path-breaking they were. By challenging the concept of a vicious circle of poverty—that is, by debunking capital fundamentalism—Bauer inflicted a strong blow to its Marxian underpinning. His analysis has an even wider relevance than his “dissent on development.” He made it clear that coercive transfers of wealth could not be justified with an appeal for economic development—nor should they, as they could not be seen as mending previous expropriations. By emphasizing the importance of trade, Bauer made his way back to the economics of Adam Smith.

References

P. T. Bauer


Determining State Preferences for the Electoral College: 1788–2016
Paul E. Godek

This article reviews some of the basic institutional elements and empirical regularities of the U.S. electoral college system. Historical election results confirm the expected dominance of the winner-take-all allocation of electors within each state. Of greater interest is the array of state preferences for the electoral college given a national popular vote as the contemplated alternative.

It is necessary to be clear about what is meant by a state’s preference. The concept explored here is sometimes referred to as a state’s a priori preference. A state prefers a priori the system that results in the greater potential influence of its voters, as represented by the state’s popular-vote majority, on the outcome of an election. In this context the term “a priori” means in consideration of only the system itself, without (a posteriori) considerations of actual or likely outcomes in particular elections (see Felsenthal and Machover 2004.) In addition, as discussed below, a state can be said to prefer an institution only in consideration of and relative to a specific alternative institution.

The analysis here demonstrates (contrary to the existing academic literature) that well over half of the states—in particular the smaller
The general consensus in the academic literature appears to be that the electoral college system reflects and maintains the greater influence of large states. For example, see Banzhaf (1965, 1968), Rabinowitz and MacDonald (1986), Strömberg (2008), and Miller (2009, 2012).

Comprehensive and relatively recent (since 1990) treatments of the history, purpose, and effects of the electoral college are found in books by Abbot and Levine (1991), Kuroda (1994), Longley and Peirce (1996), Gregg (2001), and Ross (2012). Helpful overviews and bibliographies can also be found in articles by Kimberling (1992), Boylan (2008), Strömberg (2008), and Miller (2012). The original texts and documents are The Constitution of the United States of America as Amended (Article II, Amendment XII, and Amendment XXIII), The Federalist Papers (39 and 68), The Records of the Federal Convention of 1787, and James Madison’s Notes of Debates in the Federal Convention of 1787.


The relationship between state population and electoral college preference derives from the fundamental methods—some described in the Constitution, others in century-old laws—for determining the number of congressional representatives (and consequently the number of electoral votes) assigned to each state. In addition, the inverse relationship between state size and electoral college preference, as well as the majority-of-states preference for the system, have existed throughout the entire history of U.S. presidential elections. Those characteristics are certain to persist for the foreseeable future.

An Overview

The election of the U.S. president is determined by a simple majority vote of the electoral college, the members of which are appointed by the states. The Constitution directs that the number of electors for each state shall be “equal to the whole Number of Senators and Representatives to which the State may be entitled in the Congress.” The Constitution also directs that the methods for appointing electors and for directing how those electors cast their votes are to be determined by state legislatures. At first, electors

1The general consensus in the academic literature appears to be that the electoral college system reflects and maintains the greater influence of large states. For example, see Banzhaf (1965, 1968), Rabinowitz and MacDonald (1986), Strömberg (2008), and Miller (2009, 2012).

2Comprehensive and relatively recent (since 1990) treatments of the history, purpose, and effects of the electoral college are found in books by Abbot and Levine (1991), Kuroda (1994), Longley and Peirce (1996), Gregg (2001), and Ross (2012). Helpful overviews and bibliographies can also be found in articles by Kimberling (1992), Boylan (2008), Strömberg (2008), and Miller (2012). The original texts and documents are The Constitution of the United States of America as Amended (Article II, Amendment XII, and Amendment XXIII), The Federalist Papers (39 and 68), The Records of the Federal Convention of 1787, and James Madison’s Notes of Debates in the Federal Convention of 1787.

3Article II, Section 1 of the Constitution states in part: “Each State shall appoint, in such a Manner as the Legislature thereof may direct, a Number of Electors, equal to the whole Number of Senators and Representatives to which the State may be entitled in the Congress.” Two amendments to the Constitution relate to the electoral college. Amendment XII (ratified in 1804) clarifies several aspects of the electoral college system (in particular requiring each electoral vote to indicate choices for both a president and a vice president, as well as clarifying the method of selecting a president in the event that no candidate obtains a majority of the electoral votes). Amendment XXIII (ratified in 1961) assigns three electoral votes to the District of Columbia.
were selected directly by state legislatures, but the transition to some form of popular vote to decide the assignation of electors proceeded steadily. By 1836 almost all states had changed from state legislature selection of electors to some form of popular vote selection (see Ross 2012: chap. 2; Miller 2012).

The first electoral college, for the election of 1788, consisted of 69 electors across the 10 states that by then had ratified the Constitution. The electoral college increased in size monotonically, in accordance with the increase in the number of senators and representatives, reaching its current size of 538 with the election of 1964. The number of senators has been at 100 since the admission of Hawaii and Alaska to the Union in 1959. The number of representatives has been capped at 435 since the Apportionment Act of 1911.4 The District of Columbia has neither senators nor representatives, but was granted three electoral votes in accordance with Amendment XXIII to the Constitution, ratified in 1961. Hence, the current total of 538 electoral votes.

Because every state has two senators and at least one representative, every state has at least three electoral votes. As of 2016 each of the seven smallest states, as well as the District of Columbia, has three electoral votes; the largest state, California, has 55. The median is eight electoral votes. Unless a new state is added or Congress revises the number of representatives (something it has not done in over 100 years), the current total of 538 electoral votes will prevail.

Unanimity versus Maine and Nebraska

Regardless of who was doing the selecting—state legislatures or some subset of the population—electors have been directed, with few exceptions, to cast their state’s electoral votes for president

---

4The 435 maximum was further codified by the Permanent Apportionment Act of 1929. (With the admission of Alaska and Hawaii to the Union the number of representatives increased temporarily to 437, until the subsequent reapportionment.) Each state has at least one representative and the remaining 385 (= 435–50) are apportioned among the states as a function of their populations. State populations are determined by the decennial census, as prescribed in the Constitution: see Article I, Section 2. Reapportionment of the House also occurs every 10 years, as prescribed in the Constitution: see Article I, Section 3.
unanimously: winner take all. From 1788 through 2016, there have been 58 presidential elections and 2,238 state electoral vote tabulations for the office of president of the United States; 2,186 of the 2,238 tabulations, 97.7 percent, have been unanimous. The lack of unanimity in the 2.3 percent of vote tabulations could have resulted from an explicit nonunanimous allocation method chosen by a state, or from a so-called faithless elector: one who casts a vote in violation of a state’s voting rules.\textsuperscript{5}

In any case, the dominance of a winner-take-all allocation should not be surprising. Once the majority’s preference is determined, that majority would presumably want to maximize the state’s influence on the final outcome, which is accomplished with the winner-take-all allocation of electoral votes.

Nonetheless, Maine (four electoral votes) and Nebraska (five electoral votes) have opted out of winner take all, in favor of the “congressional district” method. Under that method the popular vote winner in each congressional district receives that district’s electoral vote; the popular vote winner in the state receives the two “senatorial” electoral votes. Maine adopted the congressional district method in advance of the 1972 election and Nebraska in advance of the 1992 election.\textsuperscript{6}

When a state splits its electoral votes (as both Maine and Nebraska have done), the influence of the state’s majority on the outcome of the election is diminished. Thus, whatever the reasons for those two states to have adopted the congressional district method, it is not likely that their example will be followed by any significant number of other states. In sum, the obvious benefit of unanimity explains why so few states have ever chosen the congressional district or any other nonunanimous method for the allocation of their electoral votes.

\textsuperscript{5}Of course, states with nonunanimous rules can also generate unanimous outcomes. Seldom have states adopted such rules, however, as discussed below. The National Archives and Records Administration reports the electoral college tabulations for every presidential election. The tabulations can be found at www.archives.gov/federal-register/electoral-college/votes.

\textsuperscript{6}I have seen conflicting accounts of whether any other states have ever adopted nonunanimous elector allocation methods, but if so those methods were subsequently rescinded.
Electoral College Influence and Popular Vote Influence

Let $e_i$ represent the electoral votes in state $i$, which are all allocated to the winner of the popular vote in the state (with few exceptions, as noted). Thus, each state’s proportional contribution to the electoral vote is:

\[
E_i = \frac{e_i}{\sum_j e_j} = \frac{e_i}{538}
\]

conditional only on the winner of the popular vote in the state. I will use $E_i$ as the measure of a state’s electoral vote influence. (Other measures of influence are considered below, but as a starting point I focus on shares.)

Now consider the distribution of popular vote influence across the states. At first, a state’s popular vote influence might seem to be an artificial construct. Each vote counts the same toward the national popular vote regardless of the voter’s place of residence. Nonetheless, the state remains the relevant locus because whether a state prefers the electoral college can only be determined by the state and in consideration of an alternative. In particular, I analyze the most commonly suggested and plausible alternative—namely, whether a state would prefer to have the election determined by the electoral college or by the national popular vote. To answer that question it is assumed that a state will prefer the system that maximizes the potential influence of its voters, as represented by the state’s popular vote majority, on the outcome of an election.

With regard to a state’s popular vote influence, it will help to define some basic measures. Start with two political parties, designated as the $D$ party and the $R$ party.

\[
\begin{align*}
\nu_i^D &: \text{votes for the } D \text{ party in state } i \\
\nu_i^R &: \text{votes for the } R \text{ party in state } i \\
m_i &= |\nu_i^D - \nu_i^R| : \text{the positive vote difference (the margin) between the parties in state } i, \\
& \quad \text{which is accorded to the winning party} \\
t_i &= \nu_i^D + \nu_i^R : \text{the vote total in state } i
\end{align*}
\]
For each state the popular vote margin in percentage terms is the difference in the vote between the two parties, relative to the total votes cast:

\[ \alpha_i = \frac{m_i}{t_i} : \text{state } i \text{'s percentage popular vote margin.} \]

With

\[ \beta_i = \frac{t_i}{\sum t_i} : \text{state } i \text{'s share of the national vote,} \]

it follows that the difference in the total vote within a state relative to the total vote in the country is given by

\[ V_i = \frac{m_i}{\sum t_i} = \alpha_i \beta_i. \]

Thus, every state’s contribution to the national popular vote margin is the popular vote margin in the state multiplied by the relative size of the state. In terms of the popular vote, the state margin and state size are multiplicative. And not surprisingly, abstracting from the margin (an *a posteriori* concept), each state’s share of the national population (\( \beta_i \)) is a direct measure of its potential (*a priori*) influence on the election under a national popular vote.

**State Preferences: 2016**

The electoral vote shares and the population shares for each state are shown in Table 1. The correlations for the various metrics are shown in Table 2. Population shares and electoral shares are almost perfectly correlated. The more interesting value, however, is the ratio of the two shares. That ratio, with the electoral vote share as the

---

7In order to have a consistent population measure across all elections, the decennial census data are used to measure population shares. State population shares are, of course, only approximations for the shares of eligible voters, but the correlation should be very high.

8The correlation tables show both standard (Pearson) correlations and rank (Spearman) correlations. Standard and rank correlations are calculated with analogous formulae, except that rank correlations are derived after replacing the values of the variable pairs by their rank order, 1 through \( n \), from highest to lowest (see Bhattacharyya and Johnson 1977: chaps. 12 and 15).
# TABLE 1

**Electoral Vote Shares and Population Shares**

<table>
<thead>
<tr>
<th>State</th>
<th>Electoral Votes in 2016</th>
<th>Electoral Vote Share (%)</th>
<th>2010 Census Population</th>
<th>Population Share (%)</th>
<th>Electoral Vote Share/ Popular Vote Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wyoming</td>
<td>3</td>
<td>0.56</td>
<td>568,300</td>
<td>0.18</td>
<td>3.04</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>3</td>
<td>0.56</td>
<td>601,723</td>
<td>0.19</td>
<td>2.87</td>
</tr>
<tr>
<td>Vermont</td>
<td>3</td>
<td>0.56</td>
<td>630,337</td>
<td>0.20</td>
<td>2.74</td>
</tr>
<tr>
<td>North Dakota</td>
<td>3</td>
<td>0.56</td>
<td>675,905</td>
<td>0.22</td>
<td>2.56</td>
</tr>
<tr>
<td>Alaska</td>
<td>3</td>
<td>0.56</td>
<td>721,523</td>
<td>0.23</td>
<td>2.39</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>4</td>
<td>0.74</td>
<td>1,055,247</td>
<td>0.34</td>
<td>2.18</td>
</tr>
<tr>
<td>South Dakota</td>
<td>3</td>
<td>0.56</td>
<td>819,761</td>
<td>0.26</td>
<td>2.11</td>
</tr>
<tr>
<td>Delaware</td>
<td>3</td>
<td>0.56</td>
<td>900,877</td>
<td>0.29</td>
<td>1.92</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>4</td>
<td>0.74</td>
<td>1,321,445</td>
<td>0.43</td>
<td>1.74</td>
</tr>
<tr>
<td>Montana</td>
<td>3</td>
<td>0.56</td>
<td>994,416</td>
<td>0.32</td>
<td>1.74</td>
</tr>
<tr>
<td>Maine</td>
<td>4</td>
<td>0.74</td>
<td>1,333,074</td>
<td>0.43</td>
<td>1.73</td>
</tr>
<tr>
<td>Hawaii</td>
<td>4</td>
<td>0.74</td>
<td>1,366,862</td>
<td>0.44</td>
<td>1.69</td>
</tr>
<tr>
<td>Nebraska</td>
<td>5</td>
<td>0.93</td>
<td>1,831,825</td>
<td>0.59</td>
<td>1.57</td>
</tr>
<tr>
<td>West Virginia</td>
<td>5</td>
<td>0.93</td>
<td>1,859,815</td>
<td>0.60</td>
<td>1.55</td>
</tr>
<tr>
<td>Idaho</td>
<td>4</td>
<td>0.74</td>
<td>1,573,499</td>
<td>0.51</td>
<td>1.46</td>
</tr>
<tr>
<td>New Mexico</td>
<td>5</td>
<td>0.93</td>
<td>2,067,273</td>
<td>0.67</td>
<td>1.39</td>
</tr>
<tr>
<td>Nevada</td>
<td>6</td>
<td>1.12</td>
<td>2,709,432</td>
<td>0.87</td>
<td>1.28</td>
</tr>
<tr>
<td>Utah</td>
<td>6</td>
<td>1.12</td>
<td>2,770,765</td>
<td>0.89</td>
<td>1.25</td>
</tr>
<tr>
<td>Kansas</td>
<td>6</td>
<td>1.12</td>
<td>2,863,813</td>
<td>0.92</td>
<td>1.21</td>
</tr>
<tr>
<td>Arkansas</td>
<td>6</td>
<td>1.12</td>
<td>2,926,229</td>
<td>0.94</td>
<td>1.18</td>
</tr>
<tr>
<td>Mississippi</td>
<td>6</td>
<td>1.12</td>
<td>2,978,240</td>
<td>0.96</td>
<td>1.16</td>
</tr>
<tr>
<td>Iowa</td>
<td>6</td>
<td>1.12</td>
<td>3,053,787</td>
<td>0.99</td>
<td>1.13</td>
</tr>
<tr>
<td>Connecticut</td>
<td>7</td>
<td>1.30</td>
<td>3,581,628</td>
<td>1.16</td>
<td>1.13</td>
</tr>
<tr>
<td>South Carolina</td>
<td>9</td>
<td>1.67</td>
<td>4,645,975</td>
<td>1.50</td>
<td>1.12</td>
</tr>
<tr>
<td>Minnesota</td>
<td>10</td>
<td>1.86</td>
<td>5,314,879</td>
<td>1.72</td>
<td>1.08</td>
</tr>
<tr>
<td>Alabama</td>
<td>9</td>
<td>1.67</td>
<td>4,802,982</td>
<td>1.55</td>
<td>1.08</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>7</td>
<td>1.30</td>
<td>3,764,882</td>
<td>1.22</td>
<td>1.07</td>
</tr>
<tr>
<td>Kentucky</td>
<td>8</td>
<td>1.49</td>
<td>4,350,606</td>
<td>1.40</td>
<td>1.06</td>
</tr>
<tr>
<td>Oregon</td>
<td>7</td>
<td>1.30</td>
<td>3,848,606</td>
<td>1.24</td>
<td>1.05</td>
</tr>
<tr>
<td>Colorado</td>
<td>9</td>
<td>1.67</td>
<td>5,044,930</td>
<td>1.63</td>
<td>1.03</td>
</tr>
<tr>
<td>Washington</td>
<td>12</td>
<td>2.23</td>
<td>6,753,369</td>
<td>2.18</td>
<td>1.02</td>
</tr>
<tr>
<td>Louisiana</td>
<td>8</td>
<td>1.49</td>
<td>4,553,962</td>
<td>1.47</td>
<td>1.01</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>10</td>
<td>1.86</td>
<td>5,698,230</td>
<td>1.84</td>
<td>1.01</td>
</tr>
<tr>
<td>Maryland</td>
<td>10</td>
<td>1.86</td>
<td>5,789,929</td>
<td>1.87</td>
<td>0.99</td>
</tr>
<tr>
<td>Tennessee</td>
<td>11</td>
<td>2.04</td>
<td>6,375,431</td>
<td>2.06</td>
<td>0.99</td>
</tr>
<tr>
<td>Arizona</td>
<td>11</td>
<td>2.04</td>
<td>6,412,700</td>
<td>2.07</td>
<td>0.99</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 1 (Continued)

Electoral Vote Shares and Population Shares

<table>
<thead>
<tr>
<th>State</th>
<th>Electoral Votes in 2016</th>
<th>Electoral Vote Share (%)</th>
<th>2010 Census Population</th>
<th>Population Share(%)</th>
<th>Electoral Vote Share/ Popular Vote Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indiana</td>
<td>11</td>
<td>2.04</td>
<td>6,501,582</td>
<td>2.10</td>
<td>0.97</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>11</td>
<td>2.04</td>
<td>6,559,644</td>
<td>2.12</td>
<td>0.97</td>
</tr>
<tr>
<td>Missouri</td>
<td>10</td>
<td>1.86</td>
<td>6,011,478</td>
<td>1.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Georgia</td>
<td>16</td>
<td>2.97</td>
<td>9,727,566</td>
<td>3.14</td>
<td>0.95</td>
</tr>
<tr>
<td>Virginia</td>
<td>13</td>
<td>2.42</td>
<td>8,037,736</td>
<td>2.59</td>
<td>0.93</td>
</tr>
<tr>
<td>Michigan</td>
<td>16</td>
<td>2.97</td>
<td>9,911,626</td>
<td>3.20</td>
<td>0.93</td>
</tr>
<tr>
<td>New Jersey</td>
<td>14</td>
<td>2.60</td>
<td>8,807,501</td>
<td>2.84</td>
<td>0.92</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>20</td>
<td>3.72</td>
<td>12,734,905</td>
<td>4.11</td>
<td>0.90</td>
</tr>
<tr>
<td>North Carolina</td>
<td>15</td>
<td>2.79</td>
<td>9,565,781</td>
<td>3.09</td>
<td>0.90</td>
</tr>
<tr>
<td>Ohio</td>
<td>18</td>
<td>3.35</td>
<td>11,568,495</td>
<td>3.73</td>
<td>0.90</td>
</tr>
<tr>
<td>Illinois</td>
<td>20</td>
<td>3.72</td>
<td>12,864,380</td>
<td>4.15</td>
<td>0.90</td>
</tr>
<tr>
<td>Florida</td>
<td>29</td>
<td>5.39</td>
<td>18,900,773</td>
<td>6.10</td>
<td>0.88</td>
</tr>
<tr>
<td>Texas</td>
<td>38</td>
<td>7.06</td>
<td>25,268,418</td>
<td>8.16</td>
<td>0.87</td>
</tr>
<tr>
<td>New York</td>
<td>29</td>
<td>5.39</td>
<td>19,421,055</td>
<td>6.27</td>
<td>0.86</td>
</tr>
<tr>
<td>California</td>
<td>55</td>
<td>10.22</td>
<td>37,341,989</td>
<td>12.05</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Sources: See footnote 15.

The states with an electoral preference greater than one are overwhelmingly the smaller states. Indeed, electoral preference is largely

\[
\text{electoral preference} = \frac{E_i}{\beta_i}
\]

An electoral preference greater than one means that the state’s electoral vote share is greater than its popular vote share, and vice versa. This ratio offers a simple and direct indicator of a state’s a priori preference for the electoral college. In 2016, 32 states and the District of Columbia had an electoral preference greater than one. The median electoral preference was 1.08. The distribution of electoral preferences across states is shown in Table 1 and Figure 1.

The states with an electoral preference greater than one are overwhelmingly the smaller states. Indeed, electoral preference is largely

638
TABLE 2
CORRELATIONS FOR VOTE SHARE AND PREFERENCE, 2016

<table>
<thead>
<tr>
<th></th>
<th>Electoral Vote Share</th>
<th>Popular Vote Share</th>
<th>Electoral Vote Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral vote share</td>
<td>1</td>
<td>0.9996</td>
<td>-0.5256</td>
</tr>
<tr>
<td>Popular vote share</td>
<td></td>
<td>1</td>
<td>-0.5340</td>
</tr>
<tr>
<td>Electoral vote preference</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Electoral Vote Share</th>
<th>Popular Vote Share</th>
<th>Electoral Vote Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral vote share</td>
<td>1</td>
<td>0.9960</td>
<td>-0.9795</td>
</tr>
<tr>
<td>Popular vote share</td>
<td></td>
<td>1</td>
<td>-0.9881</td>
</tr>
<tr>
<td>Electoral vote preference</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Number of observations = 51.
SOURCE: Data in Table 1.

a function of state population. The rank correlation between population shares and electoral preference is $-0.988$; the two measures produce almost identical (inverse) rankings. That relationship between electoral preference and population share, shown in Figure 2, results in the characteristic hyperbolic form.

The inverse relationship between population and electoral preference derives from the method used to determine the number of congressional representatives per state, which determines the number of electors per state. As already described, each state has two senators and at least one representative, so each state has at least three electoral votes. The remaining 385 representatives—and their electoral votes—are allocated among the remaining states based on population, as described in the next section.

Apportioning the House of Representatives

In accordance with the Constitution, the House of Representatives is reapportioned every 10 years (see U.S. Constitution,
Sources: See Table 1 and footnote 15.

Article I, Sections 2 and 3). The reapportionment occurs with the second full congressional term following the publication of the decennial census. For example, the composition of the 112th Congress (2011–12 term) was the last apportioned according to the 2000 census. The 113th Congress (2013–14 term) was the first apportioned according to the 2010 census. As noticed earlier, the total number of representatives has been capped at 435 since 1911. The 385 additional representatives (beyond the minimum of one per state) are apportioned by state population, according to the schedule just described.

Several different apportioning methods have been used since the first census in 1790. The current method has been in place since 1941 and is known as the “equal proportions method.” The method appears to have been suggested by the mathematician Edward Huntington and by Joseph Hill, Chief Statistician of the Bureau of...
Census. Sometimes referred to as the Huntington-Hill method, it is applied as follows.\(^9\)

Each additional representative (house seat) is given a weight:

\[
\frac{1}{\sqrt{n(n-1)}}
\]

Where \(n\) represents the number of seats a state would have if it gained a seat. Because all states automatically receive one seat, the first seat gained is seat two. The weight for seat two is

\[
\frac{1}{\sqrt{2(2-1)}}
\]

\(^9\)See Huntington (1921, 1928) and Saari (1978). Helpful discussions can also be found at: www.census.gov/history/www/reference/apportionment/methods_of_apportionment.html; www.census.gov/population/apportionment/about/computing.html; and www.history.house.gov/home.
the weight for seat three is

\[ \frac{1}{\sqrt{3(3 - 1)}} \]

and so on. The weights are shown in Figure 3. The weighting formula reflects, indeed it can be said to generate, the relationship between electoral preference and population shown in Figure 2.

The weights are multiplied by the state populations \( (P_i) \) to create a series of values \( (H_{n,i}) \) for seat \( n \), for each state \( i \):

\[ H_{n,i} = \left( \frac{1}{\sqrt{n(n - 1)}} \right) P_i. \]

A sufficient number of the \( H_{n,i} \) values, say 99, are calculated for each state: Alabama \( H(2) \) through Alabama \( H(100) \); Alaska \( H(2) \) through Alaska \( H(100) \); and so on. (The number of values created for

FIGURE 3

POPULATION WEIGHTS FOR HOUSE APPORTIONMENT

![Graph showing population weights for house apportionment.](image)

SOURCE: Current apportionment formula described in text.
each state needs only to exceed by one the number of seats eventually granted to the largest state.) The resulting values are then sorted from highest to lowest. The first 385 values of that sorted sequence determine the additional seats received by each state.

The above formula for allocating those 385 additional representatives, along with the three-vote minimum for each state, generates the result described here: small states will have high electoral shares relative to their population shares, while the inverse holds for large states.

The Conventional View

While the issue is not always clearly presented, the general consensus in the academic literature appears to be that the electoral college system reflects and maintains the inherent advantage of large states. That analysis begins with the insightful distinction between voting share and voting power in systems with unequal voting shares, such as the electoral college.

Voting power derives from the likelihood that a voter will be influential, in the sense that changing just that one vote would change the outcome of the election. That likelihood is determined not just by a voter’s share, but by the distribution of all voting shares. While voting power will generally be highly correlated with voting share, the two concepts are not necessarily equal and can, under certain circumstances, be quite different. As a simple example, if a single entity controls more than half of the voting shares in a simple majority election, then only that entity’s vote matters. Its voting power is

---

10See the articles cited in footnote 1. As shown here, the small state advantage in the electoral college derives from the relationship between popular vote shares and electoral vote shares. While that relationship is sometimes described in the literature, its relevance is overlooked. Strömberg (2008) and Gelman, Silver, and Edlin (2012) might be considered exceptions, but their conclusions derive from a posteriori analysis of particular elections. I have been unable to identify any academic research that defines and quantifies the state-by-state preference for the electoral college and the resulting small state advantage, as I do here. The literature on this subject is vast, however, so that contention may be incorrect.

11Regarding such systems, scholars long ago began to distinguish between voter share and voter power. The seminal articles are Penrose (1946), Shapley and Shubik (1954), Mann and Shapley (1964), and Banzhaf (1965, 1968). For helpful surveys, see Leech (2002a, 2002b), Felsenthal and Machover (2004, 2012), and Strömberg (2008).
measured as one, and any other entity’s voting power is measured as zero, regardless of the specific shares.

Two commonly used measures of voting power are the Banzhaf index (sometimes referred to as the Penrose-Banzhaf index) and the Shapley-Shubik index. The Banzhaf index for a voter is calculated by creating all possible winning coalitions, of any size, and then calculating the percentage of those winning coalitions in which the voter at issue casts an influential vote—a vote that if changed would change the outcome of the election. Because any particular coalition can have multiple influential voters, the Banzhaf index must be normalized to sum to one.

The Shapley-Shubik index for a voter is calculated by creating all possible sequences of all of the voters and then calculating the percentage of those sequences in which the voter at issue casts the deciding vote, assuming that all prior voters in the sequence voted in the same way (that is, as a coalition). Every sequence will have only one deciding voter, the one who “tips the scales” when the votes are tabulated sequentially.12

We now compare these two indices with the population share calculations from Table 1. Table 3 displays the correlations among (1) the electoral vote shares and the two electoral vote power indices, (2) the population shares and the two population power indices, and (3) the electoral preference ratios derived from the shares and the two power indices.13

When it comes to the electoral college and state populations, shares and power indices are highly correlated (the lowest standard correlation is 0.977) and tend to produce similar rankings (the lowest rank correlation is 0.958). Although some mid-size states switch their electoral preference from one side of 1.0 to the other, based on share versus power index, the electoral preference values are also highly correlated. The lowest standard correlation is 0.954; the lowest rank correlation is 0.860.

In addition, the median electoral preference is 1.08 based on shares, 1.17 based on the Banzhaf index, and 1.09 based on the

---

12 An accessible discussion of the two indices is found at www.math.colostate.edu/~spriggs/m130/power2.pdf. The indices herein are calculated using computational methods provided by the University of Warwick at http://homepages.warwick.ac.uk/~ecaae/index.html.

13 The data underlying the correlations are available from the author.
### TABLE 3

**Correlations for 2016**

<table>
<thead>
<tr>
<th></th>
<th>Electoral Vote (EV) Share</th>
<th>Shapley-Shubik EV Index</th>
<th>Banzhaf EV Index</th>
<th>Population Shapley-Shubik Share</th>
<th>Shapley-Shubik POP Index</th>
<th>Banzhaf POP Index</th>
<th>Electoral Vote Preference (EVP)</th>
<th>Shapley-Shubik EVP</th>
<th>Banzhaf EVP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral Vote (EV) Share</td>
<td>1.000</td>
<td>0.999</td>
<td>0.998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shapley-Shubik EV Index</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banzhaf EV Index</td>
<td></td>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population (POP) Share</td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.999</td>
<td></td>
<td>0.977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shapley-Shubik POP Index</td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.979</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banzhaf POP Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electoral Vote Preference (EVP)</td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.998</td>
<td></td>
<td>0.954</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shapley-Shubik EVP</td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td></td>
<td>0.956</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>Banzhaf EVP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*(Continued)*
TABLE 3 (Continued)
CORRELATIONS FOR 2016

<table>
<thead>
<tr>
<th>Rank (Spearman) Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral Vote (EV) Share</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>Electoral Vote (EV) Share</td>
</tr>
<tr>
<td>Shapley-Shubik EV Index</td>
</tr>
<tr>
<td>Banzhaf EV Index</td>
</tr>
<tr>
<td>Population Share</td>
</tr>
<tr>
<td>Shapley-Shubik POP Index</td>
</tr>
<tr>
<td>Banzhaf POP Index</td>
</tr>
<tr>
<td>Electoral Vote Preference (EVP)</td>
</tr>
<tr>
<td>Shapley-Shubik EVP</td>
</tr>
<tr>
<td>Banzhaf EVP</td>
</tr>
</tbody>
</table>

Sources: Data available from author.
Electoral College

The existing literature adds a second step to the analysis such that the electoral power index is adjusted by a corresponding voter power index. That is, the likelihood that an individual voter in a given state will cast an influential vote (the likelihood that the popular vote is tied) is factored into the analysis. The probability of a tie vote, however trivial, is inversely related to population—smaller in larger states and vice versa. The result is that the substantial advantage for large states inherent in the electoral college is reduced to some extent, but not enough to overcome that advantage. Miller (2012) summarizes that analysis:

An individual’s voting power is the probability that his vote is decisive within the state times the probability that his state’s bloc of electoral votes is decisive in the Electoral College. . . . This effect, first noted with explicit reference to the Electoral College by Banzhaf (1968) . . . and it implies that voters in the most favored state (California) have almost three and half times the voting power as voters in the least favored state (Montana, the largest with only one House seat).

Criticisms of this two-stage analysis can be found in Margolis (1983), Gelman et al. (2002, 2004, 2012), and Katz, Gelman, and King (2004). The two-stage analysis is not relevant here (or at all, some might argue) because our interest is in the preference of the state as represented by the majority of the voters in the state.

14The existing literature adds a second step to the analysis such that the electoral power index is adjusted by a corresponding voter power index. That is, the likelihood that an individual voter in a given state will cast an influential vote (the likelihood that the popular vote is tied) is factored into the analysis. The probability of a tie vote, however trivial, is inversely related to population—smaller in larger states and vice versa. The result is that the substantial advantage for large states inherent in the electoral college is reduced to some extent, but not enough to overcome that advantage. Miller (2012) summarizes that analysis:

An individual’s voting power is the probability that his vote is decisive within the state times the probability that his state’s bloc of electoral votes is decisive in the Electoral College. . . . This effect, first noted with explicit reference to the Electoral College by Banzhaf (1968) . . . and it implies that voters in the most favored state (California) have almost three and half times the voting power as voters in the least favored state (Montana, the largest with only one House seat).

Criticisms of this two-stage analysis can be found in Margolis (1983), Gelman et al. (2002, 2004, 2012), and Katz, Gelman, and King (2004). The two-stage analysis is not relevant here (or at all, some might argue) because our interest is in the preference of the state as represented by the majority of the voters in the state.
electoral college influence relative to the influence the state would have in a national popular vote. The analysis here measures that relative influence and thereby identifies which system would be preferred *a priori* by the voters in any given state.

**State Preferences: 1788–2012**

The current distribution of populations and electoral votes implies that the majority of states—in particular, the smaller ones—prefers the electoral college system. Has that pattern always prevailed? Computing the same values (as shown in Table 1 for the 2016 election) over all of the presidential elections is straightforward. To arrive at populations for an election year, I interpolate between decennial census figures. The 1790 census is used to measure population shares for 1788. The 2010 census is used to measure population shares for 2012.\(^{15}\)

Measuring the electoral college shares from reported tabulations and the population shares from the census figures generates the results shown in Table 4. In only two of the 58 presidential elections (1876 and 1880) has the median electoral preference been lower than one. In all of the other elections the majority of the states have, by this measure, preferred the electoral college system. In addition, the strong negative relationship between electoral preference and population has always prevailed; both the simple and the rank correlations between electoral preference and population have always been significantly negative. The empirical results indicate that the current method of apportionment (in effect since 1941) did not fundamentally alter the relationship between state population and electoral preference; that relationship has always existed.

It should also be noted that there is nothing inevitable about the majority of states preferring the electoral college. It is possible to describe a distribution of populations such that the majority of states

\(^{15}\)The decennial census data can be found at: www.census.gov/population/www/censusdata/pop1790-1990.html; www.census.gov/population/www/cen2000/maps/respop.html; and www.census.gov/population/apportionment/data/2010_apportionment_results.html. For 1790 through 1860, the population figures reflect the decennial census “free” population as calculated by Gibson and Jung (2002). In any given election year, a state’s population is used (to calculate total population and state shares) only if there is an electoral vote tabulation for that state in the National Archives and Records Administration database.
TABLE 4
MEDIAN VALUES AND CORRELATION COEFFICIENTS 1788–2012

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1788</td>
<td>10</td>
<td>1.064</td>
<td>-0.761</td>
<td>-0.964</td>
<td>1904</td>
<td>45</td>
<td>1.075</td>
<td>-0.375</td>
<td>-0.969</td>
</tr>
<tr>
<td>1792</td>
<td>15</td>
<td>1.000</td>
<td>-0.560</td>
<td>-0.543</td>
<td>1908</td>
<td>46</td>
<td>1.070</td>
<td>-0.382</td>
<td>-0.880</td>
</tr>
<tr>
<td>1796</td>
<td>16</td>
<td>1.048</td>
<td>-0.561</td>
<td>-0.562</td>
<td>1912</td>
<td>48</td>
<td>1.037</td>
<td>-0.431</td>
<td>-0.978</td>
</tr>
<tr>
<td>1800</td>
<td>16</td>
<td>1.083</td>
<td>-0.467</td>
<td>-0.326</td>
<td>1916</td>
<td>48</td>
<td>1.047</td>
<td>-0.411</td>
<td>-0.959</td>
</tr>
<tr>
<td>1804</td>
<td>17</td>
<td>1.028</td>
<td>-0.426</td>
<td>-0.316</td>
<td>1920</td>
<td>48</td>
<td>1.069</td>
<td>-0.393</td>
<td>-0.938</td>
</tr>
<tr>
<td>1808</td>
<td>17</td>
<td>1.084</td>
<td>-0.379</td>
<td>-0.328</td>
<td>1924</td>
<td>48</td>
<td>1.093</td>
<td>-0.403</td>
<td>-0.915</td>
</tr>
<tr>
<td>1812</td>
<td>18</td>
<td>1.034</td>
<td>-0.565</td>
<td>-0.616</td>
<td>1928</td>
<td>48</td>
<td>1.121</td>
<td>-0.415</td>
<td>-0.885</td>
</tr>
<tr>
<td>1816</td>
<td>19</td>
<td>1.082</td>
<td>-0.496</td>
<td>-0.598</td>
<td>1932</td>
<td>48</td>
<td>1.069</td>
<td>-0.385</td>
<td>-0.980</td>
</tr>
<tr>
<td>1820</td>
<td>24</td>
<td>1.175</td>
<td>-0.580</td>
<td>-0.712</td>
<td>1936</td>
<td>48</td>
<td>1.057</td>
<td>-0.395</td>
<td>-0.969</td>
</tr>
<tr>
<td>1824</td>
<td>24</td>
<td>1.102</td>
<td>-0.574</td>
<td>-0.710</td>
<td>1940</td>
<td>48</td>
<td>1.056</td>
<td>-0.405</td>
<td>-0.945</td>
</tr>
<tr>
<td>1828</td>
<td>24</td>
<td>1.108</td>
<td>-0.519</td>
<td>-0.628</td>
<td>1944</td>
<td>48</td>
<td>1.049</td>
<td>-0.448</td>
<td>-0.969</td>
</tr>
<tr>
<td>1832</td>
<td>24</td>
<td>1.046</td>
<td>-0.539</td>
<td>-0.649</td>
<td>1948</td>
<td>48</td>
<td>1.059</td>
<td>-0.473</td>
<td>-0.947</td>
</tr>
<tr>
<td>1836</td>
<td>26</td>
<td>1.109</td>
<td>-0.464</td>
<td>-0.521</td>
<td>1952</td>
<td>48</td>
<td>1.070</td>
<td>-0.506</td>
<td>-0.984</td>
</tr>
<tr>
<td>1840</td>
<td>26</td>
<td>1.159</td>
<td>-0.434</td>
<td>-0.493</td>
<td>1956</td>
<td>48</td>
<td>1.090</td>
<td>-0.546</td>
<td>-0.970</td>
</tr>
</tbody>
</table>

(Continued)
### TABLE 4 (Continued)
### MEDIAN VALUES AND CORRELATION COEFFICIENTS 1788–2012

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1844</td>
<td>26</td>
<td>1.113</td>
<td>-0.603</td>
<td>-0.802</td>
<td>1960</td>
<td>50</td>
<td>1.118</td>
<td>-0.534</td>
<td>-0.946</td>
</tr>
<tr>
<td>1848</td>
<td>30</td>
<td>1.166</td>
<td>-0.457</td>
<td>-0.826</td>
<td>1964</td>
<td>51</td>
<td>1.110</td>
<td>-0.530</td>
<td>-0.985</td>
</tr>
<tr>
<td>1852</td>
<td>31</td>
<td>1.108</td>
<td>-0.479</td>
<td>-0.550</td>
<td>1968</td>
<td>51</td>
<td>1.092</td>
<td>-0.543</td>
<td>-0.976</td>
</tr>
<tr>
<td>1856</td>
<td>31</td>
<td>1.167</td>
<td>-0.464</td>
<td>-0.713</td>
<td>1972</td>
<td>51</td>
<td>1.088</td>
<td>-0.545</td>
<td>-0.986</td>
</tr>
<tr>
<td>1860</td>
<td>33</td>
<td>1.166</td>
<td>-0.450</td>
<td>-0.774</td>
<td>1976</td>
<td>51</td>
<td>1.050</td>
<td>-0.551</td>
<td>-0.968</td>
</tr>
<tr>
<td>1864</td>
<td>25</td>
<td>1.101</td>
<td>-0.384</td>
<td>-0.964</td>
<td>1980</td>
<td>51</td>
<td>1.035</td>
<td>-0.552</td>
<td>-0.895</td>
</tr>
<tr>
<td>1868</td>
<td>34</td>
<td>1.074</td>
<td>-0.388</td>
<td>-0.920</td>
<td>1984</td>
<td>51</td>
<td>1.101</td>
<td>-0.596</td>
<td>-0.985</td>
</tr>
<tr>
<td>1872</td>
<td>35</td>
<td>1.016</td>
<td>-0.416</td>
<td>-0.950</td>
<td>1988</td>
<td>51</td>
<td>1.117</td>
<td>-0.593</td>
<td>-0.966</td>
</tr>
<tr>
<td>1876</td>
<td>38</td>
<td>0.993</td>
<td>-0.423</td>
<td>-0.843</td>
<td>1992</td>
<td>51</td>
<td>1.086</td>
<td>-0.561</td>
<td>-0.980</td>
</tr>
<tr>
<td>1880</td>
<td>38</td>
<td>0.988</td>
<td>-0.419</td>
<td>-0.680</td>
<td>1996</td>
<td>51</td>
<td>1.061</td>
<td>-0.550</td>
<td>-0.969</td>
</tr>
<tr>
<td>1884</td>
<td>38</td>
<td>1.009</td>
<td>-0.370</td>
<td>-0.892</td>
<td>2000</td>
<td>51</td>
<td>1.063</td>
<td>-0.540</td>
<td>-0.949</td>
</tr>
<tr>
<td>1888</td>
<td>38</td>
<td>1.020</td>
<td>-0.335</td>
<td>-0.783</td>
<td>2004</td>
<td>51</td>
<td>1.067</td>
<td>-0.537</td>
<td>-0.987</td>
</tr>
<tr>
<td>1892</td>
<td>44</td>
<td>1.020</td>
<td>-0.408</td>
<td>-0.961</td>
<td>2008</td>
<td>51</td>
<td>1.069</td>
<td>-0.534</td>
<td>-0.967</td>
</tr>
<tr>
<td>1896</td>
<td>45</td>
<td>1.053</td>
<td>-0.380</td>
<td>-0.952</td>
<td>2012</td>
<td>51</td>
<td>1.079</td>
<td>-0.535</td>
<td>-0.988</td>
</tr>
<tr>
<td>1900</td>
<td>45</td>
<td>1.044</td>
<td>-0.353</td>
<td>-0.936</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sources:** See footnotes 5, 12, and 15.
would prefer a popular vote. Such an outcome would occur, for example, if the distribution of state populations were bimodal, with a group of similar small population states and a group of similar large population states, with the number of small states less than the number of large states. For example, consider if there were 20 states with small populations (20 states like, say, Wyoming) such that each was accorded the minimum number of electors, with the remaining population and electors divided among the 30 other states (30 states like, say, Florida). That distribution of populations and the implied distribution of electoral votes could generate a majority of states in favor of a popular vote system.

Conclusion

There is a powerful reason for the persistence of the electoral college: a majority of states—overwhelmingly the smaller ones—prefer it, which has been the case throughout U.S. history. Furthermore, abandoning the electoral college can only happen through constitutional amendment, which requires the assent of not just a majority but three-fourths of the states (see U.S. Constitution, Article V). Given the structure of the electoral college and the distribution of state populations, such an outcome is implausible in the foreseeable future.

The distribution of congressional representatives (and consequently of electors) was not happenstance, but rather the result of a purposeful compromise among the states, in consideration of the wide variation in state populations. The Constitution embodies several such accommodations, so as to prevent a small number of large states from dominating political outcomes. As one of those accommodations, the electoral college is likely to remain durable.

References


THE APOTHEOSIS OF THE RENTIER: HOW NAPOLEONIC WAR FINANCE KICK-STARTED THE INDUSTRIAL REVOLUTION

Martin Hutchinson and Kevin Dowd

[T]his state of affairs . . . would mean the euthanasia of the rentier, and, consequently, the euthanasia of the cumulative oppressive power of the capitalist to exploit the scarcity-value of capital . . . I see, therefore, the rentier aspect of capitalism as a transitional phase which will disappear when it has done its work.


The development of Britain’s economy in the years following the end of the Napoleonic Wars was heavily influenced by the peculiar nature of British government wartime finance. Instead of issuing bonds with higher coupons as interest rates rose, which governments normally did in wartime, British governments from the 1750s onward relied mostly on 3 percent “Consols”, i.e., perpetual bonds with a 3 percent coupon issued at deep discounts. In a world in which equity markets were almost nonexistent and in which there was a gigantic government bond market swollen by war financing, fluctuations in the prices of Consols caused swings in investor wealth that had major...
economic effects, which have been much underappreciated. After a positive wealth effect in the postwar 1780s and a negative one in the late 1790s, the large, long-lasting bond price increase after 1813 played a major role in capitalizing the Industrial Revolution’s final take-off and the acceleration of economic growth to near-modern levels. As Jeffrey Williamson (1984: 688) noted when discussing the slow start to the Industrial Revolution, “Somewhere around the 1820s Britain passed through a secular turning point.” In this article, we argue that the key to that turning point was the massive wealth effect of the postwar increase in Consols’ prices.

The Napoleonic Wars’ Financing Mechanism

Normally, when a government wants to borrow to finance a war, it knows that its borrowing will cause interest rates to rise, especially if there is inflation. For example, if peacetime interest rates are about 3 percent, the government might issue a 5 percent bond of 20-year maturity rather than say the usual 3 percent peacetime bond. When the bond matures 20 years later and peace reigns again, the government can refinance the bond at 3 percent. Investors holding the 5 percent bond receive a higher interest rate, but no capital gain if they hold the bond to maturity and only a modest gain if they sell mid-term at a premium while bond prices are higher than normal.

That approach broadly describes how the U.S. wars in Korea and Vietnam were financed, as well as both U.S. and British contributions to the two world wars. Earlier, the U.K. government had used a similar approach to finance the War of the Spanish Succession (1702–13), mostly at interest rates of 8–10 percent and with heavy use of tontines and lotteries.

The U.K. government adopted a different approach with its next major wars. In 1751, Sampson Gideon, a brilliant Jewish-British financier, persuaded Prime Minister Henry Pelham to convert most of the outstanding British government debt into 3 percent Consolidated Annuities (the famous Consols). Until their ill-advised redemption in 2015, these were perpetual—that is, they had an infinite maturity—and paid 3 percent interest each year. When the

1 The securities issued in the 1751 refinancing paid 3.5 percent interest for the first six years, then 3 percent from 1757. There was a further refinancing in 1888, carried out by Chancellor of the Exchequer George Joachim Goschen, into new securities paying 2.75 percent from 1888 to 1903, then 2.5 percent thereafter. Thus, the classic “3 percent Consols” existed without refinancing from 1757 to 1888.
government wanted to finance a war, instead of issuing 5 percent bonds at par, it issued 3 percent Consols at perhaps 60 percent of the “par” principal amount, which would then yield 5 percent (i.e., 3 percent/60 percent) on a running yield basis.

Gideon’s thinking was as follows: since all wars were temporary, people who bought 3 percent bonds at 60 percent would make generous 5 percent yields during the war and would then make a large capital gain afterwards, when yields dropped to 3 percent and the bonds went back to par. The Consols were therefore an attractive investment. The government could then find buyers even in war years, provided (as was the case) that there was confidence that Britain would later repay its debts in full.

Not only did investors enjoy strong yields, but they also made big capital gains when peace came: 21 percentage points in six years after the bottom in 1762 as the Seven Years Wars (1756–63) drew to an end, and 33 percentage points in 10 years after the wartime bottom in 1782, as the American War of Independence (1775–82) drew to an end.²

There was then strong demand for new issues of Consols throughout the major wars over the course of the following 50-plus years—the Seven Years War, the American Wars, and the French Revolutionary and Napoleonic Wars (1793–1815, with short remissions).

The Consols enabled Britain to finance heavy military expenditures when its rival France, which had no such mechanism, was unable to do so without financially crippling itself. Gideon’s clever structure, which gave windfall profits at the end of each war to inventory-holding bond dealers like himself, was central to Britain’s acquiring its empire and to not losing it again after the American colonies broke away.

The disadvantage of Gideon’s structure from the fiscal point of view was that, for each £100 of war expenditure on 3 percent Consols issued at 60 percent of par, £167 was typically added to the debt. But since the debt was perpetual and never needed to be redeemed, this feature of Consols finance was not regarded as a major fiscal disadvantage. Lord North, chancellor of the exchequer from 1767 to 1782 and prime minister through the American War of Independence,

²Homer and Sylla (2005: 57 for 1752–99; 192 for 1800–27). We take the Homer–Sylla estimate of annual average prices as the basis for these calculations.
liked the structure because “it was the interest that the people were burdened with the paying of and not the capital.”

The interest cost was indeed the same, or even slightly lower than through issuing new 5–6 percent debt at par, because the Consols’ attractiveness to speculators allowed them to be sold at a slightly lower running yield. However, over the course of the century after peace returned in 1815, Britain’s outstanding debt would decline only to £650 million in 1914, although the growth in the British economy meant that debt represented only 30 percent of GDP compared to a peak of around 260 percent of GDP in 1819.

The rise in the prices of British Consols from 1813 onward appears to have been due, in part, to a reduction in their perceived risk as the Allies approached France and peace returned, and, in part, to a drastic reduction in the annual supply of new Consols through budget deficit financing.

Table 1 shows gross public income and expenditure for the years 1812–27, with the surplus or deficit; it also shows the funded and unfunded debt outstanding at the start of each year, and the increase in debt during the year.

The deficit figures and debt increase figures in Table 1 do not tally because of the timing of government payments and debt financings. For example, the 1814 military campaign was largely financed by a £22 million debt issue (increasing the amount of 3 percent debt outstanding by £38.9 million) on the morning of the supplementary Budget of November 15, 1813. Nevertheless, the overall trend

---

4 A more elaborate explanation might go as follows. The prices of Consols surged primarily because of a decline in supply as the Napoleonic Wars ended and the government moved toward fiscal surplus. Moreover, because Consols were almost the only store of liquid wealth, their price rise would have made investors wealthier—thereby creating an even greater demand for Consols. It is noteworthy that the two main investments available to the rich were Consols and land, but land became much less profitable after 1815 as grain prices declined. Stocks and bonds were relatively unimportant (though a small foreign bond market arose in the 1820s), while direct investment in companies was risky, illiquid, and poorly diffused among the investing public.

In loanable funds terms, the primary driving factor in the market for Consols—the decline in supply—would correspond to a reduced demand by the government for loanable funds.
5 The debt figure jumps in 1818 because of the inclusion of Irish government debt from that year onward.
Apotheosis of the Rentier

### TABLE 1

<table>
<thead>
<tr>
<th>Year</th>
<th>Gross Public Income</th>
<th>Expend.</th>
<th>Surplus/Deficit</th>
<th>Start of Year Debt Funded</th>
<th>Unfunded</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1812</td>
<td>70.3</td>
<td>94.8</td>
<td>-24.5</td>
<td>583.4</td>
<td>42.6</td>
<td>24.1</td>
</tr>
<tr>
<td>1813</td>
<td>74.7</td>
<td>111.1</td>
<td>-36.4</td>
<td>607.5</td>
<td>44.8</td>
<td>70.0</td>
</tr>
<tr>
<td>1814</td>
<td>77.9</td>
<td>112.9</td>
<td>-35.0</td>
<td>677.5</td>
<td>48.1</td>
<td>7.1</td>
</tr>
<tr>
<td>1815</td>
<td>79.1</td>
<td>99.5</td>
<td>-20.4</td>
<td>684.6</td>
<td>60.3</td>
<td>49.0</td>
</tr>
<tr>
<td>1816</td>
<td>69.2</td>
<td>71.3</td>
<td>-2.1</td>
<td>733.6</td>
<td>44.7</td>
<td>-17.3</td>
</tr>
<tr>
<td>1817</td>
<td>57.6</td>
<td>58.7</td>
<td>-1.1</td>
<td>716.3</td>
<td>49.8</td>
<td>—</td>
</tr>
<tr>
<td>1818</td>
<td>59.5</td>
<td>57.6</td>
<td>1.9</td>
<td>780.6</td>
<td>62.6</td>
<td>15.0</td>
</tr>
<tr>
<td>1819</td>
<td>58.1</td>
<td>57.5</td>
<td>0.6</td>
<td>795.6</td>
<td>48.7</td>
<td>2.9</td>
</tr>
<tr>
<td>1820</td>
<td>59.9</td>
<td>58.4</td>
<td>1.5</td>
<td>798.5</td>
<td>41.6</td>
<td>6.4</td>
</tr>
<tr>
<td>1821</td>
<td>61.6</td>
<td>58.4</td>
<td>3.2</td>
<td>804.9</td>
<td>33.3</td>
<td>-6.4</td>
</tr>
<tr>
<td>1822</td>
<td>59.9</td>
<td>56.5</td>
<td>3.4</td>
<td>798.5</td>
<td>32.7</td>
<td>-1.1</td>
</tr>
<tr>
<td>1823</td>
<td>58.5</td>
<td>54.3</td>
<td>4.2</td>
<td>797.4</td>
<td>38.7</td>
<td>-4.5</td>
</tr>
<tr>
<td>1824</td>
<td>59.7</td>
<td>55.5</td>
<td>4.2</td>
<td>792.9</td>
<td>35.8</td>
<td>-10.6</td>
</tr>
<tr>
<td>1825</td>
<td>57.7</td>
<td>54.1</td>
<td>3.6</td>
<td>782.3</td>
<td>37.9</td>
<td>-3.0</td>
</tr>
<tr>
<td>1826</td>
<td>55.2</td>
<td>56.1</td>
<td>-0.9</td>
<td>779.3</td>
<td>31.7</td>
<td>5.7</td>
</tr>
<tr>
<td>1827</td>
<td>54.7</td>
<td>55.9</td>
<td>-1.2</td>
<td>785.0</td>
<td>25.0</td>
<td>—</td>
</tr>
</tbody>
</table>

**Notes:** All units are in £m. Figures from 1818 onward include Irish debt. Fiscal year-end was January 5 during this period, so figures are for January 6 of the year concerned to January 5 of the following year.

**Source:** Mitchell (2011), Tables Public Finance 3 (Income: 581), Public Finance 4 (Expenditure: 587), and Public Finance 7 (Debt: 601).

It is clear: huge deficits and increases in the supply of Consols during the war years of 1812–15 were followed by near-balanced budgets in 1816–19 and surpluses thereafter, while the supply of Consols stopped increasing from 1816 onward, except for a modest blip in 1819–20 caused by funding £10 million of the Bank of England’s holdings of Exchequer Bills in connection with the return to gold.\(^6\)

\(^6\)There is, thus, the following implied counterfactual. Reduced supply of Consols, itself associated with a move toward fiscal surplus, combined with increased demand for them and the anticipated then actual resumption of specie payments at the old parity to create a secular decline in interest rates. Falling yields also suggest that, contrary to Ashton (1948) or Williamson (1984), fiscal crowding out was no longer an issue by this period.
Table 2 sets out the £823 million of British and Irish government funded and unfunded debt that was held by the public on February 1, 1817, a date chosen so the special “emergency” financings for the Waterloo campaign were out of the way.

By far the greatest part of the outstanding debt, £562 million or 68 percent of the total, consisted of perpetual securities bearing a 3 percent interest rate, payable twice yearly. The largest single tranche, £384 million, consisted of “Consolidated Annuities,” the 3 percent Consols, which paid interest in January and July. There was also an exactly equivalent obligation, the Reduced Annuities, with principal amount of £148 million, which differed from Consols only in paying interest in April and October. Dealers would arbitrage between these two securities, with the Reduced Annuities generally trading at a small discount (subject to fluctuations before and after their interest payment dates) because of their somewhat lesser liquidity. There were also two relatively small older issues with
3 percent coupons, reflecting refinancings of debt incurred early in the 18th century, for the Bank of England and the South Sea Company, of £16 million and £14 million, respectively.

In addition to the perpetual 3 percent debt, there was £211 million of 4 percent and 5 percent debt, which would be refinanced after the war as interest rates declined. The 4 percent debt, totaling £75 million, took the form of 4 percent Consols, while the 5 percent debt, totaling £136 million, originally contracted in many cases by the Navy and Army directly, had been consolidated into 5 percent Consols. By the 1820s, the 5 percent debt was trading above par. There were then two refinancing operations carried out at the end of 1823 and in early 1824, one converting £135 million of 5 percent Consols into 4 percent Consols and the second converting £80 million of 4 percent Consols into a new issue of 3.5 percent Consols.\(^7\)

Finally, in 1817 there was £50 million of short-term debt, mostly in the form of Exchequer Bills, which were short-term instruments, generally converted into long-term debt as new issues were undertaken. The £50 million outstanding in February 1817 was high by historical standards, a residue of the emergency financings undertaken at the Napoleonic Wars’ climax in 1813–15, and was over the next few years reduced by being converted into long-term debt. Reducing the amount of Exchequer Bills outstanding, thereby mopping up excess market liquidity, was an important precondition for the subsequent return to the gold standard.

In addition to the £822 million of debt held by the public, the government had since 1786 built up a sinking fund, intended to accrue at compound interest and allow the debt to be redeemed within 45 years. This fund had been established by William Pitt the Younger because redeeming debt at par, under Gideon’s structure, was expensive and the sinking fund allowed other debt to be redeemed through market purchases. Payments had been made into the sinking fund throughout the war years, even though the government was running deficits, on the principle that the gradual accrual of the sinking fund, with each new issue of debt having an amortization provision to be paid into the sinking fund, would

\(^7\) This, and much other useful information about British government financing in 1815–25, is set out in Neal (1998).
reassure the nation’s creditors. By February 1817 the sinking fund totaled £62 million. It would be modified in 1819 and eliminated in stages in the early 1820s, with the government debt it purchased being canceled in the process.

As an example of how war finance was structured, we can examine the last major financing undertaken during the Napoleonic Wars, which was also the largest single tranche undertaken, and which was announced by Chancellor of the Exchequer Nicholas Vansittart to the House of Commons on June 14, 1815, four days before the Battle of Waterloo, and had been completed that morning. The loan raised £27 million in net proceeds, for each £100 of which buyers would be given £130 in 3 percent reduced stock, £10 in 4 percent Consols, and £44 in 3 percent Consols. The £27 million of net proceeds was obtained by issuing a total of £49.68 million principal amount of new debt, at a running interest cost of 5.62 percent and an average issue price of 54.35 percent of par. In addition to the interest payable, a sinking fund provision of 2.81 percent was made on the £27 million raised, for a total annual debt service charge of 8.43 percent.\(^8\)

Vansittart told the House of Commons that this record-sized issue, the pricing of which was determined by negotiated tender and not by competitive bid, was only just fully subscribed; a great proportion of the issue was initially left with the underwriters. Four days later the Battle of Waterloo took place, news of which was received on the late evening of June 21.

Contrary to widespread belief, there was no great immediate “bounce” in the market by which (in separate legends) Nathan Rothschild and the broker/economist David Ricardo were both supposed to have made £1 million each through trading on early (and, effectively, inside) information of the battle. However, given the price rise in Consols over the next decade, their holdings of this issue alone played a major role in generating the Rothschild and Ricardo fortunes.

As anticipated by Gideon 60 years earlier, City of London financiers did well from the postwar surges in bond prices. It is no coincidence that Rothschilds, led by Nathan Meyer Rothschild, and Barings, led by the second-generation Alexander Baring, became very powerful after 1815. Their capital base dramatically

\(^8\) Vansittart’s budget is contained in Hansard (1815: cols 795–822, June 14).
increased, from 1810 or so as their large long positions in Consols rose in value. Ricardo also became very rich by the same dynamic.

Postwar Surge in Consol Prices Causes Rentier Apotheosis

The deep discount debt issuance structure had major implications for the post–Napoleonic War economy because the volume of debt was so large, both in absolute terms and in relation to other assets in the economy:

- Britain’s debt-to-GDP ratio, at the peak in 1819, was about 260 percent, the highest of any country anywhere that has been successfully paid down—slightly higher than Britain in 1945. Of course, Britain had not borrowed 260 percent of 1819 GDP to fight its wars, but only about 170 percent in cash terms.
- Nearly all the Napoleonic War debt was issued around 60 percent of par or below, so when peace came holders would have a capital gain of 40 percent as Consols trended back toward par. Consol prices did not return to par quickly, however. They reached 77.5 percent in 1818 and almost 91 percent in 1824 at the peak of the boom (see Table 3). They were not to hit par until the 1840s.

More precisely, there was a total of £562 million of 3 percent debt outstanding on February 1, 1817, in the month preceding which Consols’ average yield had been 4.72 percent, for a price of 63.6 percent (ignoring accrued interest), with holders having already enjoyed a 7 percentage point price appreciation since the recent low of August 1815. By April 1824, Consols’ yield had dropped to 3.16 percent, so the price had risen to 94.9 percent. In the period between August 1815 and April 1824, holders of 3 percent Consols enjoyed a capital profit of 38.3 percentage points, or a total of £215 million on the £562 million of 1817’s 3 percent debt. An additional profit of perhaps £10 million would have been received on the £75 million of 4 percent Consols, as their price rose toward and above par. Total profits to Consols holders over this period would then have been around 75 percent of GDP, equivalent to a profit of $14 trillion in current U.S. dollars. In addition, investors’ 1824 pounds were worth around 20–30 percent more than their
1815 pounds had been based on any of the Gayer-Rostow-Schwartz price series, the Rousseaux price series, or the Lindert and Williamson “best guess” cost of living index (Mitchell 2011: 721–22, 737). Taking the annual figures for 1815 and 1824, these give a deflationary rise in value of 28 percent, 25 percent, and 18 percent,
respectively, primarily due to the resumption of gold payments in 1819–21.

One might say that this wealth was mostly brand new. Investors in August 1815 had put in only 57 percent for their 3 percent Consols and within a few years received a profit of 37 percentage points in capital gains plus interest at around 5 percent per annum on their initial investment. This profit was tax-free after the income tax had been abolished in 1816, and in any case capital gains had been untaxed. The capital gain was also permanent: once Consols had risen back to peacetime levels in the early 1820s, they fluctuated only moderately and indeed rose further until their price peaked late in the century.

Consequently, there was some 60 percent (from 1817) to 75 percent (from 1815) of GDP of new liquid capital in the early 1820s economy. In the early 1820s, this new money financed not only a number of subprime South American governments but also innumerable new factories and inventions that became the backbone of the Industrial Revolution.

Previous Periods of Rentier Enrichment and Impoverishment

It is interesting to observe that the same effect had occurred in the opposite direction in the 1790s. The price of Consols dropped from an average of 90 in 1792 to below 51 in 1797 and 1798. Since the nominal value of government debt in 1792 was about 120 percent of 1792 GDP, this fall caused a capital loss of about 47 percent of GDP in the first five years of the war.

Both Liverpool, the future prime minister (in 1796), and his father Charles Jenkinson (in 1798) wrote pamphlets proclaiming that Britain’s wartime economy was prospering, based on a substantial rise in exports and output over those years. However, the impoverishment of many savers during those years and the capital losses suffered by banks and other financial institutions starved the economy of capital. This latter point explains why Pitt in the 1790s had much more difficulty raising the necessary war finance than Perceval, Liverpool, and Vansittart did after 1807, and was a contributory factor to Britain’s going off gold in 1797.

One can trace this Consols’ wealth effect back further; Pitt’s benign economic conditions and growth in the 1780s were
largely caused by a rise in Consols’ prices from £55 to £90 between 1784 and 1792.

It is clear that, owing to Gideon’s financing technique, the size of the government bond market, and its importance in the national economy, the “wealth effect” of fluctuations in bond prices far exceeded any Keynesian stimulus from wartime spending.

Consols in the Context of Overall Wealth Holding

Government bonds were only one of the forms in which wealth was held in the years after 1815. However, unlike other wealth they were highly liquid, and indeed were the only security quoted on the stock exchange from its founding in 1801 until 1822. Over the course of the 18th century, government securities (“the funds”) and Consols in particular had become the principal non-landed form of wealth holding for merchants and the middle class, who tended to be more liquid than all but the richest aristocracy.

Nevertheless, agricultural land and to an increasing extent urban real estate remained an important component of British elite fortunes. Such forms of wealth storage as gold and silver plate had declined in importance since the 17th century as banks had proliferated, while bank deposits and insurance policies had appeared and at lower levels of wealth the savings bank movement was beginning its long 19th century climb.

The most important form in which wealth was held was still land, though its importance was beginning to decline during this period as mercantile relative wealth increased and landowners diversified their wealth into Consols. Prices of land had enjoyed a massive boom during the Napoleonic Wars, as corn prices had soared from an average of 43 shillings per quarter in 1794 to 127 shillings per quarter in the dearth year of 1812 (Mitchell 2011: 756). Then from 1813 to 1822 corn prices fell, bottoming out at 45 shillings per quarter in 1822 in spite of the 1815 Corn Laws, which had banned imports when the price was below 80 shillings. Jean-Baptiste Say, visiting England in late 1814, commented on the extraordinary profits made by agriculturalists during the war, and on their large investments both in bringing marginal land into cultivation and in mechanizing and up-scaling their operations (Say 1815).

Some of the landowner wealth acquired during the war was redeployed into industrialization, both during the war and in the
years after, partly because the profitability of agricultural land declined after the war and did not recover for several decades. Consider the case of John Crichton-Stuart, 2nd Marquess of Bute (1793–1848). In 1814, he inherited a very large and liquid landed estate from his grandfather, including major land holdings in South Wales acquired through the 1st Marquess’s advantageous marriage as well as substantial holdings in the funds. In 1817 he began surveying the Glamorgan coal fields, consolidating his local land holdings as he did so and building the Welsh coal mining industry during the 1820s. Between 1822 and 1848, he also developed the Cardiff Docks, opening the new Docks in 1837 at a cost of £350,000. His activity in both areas brought him huge debts of £494,000 at his death, although his assets greatly exceeded that value, and their profitability developed further, so that in the 1870s his grandson and heir was claimed (probably incorrectly, given the rise of American fortunes by that stage) to be the richest man in the world.

Between 1813 and 1822 the agricultural prosperity went into reverse. While costs fell somewhat, with the deflation attendant on returning to the gold standard in 1819–21, income from land fell considerably further, and landowners who, unlike Bute, did not have substantial outside holdings were sorely pressed. Their situation is well illustrated in an 1822 conversation between the diarist Harriet Arbuthnot, the young (28) wife of Charles Arbuthnot, a Treasury Secretary (junior minister) whose wealth was primarily in Consols, and her older (62) cousin the 10th Earl of Westmorland, Lord Privy Seal (a cabinet minister) and a large landowner:

I insisted that the cry of agricultural distress was grossly exaggerated, that the pressure felt by that class now was *nothing* compared to that suffered by the manufacturing districts in 1819, when 10,000 able and willing workmen were starving in one town (Glasgow) for want of work, and every other manufacturing town suffering in a like degree; that the farmers had got into luxurious habits which they did not choose to give up and therefore joined in the cry of “No Taxes;” that those

---

9This is largely consistent with the theoretical analysis of Ventura and Voth (2015): landholders experienced wealth gains from their Consols that led them to switch out of low-return agricultural investments, thereby lowering factor demand in the old sectors and raising profitability in the new sectors.
persons who had borrowed money at 5 per cent interest during the war to buy land would suffer as any body did who borrowed money to buy any thing; but that the great landed proprietors would weather the storm and be just as well off as ever; that there would certainly be a transfer of property, but that the country generally would be as rich as ever and that there could not be a greater proof of it than the enormous increase of the Revenue in every branch. He said I talked nonsense and was very childish (which, by the way, is no argument) and that the transfer of property I talked of so quietly would be a greater revolution than even a bankruptcy of the funds.¹⁰

Shares were a relatively insignificant store of wealth in post-1815 Britain, with total public share capital paid in of £49 million in 1827, of which £34 million related to companies formed before 1824. The 624 companies floated in the bubble of 1824–25 had a nominal capital of £372 million, but only £15.2 million of this capital was ever paid in; by 1827, when the dust settled, that had declined in value to £9.3 million. Share prices in 1815–27 were not especially buoyant, in spite of the bubble of 1824–25; on a quarterly average basis the share index rose by only 61 percent, from 3.1 in the second quarter of 1815 (1962 = 100) to 5.0 at the peak in the third quarter of 1824. The bubble thereafter deflated, with the index in the fourth quarter of 1825 being only 4.44, and 4.1 in the fourth quarter of 1827. Changes in share wealth were clearly not economically significant during these years.¹¹

There was one final form of nonland wealth that underwent a major boom in the early 1820s before collapsing in 1825–26: foreign bonds. In 1817–18 several European governments floated bonds totaling £10 million, while in the boom of 1821–25 some £43 million¹² was invested in foreign bonds, of which £37 million related to Latin American issues. These investments were a substantial absorber of British capital, especially during the 1822–25 period as declining yields on Consols caused investors to

¹⁰Bamford and Wellington (1950: vol. 1, 139), diary entry for February 2, 1822.
¹²Figures quoted on foreign bond issues are from Neal (1998: 63) and the detail is from Gayer, Rostow, and Schwartz (1953: vol. 1, 188–89).
seek higher returns elsewhere. However, the profits and losses involved do not appear to have been significant, except in the immediate aftermath of the 1825–26 crash.

Who Were the Rentiers?

Information on the owners of Consols can first be gained by examining the estates of the very wealthiest. Three died between 1809 and 1839 who are believed to have been Britain’s first millionaires: (1) Nathan Rothschild (1836), worth £5 million, whose fortune was largely in Consols and other liquid securities, and who is known to have grown very much richer in the decade after 1815; (2) the 1st Duke of Sutherland, formerly Marquess of Stafford (1833), worth £2 million, whose fortune was largely in land, but with substantial agricultural and industrial development including the Bridgewater Canal; and (3) Sir Robert Peel (1830) father of the future prime minister, worth £1.5 million, whose fortune was derived from the family cotton mill business that had employed 15,000 people at its peak, but by his death consisted largely of Consols, houses, and land.13

Lower down the scale it is notable that the five-year average of the annual number of nonland fortunes above £100,000 among the newly deceased took a sharp upward turn, from 16.5 in 1809–14 and 18.2 in 1815–19 to 25.6 in 1820–24, and 29.0 in 1825–29. It then did not rise significantly further until after 1855 (Robinson 1985: 47). Although Robinson could think of no explanation for this increase, the jump’s coincidence with the recovery in Consols’ prices suggests that Consols formed a high proportion of the wealth covered by these statistics, which derive from probate valuations that do not include real property and therefore relate primarily to mercantile rather than aristocratic fortunes. At the same time, including land changes the wealth picture considerably: more than 80 percent of people worth more than £100,000 in 1825 were primarily landowners.

Of the nonlanded rich dying in the 1820s, a high proportion were in the City of London, either as merchants, bankers, or in insurance or stockbroking—15 of the 24 with nonland wealth from £500,000 to £1 million died between 1820 and 1839. It can safely be assumed that the great majority of these fortunes would have

---

13 Data on these holdings of the very rich are contained in Robinson (1985).
been held in Consols. Of the other nine, the three in textiles, the one in publishing, and the two in public administration and defense will also have held a high percentage of their nonland wealth in Consols, since none of those occupations is capital intensive. Only three of the 24, with principal occupations in iron/steel, chemicals, and brewing may have held relatively small stocks of Consols, with the majority of their nonland wealth in their primary businesses.

The same analysis carried out on the “lesser wealthy” who died with nonlanded wealth between £100,000 and £500,000 in the years 1809–29 shows a similar pattern: 117 of the 154 such people either were merchants, worked in finance, were professionals, or worked in public administration. Since these people had little need for plant and equipment, the majority of their nonland wealth would have been held in Consols.\(^{14}\)

We can then conclude that the capital gains from the Consols’ price increase in 1815–24 had a major effect on the net worth of nearly all those with substantial nonlanded wealth, many of whom like Bute may also have had large landholdings.

In this context, the wealth accumulation patterns of early 19th century industrialists, from Sir Robert Peel downward, should also be borne in mind. There was no public equity market available to raise money for industrial companies, and bank loans were not available for long maturities except for a moderate proportion of the value of buildings and land. Hence, businesses were built up by collaboration between entrepreneurs through private partnerships, formed, and dissolved as needed.\(^{15}\)

The members of those partnerships, once the initial investment in plant had been made, kept substantial balances of liquid assets to

\(^{14}\)Data on the occupations of the dying wealthy in these paragraphs come from Robinson (1985: 82–87).

\(^{15}\)Manufacturers themselves did not borrow for long-term plant investment, but accumulated retained earnings and savings, much of them in Consols, which were more secure than the local English banks. The brokers and embryonic “merchant banks” did not invest directly in industry, but country banks lent to it, and would have been encouraged to lend more by the Consols-driven increase in their capital. The capital gains accrued to small holders as well as large and gave manufacturers more capital with which to expand their own factories or invest in other people’s. The holdings of Consols in manufacturing districts were also more significant for industrialization than those among the gentry, who had limited ability to access industrial investment opportunities and tended like Mrs. Arbuthnot to buy South American bond issues and shares, mostly in canals and early railways.
Apotheosis of the Rentier

meet the need for new investment and to protect the business against downturns, which could be severe and unexpected. Consols, utterly safe, highly liquid, and yielding a decent return (and a superb return in 1815–24) were a favorite holding for such people, who in the early 1820s found themselves through their Consols’ holdings very much richer than before. It is then fair to presume that Consols’ windfalls paid for much of the industrial expansion in the 1820s and its accompanying capital intensification.

Economic Effects of Rentier Apotheosis

Britain’s economy in 1814 was already qualitatively different from any other in the world, or from its own state 30 years earlier. The French economist Jean-Baptiste Say visited Britain in late 1814, after the peace but before Napoleon’s return from Elba, and set out his impressions in a pamphlet De l’Angleterre at des Anglais:

But it is principally the introduction of machines in the arts, which has made the production of riches more economical. There are almost no big farms in England, where for example, threshing machines are not used, by means of which, in a large operation, more work is done in a day than was done in a month by ordinary methods.

At last human labor, which has rendered the high cost of consumer goods so expensive, is in no circumstances replaced so advantageously, as by steam engines, improperly called “fire-pumps” by some. There is no work which cannot be reached to be executed by them. They go to the mills, weaving cotton and wool; they brew beer, they cut crystals. I have seen them embroider muslin and beat butter. At Newcastle, at Leeds, I have seen moving steam engines dragging after them carts of coal; and nothing is more surprising, at first sight, for a traveler, than to meet in the country long convoys which advance by themselves and without the help of any living being.

 Everywhere steam engines are prodigiously multiplied. There were no more than two or three in London thirty years ago, there are thousands at present. There are hundreds of them in the large manufacturing towns; one sees them even in the countryside, and industrial works could not be sustained profitably without their powerful help. But they must
have abundant coal, that combustible fossil which Nature seems to have reserved to supplement the exhaustion of forests, the inevitable result of civilization. Thus one could, with the help of a simple mineralogical map, trace an industrial map of Britain. There is industry everywhere there is coal in the ground [Say 1815: 29–31].

Say also noted that high British taxation and the gigantic government debt were of immense importance in raising the cost of everything in Britain and depressing the profits of business and commerce. He believed British goods would be uncompetitive in Europe once the European economies had recovered from the war. Writing while Consols’ prices were still depressed, Say failed to foresee the effect of their subsequent rise. Liverpool however foresaw that effect in the following year, during the postwar economic downturn. Writing to George Canning on February 13, 1816, and discussing the clamor to abolish the income and malt taxes (which in the event succeeded) he wrote, “I am satisfied likewise that those who raise this clamour have a narrow view of their own interest, as the restoration of public credit, the run of the funds, and the consequent fall of the interest on money will afford more relief to the existing distress of the country than any other measure of relief that could be adopted.”

Table 4 sets out U.K. (including Ireland) population, GDP, and GDP per capita figures for 1801–41, at 10-year intervals, with five-year additions for 1816 and 1826. The GDP figures are of course a backward extrapolation of a statistic (and a questionable one at that!) that was only invented a century later, but for what they are worth they show a distinct acceleration in growth of per capita GDP, from 2.16 percent per annum to 3.11 percent per annum between the decades 1811–21 and 1821–31. Note also that, with population expanding at a very rapid rate of around 1.3 percent annually, the 1820s growth rates are spectacular even by modern standards, indeed considerably higher than was to be achieved in any subsequent decade.

16 Martin Hutchinson’s translation.
17 Quoted in Yonge (1868: vol 2: 253–55). The italics are ours, and we take “the run of the funds” to be a direct reference to the price rise to be expected as public credit recovered.
Apotheosis of the Rentier

Gayer, Rostow, and Schwartz use brick production as a measure of economic activity in capital goods; this increased from an annual average of 791.2 million in 1815–18 to 992.4 million in 1819–22 to 1,501.7 million in 1823–26; the 1825 figure of 1,948.8 million was the highest in any year before 1846 (Gayer, Rostow, and Schwartz 1953: 121, 147, 186).

Other indicators show a similar picture of economic acceleration after 1820. Coal shipments from Newcastle and Sunderland rose from an annual average of 1,083 thousand chaldrons in 1814–19 to 1,249 thousand chaldrons in 1820–26 (Mitchell 2011: 243). Pig iron production rose from an annual average of 305 thousand tons in 1814–19 to an annual average of 453 thousand tons in 1820–26 (Mitchell 2011: 280). Raw cotton consumption rose from an annual average of 96 million pounds in 1814–19 to an annual average of 147 million pounds in 1820–26 (ibid., 332). Manufactured goods exports at constant prices, mostly textiles, rose from £15.3 million in 1814–19 to £17.9 million in 1820–26 (ibid., 520).

18 The chaldron was a volume measure used for coal. By a law of 1694, the Newcastle chaldron of coal was defined to weigh 5,940 pounds.

### TABLE 4

<table>
<thead>
<tr>
<th>Year</th>
<th>Nominal GDP (£m)</th>
<th>Price Level (1826 = 100)</th>
<th>Real GDP (millions of 1826 £s)</th>
<th>Population (thousands)</th>
<th>Real per capita GDP (£s)</th>
<th>Annual real per capita GDP Growth (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1801</td>
<td>257</td>
<td>155.7</td>
<td>165.1</td>
<td>16,300</td>
<td>10.1</td>
<td>—</td>
</tr>
<tr>
<td>1811</td>
<td>334</td>
<td>145.4</td>
<td>229.7</td>
<td>18,500</td>
<td>12.4</td>
<td>2.06</td>
</tr>
<tr>
<td>1816</td>
<td>328</td>
<td>118.6</td>
<td>276.6</td>
<td>19,700</td>
<td>14.0</td>
<td>2.49</td>
</tr>
<tr>
<td>1821</td>
<td>322</td>
<td>99.7</td>
<td>332.0</td>
<td>21,000</td>
<td>15.4</td>
<td>1.84</td>
</tr>
<tr>
<td>1826</td>
<td>403</td>
<td>100.0</td>
<td>403.0</td>
<td>22,500</td>
<td>17.9</td>
<td>3.09</td>
</tr>
<tr>
<td>1831</td>
<td>480</td>
<td>95.3</td>
<td>503.7</td>
<td>24,100</td>
<td>20.9</td>
<td>3.13</td>
</tr>
<tr>
<td>1841</td>
<td>520</td>
<td>97.7</td>
<td>532.2</td>
<td>26,900</td>
<td>19.8</td>
<td>−0.55</td>
</tr>
</tbody>
</table>

**Sources:** For GDP, [www.ukpublicspending.co.uk/year_spending_1841UKmm_16mc1n#ukgss302](http://www.ukpublicspending.co.uk/year_spending_1841UKmm_16mc1n#ukgss302); for population and prices, Mitchell (2011: 11, 721).
The Hoffman index of industrial production rose from an average of 9.39 in 1815–18 to 10.67 in 1819–22 and then to 12.28 in 1823–26 (ibid., 431). Perhaps most significantly, the annual average number of English patents granted fell slightly from 114 in 1815–18 to 105 in 1819–22, but then soared to 177 in 1823–26—surely a sign of entrepreneurial activity in those years, though it fell back to 153 in 1827–30 (ibid., 438).

Thus, the overall real return to Consols’ investors over the period 1813–24 was in excess of 300 percent. If they held on and reinvested income during that period, Consols holders quadrupled their money in real terms.

The mechanism by which British economic growth benefitted from the 1815–24 rise in Consols’ prices is well set out in Gayer, Rostow, and Schwartz (1953: vol. 1, 185):

The character of the (1822–26) cycle as a whole, however, can be distinguished from earlier cycles by the scale and the scope of new private investment. There was an increase in railways construction, new docks were built, and what appears to be the greatest building boom until the forties took place. Gas-light, insurance, building, trading, investment, provision companies, in addition to many others, were formed on a large scale. These, the fluctuations of foreign government and mining issues and the fabulous Stock Exchange boom and crash (1824–25) impart to these years their unique character.

The economics is clear. For the British economy to industrialize, its capital intensity had to increase. This increase accelerated in the decade after Waterloo, and Consols’ profits provided wealth enabling this increase to take place more quickly, more smoothly and with less pressure on workers’ living standards than would otherwise have occurred.19

19 There is a good counterexample in the 1831–46 period that supports our main thesis, a decade without significant Consols’ profits but with the massive capital needs for the railways. During this period, the economy underwent the “Engels pause” without rising living standards, as manufacturing was starved for capital. Engels believed living standards would never rise and had never risen. He was wrong but writing in 1844, after a decade of stagnant real wages and the 1820s ebullience had been forgotten. Our point is that, without the Consols’ profit, new capital needs for the industrializing and capital-intensifying economy were hard to come by, and could only be gotten slowly by the workers being squeezed.
Apotheosis of the Rentier

Gayer, Rostow, and Schwartz (1953: vol. 1, 185) blame the excess speculation in 1824–25 on two additional factors, both of which seem pertinent. First, in April 1822, the government had extended the privilege of country banks issuing £1 and £2 notes, which had the effect of an unexpected increase in the money supply. Second, the double refinancing of 5 percent and 4 percent Consols in 1823–24 produced “a restless feeling and a disposition to hazardous investment,”20 and a surge in foreign bond purchases.

The extent of the speculative fervor was spelled out by a contemporary stockbroker, Henry English, and his analysis has remained authoritative to this day.21 Briefly, English listed 624 companies that were floated in the years 1824 and 1825. They had a capitalization of £372,173,100. By 1827, only 127 of these existed with a capitalization of £102,781,600, of which only £15,185,950 had been paid in; but the market value had sunk even lower to only £9,303,950. Of the 127 still existing in 1827, 44 were mining companies (mostly in Latin America), 20 gas companies, 14 insurance, and 49 miscellaneous, including salt, silk, lard, bridge, emigration, glass distilling, brick, bank, and agricultural companies.

In addition, 15 railway companies were formed by Acts of Parliament in 1823–26, for a total projected construction cost of £3.64 million, of which the Stockton and Darlington, authorized in 1823 at a cost of £450,000 and opened in 1825, and the Liverpool and Manchester, authorized in 1826 at the immense construction cost of £1,832,375, more than all the rest put together, and opened in 1830 in the presence of the Duke of Wellington, are the most famous. The experimental and dangerous nature of this technology is indicated by the fact that the Stockton and Darlington used steam trains only for goods traffic, with passengers being drawn by horse carriages until 1833, while the Liverpool and Manchester accidently knocked over and killed the former Cabinet Minister William Huskisson at its opening on September 15, 1830.

Mrs. Arbuthnot was a keen observer of the stock market bubble of 1825, which she discussed in detail with Wellington, who thought it would all end in a crash (as did Liverpool, who said so

in the House of Lords). However, she could not resist a good speculation herself:

There is a railway going to be made between Liverpool and Manchester which promises to answer immensely. We have 10 shares in it for which we gave £3 a piece and which are now worth above £58 each and they are expected to be worth above £100. I am very fond of these speculations and should gamble greatly in them if I could, but Mr. Arbuthnot does not like them, and will not allow me to have any of the American ones as their value depends on political events and he thinks in his official situation it would be improper [Bamford and Wellington 1950: vol. 1, 382].

Improper or not, Mrs. Arbuthnot got in on the ground floor. Her diary entry was dated March 16, 1825, and the Liverpool and Manchester Railway Company had been formed in 1823. However, its first attempt to get Parliamentary authorization in 1825 failed (largely through opposition by the Marquess of Stafford, owner of the competing Bridgewater Canal), and it gained authorization only at a second attempt in May 1826.

Since Arbuthnot had been what today would be called the government’s Chief Whip in the House of Commons (though he had given up that job in 1823 while remaining a junior minister as Commissioner for Woods and Forests) his 10 shares (of an eventual total of 4,233, spread among 308 shareholders, with Stafford eventually owning 1,000) may have been a special allocation. It was however not quite the handout it appears. Following corporate finance practices of the time, the shares will have been partly paid, with say £10 paid on a £100 share, and Arbuthnot will have been called for the other £90 when Parliament had approved the company and construction began. While £900 was affordable for the comfortably off Arbuthnotts, larger sums would not have been, so Arbuthnot was right to stop his wife punting on South American mining companies. Nevertheless, their Liverpool and Manchester Railway investment was a successful one, if they kept it; the company paid an average dividend of 9.5 percent in its 15 years as an independent company.

Since the Bubble Act, requiring an Act of Parliament to form a new company, was not repealed until 1826, much of the new
investment in 1822–25 was informal and small-scale, in start-ups and other kinds of experimentation. The surge in patents granted in 1823–26 itself confirms that the surge in Consols’ wealth allowed exploration of projects that were not immediately profitable. One such project, far more famous in 2017 than it was at the time, was that carried out by Charles Babbage.

Babbage was a typical beneficiary of the surge in Consols’ prices. His father was a partner in the banking house of Praed & Co., whose wealth, like that of most bankers, will have been substantially augmented by the rise in Consols. Babbage was partly supported by his father and inherited a fortune of £100,000 when his father died in 1827. However, the initial funding for his difference engine project came not from his father but from Liverpool’s government, which made a grant to Babbage of £1,700 in 1823 to start work on the project. Alas, endless redesigns meant the prototype was never finished, and the government, under the unimaginative Sir Robert Peel (son of the textile millionaire) canceled further funding for the project in 1842 after it had absorbed £17,000 of public money.

In spite of Babbage’s failure, the statistics show that the pace of innovation, industrialization, and Britain’s overall economic growth quickened substantially in the early and middle 1820s, and that the new wealth produced by the return of Consols’ prices to peacetime levels substantially contributed to this quickening. Britain’s industrial lead established during this period was not to be relinquished until the last quarter of the century.

Conclusion

The rise in Consols’ prices in 1815–24, under the able economic management of Lord Liverpool created not the “euthanasia of the rentier” of Maynard Keynes’ leftist fantasies but his apotheosis, in which, combining the rise in Consols and given the decline in prices that accompanied the return to gold, rentier nonland wealth nearly quadrupled in real terms during Liverpool’s premiership. That increase in wealth coincided with and likely produced a surge in capital investment and innovation, which, apart from the follies of the Latin American bond market, produced the first great surge of the Industrial Revolution, with U.K. economic growth accelerating to unprecedented levels.
References

_________ (1817) The Parliamentary Debates from the Year 1803 to the Present Time, Vol. 36.
Untangling Hair Braider Deregulation in Virginia
Edward J. Timmons and Catherine Konieczny

We estimate the effects of removing the license requirement for hair braiding in Virginia in 2012. Using County Business Patterns and Nonemployer Statistics data from 2004 through 2014, we find evidence at the state level that deregulation has created more opportunities for smaller owner-operated beauty salons (an increase in proprietor density of more than 8 percent). A simple statistical test confirms that Virginia counties experienced beauty shop growth at a rate approximately 7 percent higher than that in contiguous counties in bordering states. Taken together, our findings support the notion that deregulation of hair braiding has enhanced economic opportunity for hair braiders in Virginia.

Thirteen U.S. states continue to require hair braiders to obtain a cosmetology license. However, the particular skills required for traditional, natural hair styles are generally not covered in the entry requirements for cosmetology licensure. Further, the unique risks to consumers from poor hair braiding practices (such as hair loss from braids that are too tight) are not addressed in cosmetology training. The result is that hair braiding licensure may not improve consumer safety, although it does lead to significant labor market costs.

In this article, we analyze Virginia’s deregulation of the hair braiding occupation in 2012. Using the states of Kentucky, North Carolina,
and West Virginia as well as bordering counties, we examine whether removing license requirements had a significant effect on the number of salons, the number of salon employees, or the level of salon employee wages. We hypothesize that removing this barrier to entry will ultimately increase the total number of beauty shops, potentially creating more opportunities for small beauty shops by decreasing the cost of becoming a practitioner and increasing the supply of braiders. Before turning to our analysis, we discuss the rationale for occupational licensing and provide some background on regulation of hair braiding.

Rationale for Occupational Licensing

The common defense for occupational licensing is that it improves the quality of services delivered to consumers. Some economists have proposed theories that occupational licensing improves the human capital of practitioners and serves as a signal of high quality (Shapiro 1986). If such theories are accurate, then although occupational licensing will increase the price of services, consumers will benefit by receiving higher-quality services. But valid arguments can be made against such theories. For example, in many instances, consumers demand proof of qualification on their own when they believe there is reasonable risk of harm. It is thus a mistake to assume that, without government-sanctioned licenses, consumers would have no indicators of service quality or would choose to ignore such indicators (Thierer et al. 2015). Indeed, advances in technology (e.g., internet rating services) have enhanced consumers’ ability to compare the quality and reputation of professional service providers.

In several professions, arguments for occupational licensing are particularly weak. In the case of florists, for example, the risks associated with unlicensed florists amount to far less than the costs of licensing for consumers and aspiring florists, and hair braiding appears to belong to this category (Carpenter 2011). By their nature, licenses do not differentiate practicing professionals within a given field, but rather only differentiate those who can practice from those who cannot practice, even excluding licensed practitioners from another state. All license holders meet the same minimum entry standards, but whether those entry standards correlate with enhanced quality remains unclear. Licensing may instead serve as a barrier to entry for the profession and may even discourage hopeful
professionals from aspiring to more than the minimum entry standards. The theory that licensing creates monopoly power for professionals by restricting entry to certain professions has emerged as the leading rationale for occupational licensing (Friedman 1962).

Empirical evidence supports the view that licensing exists because creating a barrier to entry benefits individuals already working in a given field—not because it enhances consumer welfare. Studies generally find that, although licensing is correlated with higher wages for professionals, it has ambiguous effects on the quality of services provided (White House 2015). Other studies on occupational licensing focus on employment effects in the field overall and on individual workers. Using national data from the Survey of Income and Program Participation and controlling for observable characteristics (including occupation), researchers find that licensed workers are more likely to be employed than are similar workers without licenses or certification (Gittleman, Klee, and Kleiner 2018). The same study also finds that those with federally issued credentials may earn 8.9 percent more than those without, while those with state-issued credentials may earn 6.1 percent more. An earlier study analyzing the effects of licensing and certification finds that licensed workers earn about 18 percent higher wages than do unlicensed workers (Kleiner and Krueger 2013). Licensing may also restrict mobility. A recent study finds that state-specific licensing exam requirements are associated with a 36 percent reduction in mobility (Johnson and Kleiner 2017). Licenses are generally not transferable from state to state and as a result may result in professionals experiencing “job lock.” Occupational licensing becomes particularly burdensome in the case of hair braiding because the service falls under the umbrella of cosmetology. The intent of policymakers is to ensure proper sanitation training in an effort to protect consumers, but an examination of the skills tested in the process of obtaining a license tells a different story. We now turn to a discussion of regulation of the hair braiding occupation.

Regulation of Hair Braiding Nationally

Table 1 and Figure 1 provide a snapshot of regulation of hair braiding across the United States as of October 2017. Thirteen states and Puerto Rico require aspiring hair braiders to obtain a cosmetology license. Fourteen states have specific hair braiding licensing
**TABLE 1**  
**Summary of Hair Braider Regulation, October 2017**

<table>
<thead>
<tr>
<th>Require a Cosmetology License</th>
<th>Require a Braiding License</th>
<th>No Braiding License Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>Alabama: 200 hours</td>
<td>Arizona</td>
</tr>
<tr>
<td>Hawaii</td>
<td>Florida: 16 hours</td>
<td>Arkansas</td>
</tr>
<tr>
<td>Idaho</td>
<td>Illinois: 300 hours</td>
<td>California</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>Louisiana: 1,000 hours</td>
<td>Colorado</td>
</tr>
<tr>
<td>Missouri</td>
<td>Minnesota: 30 safety and sanitation hours</td>
<td>Connecticut</td>
</tr>
<tr>
<td>Montana</td>
<td></td>
<td>Delaware</td>
</tr>
<tr>
<td>New Jersey</td>
<td>Nevada: 250 hours</td>
<td>Georgia</td>
</tr>
<tr>
<td>New Mexico</td>
<td>New York: 300 hours</td>
<td>Indiana</td>
</tr>
<tr>
<td>North Dakota</td>
<td>North Carolina: 300 hours</td>
<td>Iowa</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>Ohio: 450 hours</td>
<td>Kansas</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>Oklahoma: 600 hours</td>
<td>Kentucky</td>
</tr>
<tr>
<td>Vermont</td>
<td>Oregon: online module and written exam</td>
<td>Maine</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Pennsylvania: 300 hours</td>
<td>Maryland</td>
</tr>
<tr>
<td>Wyoming</td>
<td>South Carolina: 6 safety and sanitation hours</td>
<td>Michigan</td>
</tr>
<tr>
<td></td>
<td>Tennessee: 300 hours</td>
<td>Mississippi: requires registration</td>
</tr>
</tbody>
</table>

**NOTE:** Florida’s hair braiding license forbids licensees from providing hair extensions.  
**SOURCE:** Institute for Justice (2017).
Hair Braider Deregulation

requirements that are generally less burdensome than are the requirements for cosmetology licensing. The remaining 23 states and Washington, D.C., currently do not require hair braiders to obtain a license to work. Florida stands out as an interesting case—although a 2011 statute created a distinct hair braider license, the law prohibits hair braiders from providing hair extensions—an essential service for many braided hairstyles (Florida Department of Business and Professional Regulation). As a result, hair braiders in Florida are generally still required to obtain a cosmetology license.

Significant changes in regulation of hair braiding have come about as a result of both legislation and litigation. The Institute for Justice (IJ), a nonprofit, public-interest law firm, has championed many of the legal challenges to licensure. In these cases taken on by IJ, individual hair stylists sue states for the right to practice without a license. IJ lawsuits have led to Arizona, California, Mississippi, Utah, and Washington fully deregulating hair braiding (Avelar and Sibilla 2014). Maryland, however, is an example of the opposite

FIGURE 1
Summary of Regulation of Hair Braiding, October 2017

In 2015, a stylist actually petitioned the state to try to impose licensing requirements on new hair braiders (Maryland House Bill 1124 of 2015). The bill failed but was supported by natural hair stylists who were licensed cosmetologists and who were facing reduced business because of competition from new hair braiding businesses. One practitioner advocating for the license requirement said, “At least it will weed out those [stylists] who are really really bad” (DePhillis 2015).

A number of states have recently deregulated hair braiding. Arkansas, Colorado, and Washington removed licensing requirements for hair braiders in 2015, and Iowa and Kentucky followed soon after in 2016. A number of states (South Dakota and Indiana, for example) exempted hair braiders from cosmetology licensing requirements in 2017. Texas had a particularly interesting case that focused on the requirements for teaching hair braiding. That case, brought by IJ (Institute for Justice 2015), challenged a rule that required a hair braiding instructor to provide barber chairs and sinks, even though hair braiding courses addressed neither hair washing nor hair treatment.

### Understanding Regulation of Hair Braiding in Virginia

Before 2012, individuals offering hair braiding services in Virginia were required to obtain a full cosmetology license. Obtaining a license required an individual to attend training for 170 hours, pass an exam, and pay annual dues to Virginia’s Board for Barbers and Cosmetology. To be eligible to sit for the cosmetology licensing exam, an applicant would be required to attend training at a state-approved cosmetology or hair braiding school. Previous experience could not be substituted for the training requirement for an initial cosmetology license. Table 2 contains a complete list of the fees required to obtain a cosmetology license in Virginia before 2012, including those associated with training.

Following a statement by the Virginia Board of Barbers and Cosmetology, a law introduced in 2012 completely removed the

---

1 No distinction was found between the two types of schools. There may be additional regulations when institutions are regulated as cosmetology schools, including what training schools must offer and what certifications teachers must have, which would have indicated additional barriers to entering either profession.
cosmetology license requirement for hair braiders. The statement cited the low number of interventions the board had had to make in instances of improperly practiced hair braiding during a five-year period from 2007 to 2011. These included two fines for hair braiders, one license revocation, and one fine for a hair braiding salon with unspecified infractions (Virginia Governor’s Executive Reorganization Plan 2012).

In contrast, few changes have been made in the states surrounding Virginia between 2004 and 2014. In 2011, North Carolina introduced a specialized hair braiding license that requires 300 hours of training (Burrows 2010). West Virginia required a full cosmetology license for hair braiders until 2017. H. B. 2131, a piece of legislation introduced in 2015 that would have removed the license requirement, failed to get passed into law. Kentucky eliminated its license requirement in 2016 (Powers 2016). For the present article’s period of study (2004 to 2014), both Kentucky and West Virginia required

<table>
<thead>
<tr>
<th>Fee Type</th>
<th>Amount Due</th>
<th>When Due</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individuals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>$55</td>
<td>With application</td>
</tr>
<tr>
<td>License by endorsement</td>
<td>$55</td>
<td>With application</td>
</tr>
<tr>
<td>Renewal</td>
<td>$55</td>
<td>With renewal card before expiration date</td>
</tr>
<tr>
<td>Reinstatement</td>
<td>$55</td>
<td>With reinstatement application</td>
</tr>
<tr>
<td><strong>Salons</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>$90</td>
<td>With application</td>
</tr>
<tr>
<td>Renewal</td>
<td>$90</td>
<td>With renewal card before expiration date</td>
</tr>
<tr>
<td>Reinstatement</td>
<td>$90</td>
<td>With reinstatement application</td>
</tr>
<tr>
<td><strong>Schools</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>$120</td>
<td>With application</td>
</tr>
<tr>
<td>Renewal</td>
<td>$120</td>
<td>With renewal card before expiration date</td>
</tr>
<tr>
<td>Reinstatement</td>
<td>$120</td>
<td>With reinstatement application</td>
</tr>
</tbody>
</table>

Source: Fee information is from 18 VAC 41-30-110.
hair braiders to have a full cosmetology license. We therefore use Kentucky, North Carolina, and West Virginia, as well as bordering counties, as a control group to examine the employment effects of Virginia’s removal of hair braiding regulation.²

Few studies have been conducted on licensing for hair braiders. A recent study performed by IJ explores whether hair braiding presents risks to consumers (Erickson 2016). A review of data for nine states and Washington, D.C., from 2006 to 2012 indicated that just 95 complaints were filed against hair braiders (approximately 1 percent of the population of licensed hair braiders). In addition, virtually all of those complaints (94 of the 95) were filed by competing cosmetologists as opposed to consumers. The study also notes significant differences between Mississippi and Louisiana with respect to the number of professional hair braiders. Mississippi (no licensing requirements for hair braiders) had 1,200 hair braiders in 2012, whereas Louisiana (licensing requirements for hair braiders) had only 32 hair braiders.

As demonstrated by Maryland’s 2015 attempt to protect currently licensed hair stylists, there is some substitution among braiders and other cosmetology establishments. Cosmetology and barbering establishments cater to a broader clientele, and there are more kinds of establishments (Timmons and Thornton 2010). The market has also adapted to satisfy varying consumer preferences. Consumers can now frequent barbers dedicated to beard styling, blow-dry bars, unisex salons, traditional salons that offer complimentary alcoholic beverages with the purchase of a service, or hair braiding salons. Some customers prefer traditional salon services bundled with hair braiding. Other consumers prefer (and should be able to visit) salons that specialize solely in hair braiding—and that offer their services at a reduced cost because they have to neither outfit their facility for hair washing nor pay for cosmetology licenses for their stylists. Virginia’s deregulation of hair braiding allows for this very type of salon, and we seek to expand the existing literature by estimating the economic effects of the 2012 regulatory change on hair braiding in Virginia. Before discussing the results of our investigation, we turn to a discussion of the data.

²It is also possible to measure the effect of changes in states that have not completely removed licenses. Some may remove the fee for a license, decrease the number of exams, or decrease the hours of training required. We chose to focus instead on the complete removal of regulation.
Hair Braider Deregulation

The Data

We consider recent changes to Virginia law to empirically review the effects of deregulation of the hair braiding occupation. The investigation is divided into two parts: the first part is a state-level comparison of Virginia with contiguous states with different levels of hair braiding regulation; the second part discusses a simple statistical test to confirm the significance of our comparison in the previous section.\(^3\) Within these parts, we differentiate between trends in proprietor establishments that are run by the owner and trends in employer establishments with multiple employees.

Our analysis is based on data retrieved from the County Business Patterns (CBP) report and the Nonemployer Statistics (NES) reports published by the U.S. Census Bureau. Both are broken down by NAICS code: we focus on code 812112—beauty salons. Beauty salons (also referred to as “beauty shops” in this article) include all establishments engaged in “(1) cutting, trimming, shampooing, coloring, waving, or styling hair; (2) providing facials; and (3) applying makeup (except permanent makeup).” Excluded from this category are establishments catering to men specifically, which are classified under NAICS 812111—barber shops. Although it would be ideal to focus solely on hair braiding, we are not aware of any data that focus solely on that specialization. Instead, we rely on the broader beauty salon category while acknowledging the limitations of this approach—most notably, that we may be identifying exogenous changes in the cosmetology market that are unrelated to hair braiding.\(^4\) To our knowledge, there were no changes in the requirements for cosmetology licenses in our sample areas, but there may be exogenous changes to demand that are otherwise unaccounted for. We use four outcomes from the CBP report at the state level: number of establishments, number of employees, number of employees per establishment, and annual earnings per employee. Measures were adjusted for inflation and converted into 2009 U.S. dollars. At the county level, because of limited data, we are able to observe only

\(^3\)Tennessee also borders Virginia, but we were unable to find any data at the county level on bordering counties in Tennessee. We therefore omitted Tennessee from this part of the analysis.

\(^4\)We are not aware of any significant changes in cosmetology licensing (besides changes to hair braiding) over our time period of study.
the number of establishments. The NES reports fewer variables, and we can track only the number of sole proprietors of beauty salons.

Analysis of Removal of Hair Braiding Regulation

State Level

Figures 2 through 5 plot trends in the number of beauty shop establishments, the number of beauty shop employees, the number of beauty shop proprietors, the number of employees per beauty shop, and the average earnings per beauty shop employee from 2004 to 2014 in Virginia (which fully deregulated hair braiding in 2012) and in Kentucky, North Carolina, and West Virginia (which maintained some form of regulation of hair braiding throughout that time period). The numbers of beauty shop establishments, employees, and proprietors are weighted by the state population. In Figure 2, we observe little change in the number of beauty shops with employees (per 10,000 residents) in Kentucky and North Carolina in 2012. Both Virginia and West Virginia experienced more notable decreases. Most hair braiding salons are smaller beauty shops; thus, we may not expect to observe much difference with respect to the number of beauty shops with employees following the deregulation of hair braiding. Turning to Figure 3A, the number of beauty shop employees (per 1,000 residents) falls slightly in Virginia and West Virginia following deregulation and stays mostly consistent in North Carolina and Kentucky. In percentage terms, the decline in the number of beauty shop employees is 3.19 percent from 2012 to 2014, while the number of employee establishments increases by a modest 0.04 percent. Many hair braiding establishments are small and may not have employees, so perhaps it is not surprising to see little evidence of a positive trend in the number of employees following deregulation. In fact, the drop in the number of employees in Virginia may be a result of an overall decrease in the size of beauty shops following deregulation. In Figure 3B, we plot trends in the number of sole proprietors/owners of beauty shops (per 1,000 residents) in each state. Once again, these data are obtained from the NES (as opposed to the CBP). Virginia, Kentucky, and North Carolina experience increases in the number of beauty shop sole proprietors, and in West Virginia the number remains relatively flat. In percentage terms, the number of beauty shop proprietors increased by 8.2 percent from 2012 to 2014 in Virginia.
Figure 4 displays raw trends in the number of employees per beauty shop in each state. Three of the four states (all but West Virginia) experience a decrease in the number of employees per shop before Virginia’s deregulation in 2012. Following deregulation, the number of employees per shop falls sharply in West Virginia (by almost a full employee from 2012 to 2014) and increases modestly in both North Carolina and Kentucky. In Virginia, the primary state of interest, the number of employees per beauty shop declines modestly following deregulation in 2012 after several years of staying relatively constant. This trend seems to support increased economic opportunities for hair braiders, as new salon proprietors may have once been salon employees. Hair braiding establishments are generally smaller than are salons offering other cosmetic services; thus, the drop in the number of employees per shop in Virginia may reflect either an increase in the number of new small beauty shops with few employees or an increase in the number of owner-operated salons (with no employees) opening in the state following deregulation.

FIGURE 2
Number of Beauty Shops with Employees per 10,000 Residents (2004–14)

FIGURE 3A
Number of Beauty Salon Employees per 1,000 Residents (2004–14)

![Graph showing the number of beauty salon employees per 1,000 residents from 2004 to 2014.](image)

Source: County Business Patterns data, 2004–14.

FIGURE 3B
Number of Beauty Salon Sole Proprietors/Owner Operators per 1,000 Residents (2004–14)

![Graph showing the number of beauty salon sole proprietors/owner operators per 1,000 residents from 2004 to 2014.](image)

FIGURE 4
NUMBER OF EMPLOYEES PER BEAUTY SALON (2004–14)

WORKERS PER SALON


Kentucky North Carolina Virginia West Virginia


FIGURE 5
AVERAGE ANNUAL EARNINGS OF BEAUTY SALON EMPLOYEES (2004–14)

REAL 2009 DOLLARS


Kentucky North Carolina Virginia West Virginia

In Figure 6A, we plot the distribution of beauty shops in the state of Virginia. More than 65 percent of counties in the state have fewer than 13 beauty shops with employees. And, as shown in Figure 6B, most Virginia counties (a few more than half) have a relatively small number of beauty shops (fewer than 35) without employees (owner-operated shops). Although a state-level comparison is interesting, a county-level analysis allows for an opportunity to better isolate the effects of deregulation by focusing on geographic areas that may be economically similar despite being in different states—and that are thus subject to different regulation. We turn to a county-level analysis in the next section.

County-Level Analysis

To further investigate the effect of Virginia’s deregulation of hair braiders, we present six simple tests, three testing employee establishments and three testing nonemployee (proprietor) establishments. To facilitate interpretation, we examine the number of proprietorships per 1,000 people and employee establishments per 10,000 people. No Virginia counties with reporting beauty shops are excluded in the regressions—the total sample includes 147 counties from 2005 to 2014. The equation that we estimate is:

\[
\ln(\text{Beauty Shop Density}) = \alpha + \beta_1 (\text{DID}) + \beta_2 (\text{county control variables}) + \beta_3 (\text{county fixed effects}) + \epsilon
\]

We control for county unemployment rate and real personal income per capita to account for possible county-level differences in demand for hair braiding services. In addition, all regressions include county fixed effects and standard errors that are adjusted for clustering at the county level.

We take the natural log of each beauty shop density (the number of proprietorships and employee establishments per capita) to facilitate interpretation of each coefficient as a rate of change. A summary of each test is presented below:

- Test 1: Proprietorships in Virginia border counties compared with Virginia inner counties
- Test 2: Proprietorships in Virginia border counties compared with out-of-state border counties
FIGURE 6A
DISTRIBUTION OF VIRGINIA COUNTIES BY NUMBER OF BEAUTY SALONS WITH EMPLOYEES


FIGURE 6B
DISTRIBUTION OF VIRGINIA COUNTIES BY NUMBER OF BEAUTY SALONS WITHOUT EMPLOYEES (OWNER OPERATED)

Test 3: Proprietorships in Virginia counties compared with out-of-state border counties
Test 4: Employee establishments in Virginia border counties compared with Virginia inner counties
Test 5: Employee establishments in Virginia border counties compared with out-of-state border counties
Test 6: Employee establishments in Virginia counties compared with out-of-state border counties

Our primary variable of interest in each regression is a simple interaction term of dummy variables. In tests (columns) 1 and 4 of Table 3, we interact a dummy variable for deregulation of hair braiding (equal to 1 after 2012) with a dummy variable denoting Virginia border counties. The resulting variable is labeled “Dereg|Border.”

In addition, the hair braiding deregulation and Virginia border county dummy variables are both included separately as additional independent variables—fully specifying the difference-in-difference (DID) coefficients. Both interaction term coefficients are negative (the coefficient on proprietors is statistically significant), suggesting that deregulation resulted in more growth in the number of proprietor beauty shops within the inner counties of Virginia than in the border counties. Most important, it does not appear that the number of beauty shop establishments grew more quickly in Virginia border counties relative to inner counties.

For the remainder of the tests (columns 2, 3, 5, and 6), we compare Virginia counties to contiguous counties in Kentucky, North Carolina, and West Virginia. Our main variable of interest is also a simple interaction of dummy variables—in this case, a dummy variable denoting a Virginia county interacted with the same dummy variable from tests 1 and 4 denoting the period of hair braiding deregulation. This variable is labeled “Dereg|VA.” As in the previous tests, we also include each dummy variable from the interaction term separately. Tests 2 and 5 focus exclusively on Virginia border and contiguous out-of-state counties (resulting in the noticeably smaller sample size). Although we do estimate that Virginia border counties had a higher rate of growth in the number of beauty shops (particularly for the employee shops in test 5), neither of the coefficients is statistically significant. Considering our results from tests 1 and 4, we reran this test including all Virginia (border and inner) counties. The results from this final test are

The full results of the regressions are available upon request.
### TABLE 3

**Simple Tests of Virginia Border and Contiguous Counties**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dereg</td>
<td>Border</td>
<td>-0.0838***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0227)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dereg</td>
<td>VA</td>
<td>0.00296</td>
<td>0.0704***</td>
<td>0.0512</td>
<td>0.0704**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0247)</td>
<td>(0.0190)</td>
<td></td>
<td>(0.0429)</td>
<td>(0.0320)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,207</td>
<td>438</td>
<td>1,425</td>
<td>1,210</td>
<td>423</td>
<td>1,421</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.924</td>
<td>0.888</td>
<td>0.925</td>
<td>0.934</td>
<td>0.917</td>
<td>0.939</td>
</tr>
</tbody>
</table>

Notes: * p < 0.10; ** p < 0.05; p < 0.01. Standard errors are in parentheses.
reported in columns 3 and 6. In this case, we consistently estimate that Virginia counties had more growth in the number of beauty shops (approximately 7 percent) than did contiguous counties in bordering states after deregulation. This result is very similar to the difference we noted in our state-level comparison in the preceding section.

Summary and Limitations

In the preceding sections, we find that deregulation has expanded opportunity for smaller beauty salons (measured by as much as an 8 percent growth rate in owner-operated salons after deregulation). A county-level regression analysis produces similar results that the number of beauty shops grew approximately 7 percent faster in Virginia relative to contiguous counties in bordering states. In short, the evidence presented seems to support the hypothesis that deregulation of hair braiding has resulted in more opportunity for hair braiders in Virginia relative to bordering states. We should note that there are limitations to our analysis. First, there are several cases of missing data that make it difficult to present a more complete comparison. The CBP publication often does not report data because either the data are redacted to protect the privacy of businesses or the data have not met certain standards. Such reporting irregularities result in many gaps in the data, with some counties having to be excluded from the model because too many years of data were missing. The NES is more robust with respect to the number of observations but includes only counts of establishments.

Another limitation of our analysis is that the beauty shop category is broad. Ideally, we would have specific data on hair braiding salons. Unfortunately, to the best of our knowledge, such data are not available. Finally, having data that end in 2014 limits our ability to identify longer-term effects of deregulation. It is possible that there is a lag in seeing the effects of deregulation. Hair braiders who offer their services unofficially in the underground economy may continue to not report their activity, which may limit our ability to isolate the effects of deregulation.

Conclusion

In this article, we have estimated the effects of Virginia’s deregulation of the hair braiding profession in 2012. Our results suggest
that beauty shops in Virginia at the county level experienced approximately 7 percent higher growth rates than did those in contiguous counties in bordering states. State-level analysis confirms this result and is also supportive of the view that deregulation has expanded opportunity for smaller beauty salons (measured by as much as an 8 percent growth rate in owner-operated salons after deregulation).

As policymakers reconsider regulation of hair braiding, our results should provide very clear guidelines. Having no regulation of the profession seems to provide more opportunities for business development than burdensome regulation (as in Kentucky and West Virginia, where hair braiders were previously required to obtain a cosmetology license) and slightly less burdensome regulation (as in North Carolina, where hair braiders and cosmetologists obtain separate licenses) improves business outcomes. The past several years have proved to be fruitful, with many states choosing to deregulate the hair braiding profession. Nevertheless, 13 states continue to require hair braiders to obtain a cosmetology license—a process that seems unnecessarily onerous to aspiring practitioners and may not be generating benefits for consumers.

References


Hair Braider Deregulation


Virginia Governor’s Executive Reorganization Plan, Section 19 (2012) House Joint Resolution No. 49 (March 10).


Restoring the Rule of Law in Financial Regulation

Charles W. Calomiris

The financial crisis of 2007–08 ushered in the most sweeping changes in financial regulations since the Great Depression. Unlike the reforms wrought in 1932–35, which remained in place for decades, much of the post–2008 legislation is already a likely target for repeal or at least significant modification.

Critics point to many shortcomings. Some focus on the costs of regulation, arguing that regulatory reform has benefited large Wall Street banks by codifying their status—as “too big to fail”—and by creating new regulatory costs that big banks can bear more easily than competitors. Small banks face a morass of new rules and compliance burdens. Lux and Greene (2015) find that the “increasingly complex and uncoordinated regulatory system has created an uneven regulatory playing field that is accelerating consolidation for the wrong reasons,” producing a declining market share for community banks. These various costs are being passed on to bank customers. Many Americans are finding it increasingly difficult to access banking services on favorable terms. For example, the share of banks offering free checking accounts fell from 75 percent prior to Dodd-Frank to 37 percent in 2015. Monthly service fees charged by banks have grown 111 percent over the same time, while the

Cato Journal, Vol. 38, No. 3 (Fall 2018). Copyright © Cato Institute. All rights reserved.

Charles W. Calomiris is the Henry Kaufman Professor of Financial Institutions and Director of the Program for Financial Studies at Columbia Business School. He is also a Research Associate of the National Bureau of Economic Research, a Fellow of the Manhattan Institute, and Co-Director of the Hoover Institution Program on Regulation and the Rule of Law.
number of “unbanked” Americans has grown. Credit card interest rates are 2 percent higher, and the number of credit card accounts has fallen by 15 percent. A Goldman Sachs Global Markets Institute (2015) study on the consequences of financial regulation for small businesses also found major costs: “The tax from increased bank regulation falls disproportionately on the smaller businesses that have few alternative sources of finance. We see this in the muted recovery in bank lending to small businesses: outstanding commercial and industrial (C&I) loans for less than $1 million are still well below the peak 2008 level and are only 10 percent above the trough seen in 2012.”

Other critics have focused on the current or prospective failings of regulation to achieve its desired prudential objectives. The continued reliance by capital regulation on book values of tangible net worth as a measure of loss absorbing capacity is one obvious weakness. That approach is not likely to work better in the future to prevent too-big-to-fail banks from failing because it does not reliably track the true economic value of bank equity. Risk measurement under the Basel approach employed in the United States and many other countries notoriously creates opportunities for circumvention through the understatement of risk. New bank liquidity requirements are extremely complex, easy to circumvent, and lacking in any fundamental grounding in economic theory. Title II of Dodd-Frank is viewed by many academic critics as unworkable and unlikely to produce orderly resolution of nonbank institutions or large bank holding companies.

In this article, I focus on another, even more fundamental, problem. Increasingly, our regulatory structure has been adopting processes that are inconsistent with adherence to the rule of law. These process concerns are rarely voiced by academics, but that is a strange omission. Appropriate regulatory process is fundamental to the ability of regulation to succeed because process defines the incentives of regulators, which are crucial to ensure that regulators act diligently in pursuit of bona fide objectives. Relying on regulatory

---

1See Calomiris (2017a) for a detailed review of each of these issues and proposals for regulatory standards that would be more likely to succeed.

2There are many definitions of “rule of law.” I use the term to refer to the predictable and nondiscriminatory enforcement of laws and regulations.
processes that avoid transparency, accountability, and predictability increases regulatory risk and is likely to lead to poor execution of regu-
larly obligations, as well as create unnecessary regulatory costs and opportunities for politicized mischief. This is not merely a theo-
retical concern. As I will show, because recent regulation has increased regulators' discretionary authority, and has reduced the predictability and transparency of regulatory standards, it has reduced the accountability of regulators. This has already resulted in abuses that not only deform our democracy but also impose unwarranted costs on the financial system and distract from legitimate problems that should be the focus of prudential and consumer pro-
tection regulation.

The Demise of the Rule of Law in Financial Regulation

**CFPB Structure, Process, and Policies**

Barney Frank has said that he regards the creation of the Consumer Financial Protection Bureau (CFPB) as the greatest achievement of the Dodd-Frank Act. But the CFPB has been a lightning rod for controversy, both about its policies and with respect to its structure and process. With respect to its structure and process, the CFPB was given a unique position within the govern-
ment. Its budget is determined without the possibility of congres-
sional limitation (its expenses are assessed against the Federal Reserve System, prior to the Fed rebating its surplus to the Treasury), its mandate is extremely broad, and unlike similar regulatory authorities (such as the Securities and Exchange Commission [SEC]), it is run by an individual director rather than a bipartisan panel. In October 2016, a three judge panel of the U.S. Court of Appeals for the District of Columbia found not only that the CFPB was incorrect in its interpretation of a law that it used to justify the imposition of a $109 million penalty, but also that the CFPB “violated bedrock due process principles” and that its structure was unconstitutional because Congress gave the CFPB “more unilateral authority than any other officer in any of the three branches of the

---

U.S. government, other than the president” and that consequently the CFPB “possesses enormous power over American business, American consumers and the overall U.S. economy” (U.S. Circuit Court for the District of Columbia 2016). The full appellate court overturned that ruling in January 2018, but the conflict over the CFPB’s structure continues.

With respect to its policies, the CFPB has been aggressive in promoting unprecedented interpretations of consumer protection regulation. Perhaps its most controversial decision was the use of “disparate impact” theory to gauge discrimination against minorities. According to this theory, if one group of people (identified on the basis of racial or ethnic identity) receives different average outcomes (different approval/denial rates or different terms for lending), even in the absence of any evidence of differences in treatment by a lender on the basis of race or ethnicity, then that disparate impact constitutes evidence of illegal discrimination. Furthermore, the CFPB’s (2014) information about race and ethnicity was derived not from actual knowledge of individuals’ race and ethnicity, but rather from “a Bayesian Improved Surname Geocoding (BISG) proxy method, which combines geography- and surname-based information into a single proxy probability for race and ethnicity.” In other words, discrimination was punished based on forecasted probabilistic racial or ethnic identities, not actual ones.

The report of a congressional investigation into CFPB’s practices by the U.S. House Committee on Financial Services (2015) found that the CFPB had knowingly failed to control for influences other than discrimination that cause differences in outcomes, and that its actions were inconsistent with congressional intent in creating the CFPB, with the law (which specifically exempted certain automobile financing from CFPB authority), and with Supreme Court definitions of what constitutes discrimination. Racial and ethnic forecasting was also unreliable. The Executive Summary of the report is a scathing indictment of CFPB practices:

Since at least February 2012, the Bureau of Consumer Financial Protection (Bureau), and in particular its Office of Fair Lending and Equal Opportunity, has engaged in an aggressive effort to enforce the Equal Credit Opportunity Act (ECOA) against vehicle finance companies using a controversial theory of liability known as disparate impact. In doing so, it has attempted to implement a “global solution” that enlists
these companies in an effort to alter the compensation of automobile dealers, over which the Bureau has no legal authority. As internal documents obtained by the Financial Services Committee and accompanying this report reveal, the Bureau’s ECOA enforcement actions have been misguided and deceptive. The Bureau ignores, for instance, the lack of congressional intent to provide for disparate impact liability under ECOA, just as it ignores the fact that indirect auto finance companies are not always subject to ECOA and have a strong business justification defense. In addition, memoranda reveal that senior Bureau officials understood and advised Director Richard Cordray on the weakness of their legal theory, including: (1) that the practice the Bureau publicly maintained caused discrimination—allowing auto dealers to charge retail interest rates to customers—may not even be recognized as actionable by the Supreme Court; (2) that it knew that the controversial statistical method the Bureau employed to measure racial disparities is less accurate than other available methods and prone to significant error, including that for every 100 African-American applicants in a data set for which race was known, the Bureau’s proxy method could only identify roughly 19 of them as African-Americans; and (3) that the Bureau knew that factors other than discrimination were causing the racial disparities it observed, but refused to control for such factors in its statistical analysis. Notwithstanding the weakness of its case, the Bureau pursued its radical enforcement strategy using “unfair, abusive, and deceptive” tactics of its own, including by making an example of a company over which it had significant political leverage and concealing other aspects of its efforts from public scrutiny. The purpose of this report is to provide the public with a better understanding of the Bureau’s activities.

Thus, the CFPB appears to have created and enforced a new theory of discrimination, one that was inconsistent with economic evidence about the causes of disparate impact, and one that lawmakers characterize as contrary to statutory language and Supreme Court opinions about what constituted illegal discrimination. Clearly, there is a connection between the unconstitutional structure and process that created the CFPB, which insulated its imperious director from any budgetary or administrative discipline, and its abuse of power. The broad lesson—which applies to the regulatory reliance
on guidance in general—is that financial regulatory power is easily politicized and abused when it is not required to adhere to statutory authority, or at least to a formal rule making process.

Operation Choke Point

Imagine that you are operating a business and you get a call from your banker explaining that she can no longer provide services to you. Your accounts at the bank must be closed immediately, despite the fact that your business is thriving and you have done nothing unlawful. When you call another banker to try to open an account, he turns you down, too. The bankers all tell you the same story: bank regulators have told them that they should not serve you, and they must obey or will face significant regulatory penalties. Welcome to the Obama Administration’s main post-Dodd-Frank contribution to financial regulation, known as “Operation Choke Point.”

Alongside a litigation initiative that began in the Justice Department, in 2011, the Federal Deposit Insurance Corporation (FDIC) and other bank regulators warned banks of heightened risks from doing business with certain merchants. Purveyors of “pornography” or “racist materials” may enjoy First Amendment rights, but not the right to a bank account. Gun and ammunition dealers were targeted despite their Second Amendment rights. Firms selling tobacco or lottery tickets were persona non grata, too. Payday lenders also were targeted on the basis of the presumption that they prey on the poor. A total of 30 undesirable merchant categories were deemed to be “high-risk” activities. In 2012, the FDIC explained that having the wrong kinds of “risky” clients can produce “unsatisfactory Community Reinvestment Act ratings, compliance rating downgrades, restitution to consumers, and the pursuit of civil money penalties.” Other regulatory guidelines pointed to difficulties banks with high “reputation risk” could have consummating acquisitions (Calomiris 2017b).

A report by the House Committee on Oversight and Government Reform (2014) unearthed internal FDIC emails voicing intent to “take action against banks that facilitate payday lending” and “find a way to stop our banks from facilitating payday lending,” which highlighted the FDIC’s use of Memoranda of Understanding with banks and Consent Orders to implement its campaign against payday lending. The report concluded that “senior policymakers in FDIC headquarters oppose payday lending on personal grounds,” and that FDIC’s campaign against payday lenders reflected “emotional
intensity” and “personal moral judgments” rather than legitimate safety and soundness concerns, and was “entirely outside of FDIC’s mandate.” The Inspector General of the FDIC (2015) issued a report substantiating those judgments. It found that FDIC staff had been working with the Department of Justice to identify banks’ relationships with payday lenders, which (contrary to the FDIC’s financial interests and duties) served to make litigation risk from the Justice Department greater for banks with payday lending relationships. Operation Choke Point is not grounded in bank regulators’ expertise or mission, just their willingness to harass bank clients whose activities they dislike.

Some observers may agree with Mr. Obama’s list of disfavored industries. But now that Mr. Trump has taken office, will they agree with his list? Do we want our regulatory system to be a tool for attacking those our president dislikes? If not, it’s worth asking why the political abuse of regulation has become easier than in the past, and what can be done to stop it.

There was never legislation defining the 30 industries as undesirable, nor did regulators establish rules to set clear standards for what constituted undesirable behavior by a bank’s client, or announce penalties for banks serving undesirables. Such legislation or formal rule making likely would have been defeated owing to the checks and balances inherent in congressional debate or formal rule making under the Administrative Procedures Act. Instead, regulators relied on guidance—which requires no rule making, solicits no comments, entails no hearings, avoids defining violations, specifies no procedures for ascertaining violations, and defines no penalties that will be applied for failure to heed the guidance.

Communications between regulators and banks are private; banks often aren’t permitted to share them with outsiders. Regulators avoid public statements explicitly requiring banks to terminate undesirables, but privately threaten banks with an array of instruments of torture that would have made Galileo faint, using secrecy to avoid accountability.

As DeMuth (2014), Epstein (2014), Hamburger (2014), and Baude (2016) have documented, and as financial regulatory practice illustrates, there has been a dramatic increase in reliance on guidance in recent years. Financial regulators can find it particularly useful to rely on vaguely worded guidance and the veil of secrecy to maximize
discretionary power, although doing so imposes unpredictable and discriminatory costs on banks and their customers.

The regulators’ campaign against payday lenders produced a wave of bank relationship terminations, with dire consequences for the industry since 2013. That not only victimized payday lenders, it also imposed significant costs on consumers by reducing competition. A large, and very one-sided, academic literature convincingly shows that payday lenders serve customers’ interests and perform competitively (see Calomiris 2017b: Appendix A). Their presence reduces borrowing costs for customers. If the prejudiced views of bureaucrats about payday lending had been held up to scrutiny during public hearings, their jaundiced portrayals of the industry would have been disproven. But using guidance avoids having to defend one’s prejudices in public. Once the government and its regulators decided to strip the industry of its ability to transact with banks, the view that payday lenders were “risky” became self-fulfilling.

Payday lenders are now suing bank regulators for the harm they have suffered (a lawsuit in which I have filed a report on Plaintiffs’ behalf—Calomiris 2017b). In that lawsuit there is more at stake than the fate of payday lenders or their customers. Regulators’ reliance on vague guidance and discretionary judgments about ill-defined violations under a veil of secrecy constitute a major departure from the rule of law, with far-ranging adverse consequences for our economy, our political institutions and our society.

Operation Choke Point has been discontinued. In an August 16, 2017, letter by Assistant Attorney General Stephen Boyd to Senators Tillis and Crapo, Secretary Boyd writes:

We share your view that law abiding businesses should not be targeted simply for operating in an industry that a particular administration might disfavor. Enforcement decisions should always be made based on the facts and the applicable law.

. . . The [2013] FDIC guidance included a footnote listing certain “elevated-risk” merchants, including short-term lenders and firearms dealers. The FDIC subsequently rescinded its list . . . The Department of Justice (Department) strongly agrees with this withdrawal. All of the Department’s
bank investigations conducted as part of Operation Chokepoint are now over, the initiative is no longer in effect, and it will not be undertaken again.

. . . We reiterate that the Department will not discourage the provision of financial services to lawful industries, including businesses engaged in short-term lending and firearms-related activities [Boyd 2017].

Despite that assurance, another administration might very well decide to encourage similar regulatory abuses of power in the future. Congress should prevent that from being possible. Or, if the Department of Justice believes that such actions were already illegal, perhaps it will consider criminal legal action against its own former officials responsible for this disgrace.

Financial Stability Oversight Council

Dodd-Frank created a new macroprudential mandate for the newly established Financial Stability Oversight Council (FSOC) and Office of Financial Research (OFR). The OFR is supposed to identify potential systemic risks, using its unprecedented access to the proprietary data of financial regulators and financial institutions, and inform the FSOC of risks. The FSOC, chaired by the Secretary of the Treasury, has a statutory duty to facilitate information sharing and regulatory coordination by the various financial regulators. It is charged to respond to systemic risks by recommending appropriate strengthening in regulatory standards, and designating, as appropriate, certain financial market utilities and nonbank financial institutions (or other firms) as systemically important (and therefore subject to new regulations). It is also empowered to break up any firms in the United States that it deems to be a “grave threat” to systemic stability.

Critics of FSOC (and OFR) have pointed to two primary problems in its structure and operation: procedural shortcomings and politicization. The two problems are closely related. With respect to procedural shortcomings, at least one SEC commissioner—Michael Piwowar—has complained publicly about being shut out of FSOC deliberations. Piwowar (2014) notes many concerns about FSOC, which he labels, among other things, the “Unaccountable Capital Markets Death Panel.” Piwowar has
identified an important problem. Not only is FSOC unaccountable, it also is composed of the heads of the various financial regulatory agencies, all of whom are appointed by the same political party. The diversity that is either required by statute, or is a function of staggered appointments over time, for the SEC, the Commodity Futures Trading Commission (CFTC), the Federal Reserve Board (FRB), the FDIC, and others does not apply to the FSOC’s deliberations, which also remain largely secret.

Even worse, the FSOC has not established standards to guide its designations of firms as systemically risky or “grave threats.” Creating the power to regulate anyone in the U.S. economy, and to shut down any business operating in the U.S. economy, is worrying enough, but when that authority can be exercised by members of one political party, acting in secret without any specified standards to guide them, it must be considered outside the realm of actions that should occur in a democracy governed by the rule of law.

On December 18, 2014, Metlife was notified by FSOC that it had been designated a nonbank systemically important financial institution (SIFI), which implied new regulatory burdens and risks. Metlife challenged the FSOC’s decision in federal court and on March 30, 2016, U.S. District Court Judge Rosemary Collyer ruled in Metlife’s favor and rescinded its SIFI designation. The Judge’s opinion is interesting because it pointed to the shortcomings of FSOC’s procedures as central to the case. In doing so, the Judge’s opinion opened a broader debate about the abuse of guidance by regulators.

Judge Collyer’s opinion was, therefore, especially noteworthy as one of the first attempts by a federal court to disallow the unlimited use of regulatory discretion in the administration of regulatory guidance. The judge did not disallow guidance, per se, but she rejected unlimited and inconsistent discretion in its use as a regulatory tool, “[h]aving found fundamental violations of established administrative law”:

During the designation process, two of FSOC’s definitions were ignored or, at least, abandoned. Although an agency can change its statutory interpretation when it explains why, FSOC insists that it changed nothing. But clearly it did so. FSOC reversed itself on whether MetLife’s vulnerability to financial distress would be considered and on what it means to threaten the financial stability of the United States. FSOC
Rule of Law

also focused exclusively on the presumed benefits of its designation and ignored the attendant costs, which is itself unreasonable under the teachings of Michigan v. Environmental Protection Agency, 135 S. Ct. 2699 (2015). While MetLife advances many other arguments against its designation, FSOC’s unacknowledged departure from its guidance and express refusal to consider cost require the Court to rescind the Final Determination.4

In addition to the potential for abusive actions, there is also reason to be concerned about FSOC inaction. Strangely, FSOC and OFR have been largely silent about the mounting systemic risks in U.S. real estate, which many observers believe may be substantially over-priced. Indeed, it is not an exaggeration to say that FSOC seems to be uninterested in the only obvious and legitimate systemic risk facing the U.S. economy today.

The unprecedented pandemic of financial system collapses over the last four decades around the world is largely a story of real estate booms and busts (Jordà at al. 2015; Calomiris 2018). Real estate is central to systemic risk in many countries because of four facts. First, exposures to real estate risk inherently are highly correlated with each other and with the business cycle, which means that downturns in real estate markets can have large and sudden implications for massive amounts of loans and securities backed by real estate.

Second, real estate assets are unique and generally cannot be liquidated quickly at their full long-term value, which can imply large losses to holders who are forced to sell real estate quickly. Those losses can further exacerbate financial losses and magnify systemic risk.

Third, over the past 40 years worldwide, and especially in the United States, real estate is increasingly funded by government-protected and government-regulated entities. That protection encourages the politicization of real estate funding (given the strong short-term political incentives to subsidize mortgage risk).

Fourth, throughout the world, a large amount of commercial and/or residential real estate investment is being funded increasingly

within banks, which rely primarily on short-term debt for their funding. As we witnessed during the Subprime Crisis in the United States, real estate losses produced substantial liquidity risk (beginning in August 2007 in the asset-backed commercial paper market, and continuing through September 2008 in the repo and interbank deposits markets), which deepened the losses during the crisis and magnified the general contraction in credit that ensued. But this is not just a problem of large banks. The loan portfolios of small banks in the United States are also highly exposed to residential and commercial real estate risk, which over the past two decades averaged about three-quarters of total lending by small banks.

Many observers see large banks as the only source of systemic risk in the economy, but that mistaken view forgets that the United States has been the most financially unstable developed economy in the world for more than a century, despite the fact that large banks are a recent development in the United States (Calomiris and Haber 2014: chaps. 6–7). The 1980s banking crises were all about real estate losses incurred by small banks—not just in housing, but also in commercial real estate, especially in the southwest and the northeast, and in agricultural real estate throughout the country.

It is not hard to see why FSOC has been silent about the excessive exposure to real estate in the banking system, the increased risk taking by the GSEs and the FHA, the failure to reform the GSEs, and the increasing riskiness of mortgages over the past three years. Any discussion about these important systemic risks would be politically inconvenient.

For example, no one expected Jacob Lew, the Treasury secretary in the same administration that appointed Mel Watt to head the Federal Housing Finance Agency (FHFA), to criticize Mr. Watt’s decisions to increase mortgage risk after his appointment as head of FHFA. Immediately upon assuming authority, Mr. Watt reduced the down payment limit on GSE-eligible mortgages from 5 percent to 3 percent. The GSEs remain in conservatorship, and the combination of the watering down of the “Qualified Mortgage” (QM) and “Qualified Residential Mortgage” (QRM) rules and exemptions (Gordon and Rosenthal 2016), alongside these lax FHFA standards, and the government’s funding of the GSEs and the FHA and VA, continue to ensure that government subsidization of housing finance risk—the central problem highlighted by the 2007–08 crisis—will continue.
The continuation of the government’s push for risky housing finance has resulted in a continuing escalation of mortgage risk. For first-time buyers, combined loan-to-value ratios have risen from 90.7 percent in February 2013 to 91.9 percent in January 2017, and the average debt service-to-income ratios rose over that period from an average of 36.4 percent to 37.7 percent. As of the end of January 2017, 28 percent of first-time buyers had debt service-to-income ratios in excess of the QM limit of 43 percent, which is four percentage points higher than it was two years earlier. Fannie Mae, Freddie Mac, FHA, and VA hold riskier mortgage portfolios than banks, and they account for about 96 percent of purchased mortgage volume.

Given the political push for providing financing subsidies in the form of government-sponsored encouragement of systemic mortgage risk, the FSOC prefers to focus its attention on “interconnectedness” when discussing systemic risk, rather than recognize the central importance of real estate finance in producing systemic shocks (Scott 2016, Calomiris 2018).

Nor does the FSOC care to focus on the potential for small banks’ funding of commercial and residential real estate to create systemic risk. Small banks are politically popular in Congress (which rightly worries that regulatory burdens are putting many of them out of business). Builders and realtors also are popular with both political parties (Calomiris and Haber 2014: chaps. 7–8; Gordon and Rosenthal 2016) and no one is going to point toward small banks’ outsized exposures to real estate, or any other exposures to real estate, as a problem. When I did so in congressional testimony (Calomiris 2015), I was attacked from both sides of the aisle for failing to appreciate the importance of the “American dream.” Of course, mortgage subsidies have little effect on housing affordability (currently at a long-term low in the United States) because they not only expand credit but also prop up home prices, but honesty about housing markets has always been in short supply in Washington.

When I talk to OFR economists about the need to focus attention on real estate risks, it often seems that people start looking at their shoes. I have found the economists at the OFR to be skilled and diligent, and I find much of the OFR’s research output quite useful,

5The facts noted in this paragraph are taken from Pinto and Peter’s (2017) Power Point presentation.
but they have a blindspot when it comes to the risk-creating policies of the Administration.

**Stress Tests**

As part of the resolution of the financial crisis, U.S. banks were stress tested by the Federal Reserve in 2009 (the Supervisory Capital Assessment Program). Beginning in 2011, stress tests became a regular feature of the regulatory apparatus and, beginning in 2014, stress tests were required as part of Dodd-Frank for all banks with more than $10 billion in assets.

In concert with reformed capital ratios, stress tests could be an effective means of encouraging bankers to think ahead—leading them to consider risks that could cause sudden losses of value, and prodding them to increase capital buffers and improve their risk management practices. As they are currently structured, however, stress tests violate basic principles of the rule of law that all regulations should adhere to. Banks that fail stress tests are punished for falling short of standards that are never stated (either in advance or after the fact). This makes stress tests a source of uncertainty rather than a helpful guide against unanticipated risks.

Fed officials have justified the lack of transparency and accountability in stress testing because of the need to ensure that banks do not game the test, but that is not a reasonable argument. Changing economic circumstances imply that every year the scenarios that are relevant for stress testing should change, and therefore scenario modeling should not be highly predictable on the basis of past years’ tests. Ex post disclosure of the tests, combined with learning over time, changes in scenarios that track changing market circumstances, the use of multiple scenarios designed by multiple teams of experts, and rotation of the people designing scenarios should provide more than adequate unpredictability about the specifics of any test to prevent gaming of the test by bankers. Keeping the details of the methodology of stress testing a permanent secret has very undesirable features: it makes it impossible for market participants to learn what regulators regard as appropriate modeling techniques and assumptions, thereby insulating the regulators from any accountability for poor test design.

Regulators not only impose unstated quantitative standards for meeting stressed scenarios, but they also retain the option of simply deciding that banks fail on the basis of a qualitative judgment
unrelated even to their own secret model’s criteria. It is hard to believe that the current structure of stress tests could occur in a country like the United States, which prizes the rule of law, the protection of property rights, and adherence to due process.

Process Reform

How can we reform financial regulatory process to restore adherence to the rule of law? Some solutions are simple and obvious: Operation Choke Point is a sad episode in our nation’s history and its revival should be prevented by legislation. But, as the above examples show, there is a more general problem illustrated by Operation Choke Point, one that arises because of the increasing reliance on guidance, which allows regulators to avoid accountability for their actions.

I favor requiring regulators to rely on formal rule making rather than guidance, so that regulation can be based on clearly defined standards, debated in public, and enforced transparently. Over a short period of time, I propose that existing guidance should be phased out entirely and replaced by formal rules. This will be a massive undertaking, and I do not recommend it lightly. But it is crucial for restoring the rule of law to our financial system. Eliminating the reliance on guidance also will reduce regulatory risk substantially, with favorable consequences for both growth and stability.

In addition to this overarching reform, below I propose specific reforms that address the specific problems outlined above in the FSOC, the CFPB, and stress testing.

FSOC Reform

FSOC should be made as politically independent as possible, especially because it (and the OFR that advises it) should retain necessary access to privileged data. FSOC’s mission should be identifying problems related to systemic risk, especially potential shortcomings in the enforcement of regulatory standards.

Barth, Caprio and Levine’s (2012) proposal for a “Sentinel” is a potential model. In their formulation, this body would be administered independently but have access to privileged data, including information about the actions of regulators and supervisors. To accomplish that mission, the FSOC would have to be removed from the Treasury and established on other, independent footings. It may still make sense to have the FSOC meet with regulators (such as Fed governors, SEC
commissioners, and FDIC officials) but to be able to oversee the actions of those parties effectively it must be separate from them.

The designation of SIFI status, like other regulatory designations, should follow from clear rules, not opaque discretionary judgments that invite the abuse of power. For example, in addition to size thresholds (which measure an institution’s systemic importance), the degree of a nonbank institution’s reliance on short-term debt, and the degree to which it uses short-term debt to fund illiquid investments, such as commercial real estate loans, could be taken into account explicitly (and quantitatively) when formulating a rule for what constitutes systemic importance.

**CFPB Reform**

The CFPB could play a constructive role in monitoring compliance with consumer protection laws, such as disclosure requirements, mortgage brokerage standards, fair lending requirements, and anti-discrimination statutes. It should focus on monitoring and enforcing compliance of the laws that exist (e.g., by using testers to root out discriminatory treatment of consumers), advise Congress on the creation of new laws, and engage in formal rule making that is consistent with specific powers delegated to it by Congress. These functions are analogous in the banking sphere to some of the activities of the SEC, and it seems natural for the CFPB to adopt a similar bipartisan commission structure, which will help to insulate it from counterproductive political pressures. In keeping with this new structure, the CFPB’s activities should no longer be funded by special access to Federal Reserve surplus.

**Reforming Fed Stress Tests**

Two important sets of reforms are needed to address the current deficiencies in the process governing stress tests. First, the criteria for stress tests must be clarified, and the stress tester (the Fed) must be made accountable for its approach to measuring compliance. The Fed does not need to predisclose the specific models it will use, but it does need to explain, and demonstrate that it is adhering to, a reasonable and transparent process to build the models that will be used to measure compliance. And it must disclose the models it employs with a lag, to ensure accountability for the quality of its testing standards.

One practice that would improve accountability and reduce the ability of banks to game the test would be for the Fed to invite
independent teams to assist it in building models (perhaps using multiple models rather than one). The Fed could rotate its model-building personnel and alter its scenarios in light of changing economic circumstances, which would ensure that its models conform to best practice while also remaining somewhat unpredictable. Each year, the Fed could disclose the models that were used previously, which would ensure accountability by permitting detailed criticisms by academic and industry observers.

Improved regulatory accountability likely would produce improvements in stress testing. Currently, there is great room for improvement. Stress tests should focus realistically on the true loss of economic value under various forward looking scenarios, based on defensible cash flow forecasts, not just tangible asset loss projections or forecasts of broad financial accounting measures. To accomplish that objective, bank cash flows must be analyzed properly. Managerial accounts of revenues and expenses should be separated by line of business, and cash flow projections for each line of business under each scenario should be justified by reference to observable historical patterns. For example, under a scenario of severe housing finance decline, mortgage servicing income is likely to be more affected than asset management fees. To be able to realistically capture the effects of macroeconomic scenarios on bank condition the data used in stress testing must be improved dramatically.

Conclusion

Some readers will regard the proposed reforms presented here as quixotic. I recognize that politics, not just principled thinking, will guide regulatory reform. Nevertheless, in spite of the partisan acrimony that rages in Washington, there may be cause for some optimism when it comes to process reform. President Trump has called for an overhaul of financial regulation, and congressional Republicans have enunciated principles of reform that echo many of the process concerns discussed above. Furthermore, there is evidence that Democrats are also becoming increasingly concerned about principles of due process now that they no longer control the administration of regulation. Despite the partisan battles that have defined financial regulation throughout its history, it seems that now might be an ideal time for both parties to find common ground supporting a policy platform that depoliticizes financial regulation and strengthens the rule of law.
References


U.S. House of Representatives Committee on Oversight and Government Reform (2014) “FDIC Involvement in ‘Operation Choke Point’” (December 8).
When Leland Yeager (1924–2018) passed away on April 23, at the age of 93, the world lost a brilliant mind, a devoted teacher, a dedicated scholar, and a man of integrity. I had the privilege of having Leland as a professor at the University of Virginia during my graduate studies in economics in the late 1960s, and later working closely with him as editor of the *Cato Journal* and director of Cato’s annual monetary conference. Of course, I didn’t need to edit his beautifully crafted papers and I studiously avoided sending his papers to a copy-editor! When he retired from the University of Virginia as the Paul Goodloe McIntire Professor of Economics, he left a legacy of excellence, as well as a host of stories about his legendary persona. All of his students know about “the yardstick” and “the stare.”

---

1See “The Yeager Mystique” by William Breit, Kenneth Elzinga, and Thomas Willett (1996). The first two authors were Leland’s colleagues at Virginia, and the third was a former student. Yeager often brought a yardstick to class to draw complex diagrams. He sometimes stood on a chair and extended the graph onto the cinder blocks to get it perfectly to scale. (I am the safe-keeper of “the yardstick,” which I received from Ed Olsen when I hosted a retirement party for Leland at Cato.) “The stare” refers to Yeager’s penchant for avoiding small talk and simply staring at you intensely until you made a fool of yourself and departed his office totally embarrassed. As one graduate student recalled, “From the moment you entered his office you knew you were in trouble. Your simple question like ‘Will you be offering International Trade in the fall?’ was met by this stunned look of disbelief and a penetrating look straight into your eyes” (ibid., 222). Yet, those who got to know Leland found that he was an engaging conversationalist, a considerate mentor, and an excellent dinner host who enjoyed fine wines and knew their provenance (ibid., 224–25).
In this memorial essay, I wish to paint a picture of Leland as a “market grandmaster,” in the sense of his keen understanding of markets and prices along with the role of money in facilitating exchange, and the importance of property rights in shaping incentives and behavior. Along with James Buchanan, Gordon Tullock, Ronald Coase, G. Warren Nutter, Roland McKean, and others, he was an important member of the Virginia School of Political Economy.

The Centrality of Market Exchange and Coordination

James Buchanan, who was instrumental in bringing Leland to the University of Virginia from the University of Maryland (Koppl 2006: 7), held that “economists should be ‘market economists’”—that is, “they should concentrate on market or exchange institutions . . . in the widest possible sense” (Buchanan [1964] 1979: 36). Buchanan also argued that the “most important social role [of economists] is that of teaching students,” and “the most important central principle in economics is . . . that of the spontaneous coordination which the market achieves”—namely, “the principle of spontaneous order” (Buchanan [1976] 1979: 81–82).
Leland excelled in the tasks that Buchanan thought most important. He was an outstanding teacher who placed market exchange and coordination at the center of his lectures. He wanted us to understand how the voluntary actions of individuals, guided by free-market prices and limited government, would lead to mutually beneficial exchanges and a harmonious market order in which people were free to choose. He also wanted us to recognize the significance of money in facilitating exchange and what happens when monetary disorder upsets the market system by distorting relative prices.

In the realm of public policy, he took a principled approach and sought to trace out the long-run effects of alternative policies, whether in the field of international trade or monetary policy. In his presidential address, “Economics and Principles,” at the 1975 meetings of the Southern Economic Association, he reiterated what his students knew well:

> The principled approach to economic policy recognizes that the task of the policymaker is not to maximize social welfare, somehow conceived, and not to achieve specific patterns of outputs, prices, and incomes. It is concerned, instead, with a framework of institutions and rules within which people can effectively cooperate in pursuing their own diverse ends through decentralized coordination of their activities. In the macroeconomic field, it shuns activist “fine-tuning” and aims instead at a steady monetary framework [Yeager 1976a: 560].

Leland espoused “the principle of limited government” and argued that “even when no disadvantages are obvious, there is a presumption (defensible, to be sure) against a new or expanded government activity” (ibid., 562). He hoped that voters “might come to appreciate the value of avoiding myriad interventions and of orienting economic policies, instead, toward legal and monetary frameworks within which decentralized decisions are coordinated by market processes”—but he was not optimistic that voters “might come to judge policies and candidates in the light of the principle of limited government” (ibid., 565).

Yeager, nevertheless, saw it as his duty to “help explain the value of respecting principles not only in the realm of economic policy but also in other interactions among human beings” (ibid.). In thinking about “relevant strands of economic theory,” he pointed to the liberal
the importance, for a functioning society, of people’s having some basis for predicting each other’s actions; the inevitable imperfection, incompleteness, dispersion, and costliness of knowledge; the costs of making transactions and of negotiating, monitoring, and enforcing agreements, and the consequent usefulness of tacit agreements and informally enforced rules; applications of methodological individualism and of property-rights theory to analysis of nonmarket institutions and activities; concepts of externalities and collective goods and the supposed free-rider problem; and the principle of interdependence [Yeager 1976a: 565].

Like Buchanan, Leland was a fellow traveler of the Austrian School of Economics. Indeed, after he left the University of Virginia, he took a position at Auburn University as the Ludwig von Mises Distinguished Professor of Economics. Although he neither subscribed to the radical subjectivism of some Austrians nor fully accepted the Austrian theory of the business cycle, he noted the significance of thinking in terms of individuals, not aggregates; viewed competition as a market process; recognized the subjective nature of value; and understood the importance of sound monetary institutions and private property rights for maintaining a vibrant market system.

In his advanced price theory class, Yeager took time to review the socialist calculation debate and the absurdity of adhering to gross output targets, which would lead to the production of large nails that were useless. When he gave this example of the experience under Soviet central planning, Professor Yeager stopped in the middle of his lecture to leave the room because his face turned red with laughter at the absurdities that resulted from the lack of a real price system. However, he soon regained his composure and returned to carry on his lecture.

Throughout his career as a teacher and scholar, Leland always returned to fundamentals. In his keynote address at a meeting of the Chesapeake Association of Economic Educators, which I helped organize in 1979, he began by saying, “The economic ignorance that is so painfully evident in public-policy discussions is ignorance not of the subtleties of technicalities but of the basic truths. Economists
should make an effort to communicate these basics, and not only to their students but also to a wider audience” (Yeager 1979). Moreover, “in our teaching we ought to point out not only truths but also ancient fallacies that keep being rediscovered in new guises with an air of triumphant novelty, e.g., [the] real-bills doctrine.”

Some of the basics Leland thought should be emphasized include:

- Scarcity, rivalry among uses of resources, opportunity cost, the need for choice, choices made in the light of perceived costs and benefits (prices), the laws of supply and demand, specialization, decentralization, the general interdependence of economic activities, the coordination problem and the way that a market price-and-profit system handles it, the role of money in facilitating the operation of such a system (facilitating the fundamental exchanges of goods and services against goods and services), and the snarl that occurs when the quantity or growth rate of money changes erratically (here is the chief point of intersection between micro and macroeconomics [ibid.].

Finally, Yeager argued that, “in understanding how the economic system and the individuals that compose it behave, it is important to understand the opportunities open to individuals and the signals and incentives that impinge on them. This holds true not only of consumers, workers, and businessmen but also of people in government—politicians, economic regulators, and so forth.” He also warned, wisely, that:

- Failure to appreciate the roles of prices and profits is tied up with failure to appreciate the problems of mobilizing knowledge, of coordinating decentralized activities, and of coping with obstacles to carrying out transactions. There is a tendency, perhaps found even more among high-powered theorists than among laymen, to regard the imaginary extreme case of pure and perfect competition as a standard for judging the real world. Relevant information is assumed to be

---

2 Although his speech, which was given at the University of Baltimore on December 7, 1979, was never published, he prepared note cards that were neatly typed and, jointly, read like an article. The quotes are from those cards, which were preserved in my files.
automatic at hand; information and transactions costs tend to be forgotten. Hard facts of reality—including aspects of scarcity, the fact that resources have to be used in acquiring and transmitting knowledge and in conducting transactions—are seen as “imperfections” of our particular economic system. Reality is blamed for being real. Well, we should beware of planting any such misconceptions in the minds of our students [ibid.].

Because Leland was an authority on monetary economics and international monetary relations, as well as on alternatives to our present system of discretionary government fiat money, I now turn to those topics.

Money Is the Centerpiece of Macroeconomics

In his Chesapeake lecture, Leland called money “the centerpiece of macroeconomics.” He then went on to explain why. I will give a brief overview of some of Leland’s fundamental thoughts about the role of money in a market system.

Say’s Law Is Fundamentally Right

According to Yeager (1979), “There has been too much aggregation in macroeconomics, theoretical and applied—too much of the notion of aggregate demand confronting aggregate supply. Fundamentally, Say’s Law is right: supply of some goods and services constitutes demand for other goods and services; fundamentally there can be no problem of deficiency of aggregate demand.” However, “the exchange of goods and services against goods and services takes place through money.” When the supply of and demand for money do not mesh, monetary disequilibrium can upset the smooth operation of the market mechanism and Say’s Law must be qualified. This is especially true when price and resource adjustments are sluggish.

Consequently, Yeager emphasized that students need “to understand the tremendous importance of money in facilitating exchange and thus in facilitating the division of labor in producing the goods to be exchanged.” In particular, they need to recognize that “money facilitates economic calculation and the comparison of costs and benefits and the signaling function of price and profit” (ibid.).
Yeager goes on to argue that it is “precisely because money is so important to the working of the economic system [that] monetary disorders can have fateful consequences.” Thus, there is a “hitch in Say’s Law: Although ‘fundamentally’ goods and services exchange against goods and services, money is the intermediary in this process; and if the demand for and supply of money get out of balance, these fundamental exchanges are impeded” (ibid.).

Yeager elaborated on this idea elsewhere, explaining that an imbalance between the actual quantity of money and the total of desired cash balances cannot readily be forestalled or corrected through adjustment of the price of money on the market for money because money, in contrast with all other things, does not have a single price and single market of its own. Monetary imbalance has to be corrected through the roundabout and sluggish process of adjusting the prices of a great many individual goods and services (and securities). Because prices do not immediately absorb the full impact of the supply and demand imbalances for individual goods and services that are the counterpart of an overall monetary imbalance, quantities traded and produced are affected also. Thus, the deflationary process associated with an excess demand for money, in particular, can be painful [Yeager 1983: 307].

It is to this theory of monetary disequilibrium that I now turn.

Monetary Disorder

Leland was influenced by the work of Clark Warburton and was put in charge of Warburton’s papers after his death. Those papers are now housed in the special collections at George Mason University. Like Warburton (1966), Yeager (1979) understood monetary disorder as “erratic disturbances to the relation between the actual quantity of money and the demand for money balances.” Those disturbances mean

not only wrong relative prices but also a wrong price level or purchasing power of the dollar, in relation to the quantity of

---

3For a summary of the Warburton collection, see Dorn (1987a).
money, [which] can snarl up exchanges. Anything that snarls up exchanges snarls up the production of goods to be exchanged, with cumulative consequences. Anything that undercuts the reliability of the dollar as a unit of account snarls up the accuracy of economic calculation, with fateful consequences [Yeager 1979].

This “theory of monetary disequilibrium” or “erratic money” has a long history that Warburton (1966: 1–35) resurrected (see Dorn 1987b: chap. 1). In his empirical studies after 1945, he emphasized “an erratic money supply as the chief originating factor in business recessions and not merely an intensifying force in the case of severe depressions” (Warburton 1966: 9). In May 1945, Warburton (ibid., 257) argued that “the theoretical developments since publication of Lord Keynes’s General Theory . . . have shifted attention away from the policies which produced the Great Depression and other cases of large departure from full employment, and have laboriously diverted the energies of economists into fruitless directions.”

Yeager followed Warburton’s lead and wrote an important article on “The Keynesian Diversion” (1973) and another on “The Significance of Monetary Disequilibrium” (1986), both of which are reprinted in The Fluttering Veil: Essays on Monetary Disequilibrium (Yeager 1997).

Money and Freedom: In Search of a Monetary Constitution

Leland was a member of the Mont Pelerin Society and a libertarian; he was a proponent of limited government and economic freedom, which he viewed as part of personal freedom. He was interested in examining a broad range of monetary alternatives to advance monetary theory and offer ideas for fundamental reform. In the fall of 1960, he organized an innovative conference for the Thomas Jefferson Center for Studies in Political Economy at the University of Virginia, which produced a book titled In Search of a Monetary Constitution, published by Harvard University Press in 1962.

That book has had a major influence in thinking about monetary reform from a constitutional perspective. In April 2012, to mark the book’s 50th anniversary, the Liberty Fund held a symposium in Freiburg-im-Breisgau, Germany, which was organized by
Lawrence H. White, Viktor Vanberg, and Ekkehard Köhler—who acted as editors for *Renewing the Search for a Monetary Constitution* (2015). Leland, who could not attend the conference due to health issues, contributed a paper titled “The Continuing Search for a Monetary Constitution” (Yeager 2015).

By “constitution” Leland meant “rules of the game” in a broad sense—that is, both formal and informal rules that would lend predictability to human interactions and help bring about social and economic order. With respect to a monetary constitution, he was critical of the lack of a rules-based monetary policy and what he called “our preposterous dollar”:

> On reflection, our existing monetary system [discretionary government fiat money] must seem preposterous. It is not difficult to understand how individually plausible steps over years and centuries have brought us to where we now are, but the cumulative result remains preposterous nevertheless. Our unit of account—our pervasively used measure of value, analogous to units of weight and length—is whatever value supply and demand fleetingly accord to the dollar of fiat money.

> If balance between demand for and supply of this fiat medium of exchange is not maintained by clever manipulation of its nominal quantity at a stable equilibrium value of the money unit, then any correction of this supply-and-demand imbalance must occur through growth or shrinkage of the unit itself. Money’s purchasing power—the general price level—must change. This change does not occur swiftly and smoothly. Money’s value must change, when it does, through a long-drawn-out, roundabout process involving millions of separately determined, though interdependent, prices and wage rates. Meanwhile, until the monetary disequilibrium has been finally corrected in this circuitous way, we suffer the pains of an excess demand for or excess supply of money [Yeager 1983: 305–6].

Leland argued that “some properties of actual monetary systems are illuminated by contrasting them with imaginary systems” (ibid., 305). He was interested in nongovernmental approaches to correcting deficiencies in the present monetary system. As he stated, “proposals for nongovernmental remedies intrigue me more” than
“remedies within the realm of government money” (ibid., 308–9). In discussing “private money,” Yeager wrote:

As a libertarian, I favor allowing free banking—the competitive private issue of notes and deposits redeemable, presumably, in gold. Notes and deposits would be backed by merely fractional reserves, for efforts to enforce 100 percent banking in the face of contrary incentives and private ingenuity would require unacceptably extreme government interference [ibid., 318].

His preferred scheme, which was designed to eliminate or greatly reduce monetary disorder, was tentatively called the “BFH system,” after Fischer Black, Eugene Fama, and Robert Hall whose writings influenced Yeager and his coauthor Robert Greenfield (see Greenfield and Yeager 1983). Their proposed system is rather complex and is best summarized by Leland in his 1983 Cato Journal article:

Like the reform proposed by Hayek [1976, 1978], [the BFH system] would almost completely depoliticize money and banking. By the manner of [the state’s] withdrawal from its . . . domination of our current system, the government would give a noncoercive nudge in favor of the new system. It would help launch a stable unit of account free of the absurdity of being the supply-and-demand-determined value of the unit of the medium of exchange. The government would define the new unit, just as it defines units of weights and measures. The definition would run in terms of a bundle of commodities so comprehensive that the unit’s value would remain nearly stable against goods and services in general. The government would conduct its own accounting and transactions in the new unit. Thanks to this governmental nudge, the public-goods or who-goes-first problem of getting a new unit adopted would largely be sidestepped. The government would be barred from issuing money. Private enterprise, probably in the form of institutions combining the features of today’s banks, money-market mutual funds, and stock mutual funds, would offer convenient media of exchange. Separation of a unit of account of defined purchasing power from the medium—or rather media—of exchange, whose quantity would be appropriately determined largely on the demand side, would go far toward avoiding macroeconomic disorders.
Yeager: Market Grandmaster

and facilitating stable prosperity. Lacking any base money, whether gold or government-issued money, on which ordinary money would be pyramided on a fractional-reserve basis, the BFH system would not share the precariousness and vulnerability of ordinary monetary systems [Yeager 1983: 323–24].

The question inherent in the Yeager-Greenfield proposal is, Why would the government “give a noncoercive nudge in favor of the new system?” 4 Perhaps Leland would think that question irrelevant, because his aim was to imagine a totally different type of system and to advance monetary theory—not to think that his idealized system would ever emerge, given the political economy/public-choice problems involved.

Leland did not believe the BFH system was the sole or best path to privatization; it was not intended to be “a definitive proposal.” Rather, it was to serve “as an example of one route to privatization” (Yeager 2015: 12). He offered four reasons for studying private alternatives to government money, even if they may not yet be politically feasible:

First, what is politically realistic can evolve. Second, although horrible to contemplate, a collapse of the government dollar would call for drastic reform. Third, keeping the government (or its agent, the Federal Reserve) from printing money will impose some discipline on its fiscal policies. . . . Fourth, considering how private money might work provides opportunities for progress in monetary theory [ibid., 9–10].

Yeager concludes by returning to a general theme, namely:

It is preposterous to try to remedy economic discoordination without even understanding coordinating processes in the first place and without understanding what obstacles and inhibitions sometimes impede productive transactions. Evident ignorance of economics, even at the highest levels of government, must itself sap the business and consumer confidence necessary for business recovery [ibid., 19].

4Gerald P. O’Driscoll Jr. (1983) points to this and other problems with the BFH system.
In his Chesapeake lecture, Yeager points to the importance of the economic way of thinking and methodological individualism in helping us recognize why government tends to overexpand:

The failure to appreciate the fact that decisions are made by individuals pursuing purposes of their own is perhaps worst when it comes to government. . . . The government, instead of being a philosopher-king, is a congeries of persons—politicians, legislators, bureaucrats, judges, and, once in a while, voters—all pursuing purposes of their own and operating with the fragmentary information that comes to their attention. Furthermore, nothing in government corresponds to the market process of spontaneous coordination of decentralized decisions. Nothing corresponds to the market’s way of bringing even remote considerations to the attention of each decentralized decision-maker in the form of prices. Knowledge, authority, incentives, and responsibility are largely fragmented and uncoordinated in the political and governmental process. Far-reaching and long-run consequences of government decision-making receive scant attention. There are reasons for thinking that these characteristics of government result in too much government—too much taxing and spending and regulation [Yeager 1979].

Leland goes on to argue that the free-market system allows for great diversity, which he deems a good thing. The voluntary nature of the market system provides for myriad opportunities for exchanges that reflect the diverse preferences of millions of traders. As he puts it, “A subjectivist and individualistic approach appreciates the diversity of people’s personalities and tastes and the attendant diversity of consumption and work patterns and of lifestyles generally. If we value the uniqueness of each human being, we are led to value a system that caters to diversity” (ibid.). He cautions, however that “recognizing the legitimacy of self-interest does not mean . . . approving of any and all self-interested behavior. It doesn’t justify lying, cheating, and stealing.”

Leland’s understanding of the free-market system made him a strong proponent of limited government and wary of government interventions under the guise of serving the “public interest.”
International Monetary Relations and the Necessity for Choice

In Yeager’s magnum opus, *International Monetary Relations: Theory, History, and Policy*,\(^5\) he notes: “It would seem odd that economists have not yet reached near-unanimity on the need, in framing international monetary policy, to sacrifice one or more individually desirable things. The necessity for choice is, after all, the most fundamental principle of economics” (Yeager 1976b: 651).

To illustrate the need for choice, Yeager (ibid.) turns to the “Doctrine of Alternative Stability,” which posits that among three desirable alternatives only two can be fully obtained. The three alternatives are (1) “freedom for each country to pursue an independent monetary and fiscal policy for full employment without inflation” (internal stability); (2) “freedom of international trade and investment from controls wielded to force equilibrium in balances of payments” (capital freedom/external balance without controls); and (3) “fixed exchange rates.” If a country chooses capital freedom (i.e., open capital markets), it “can maintain a fixed exchange rate only by sacrificing its monetary independence and allowing its domestic business conditions and price level to keep in step with foreign developments.”

Yeager goes on to say, “If countries do not openly choose which one to sacrifice among monetary independence, external balance without controls, and fixed exchanges, the choice gets made unintentionally.” For example, “in the early years after World War II—the ‘transition period’ of the IMF Charter—freedom from controls was sacrificed to national pursuit of full employment. Since then, the choice has become fuzzier as countries seek a compromise in the merely partial sacrifice of each of the three objectives” (ibid., 652).

China is a textbook case in this regard: policymakers have made the renminbi convertible for current account transactions and have been slowly opening the capital account, while making the exchange rate more flexible and using domestic monetary policy to spur the economy. If China is to establish a world-class capital market, it needs the free flow of capital, which means policymakers will have to

---

choose either monetary independence under a floating rate system or a genuine fixed-rate system. Because China is a large player in the global trading system, it would make sense for Beijing to float the renminbi and remove capital controls, using domestic monetary policy to stabilize nominal GDP and maintain long-run price stability. However, in a political system dominated by the Chinese Communist Party, lacking both a genuine rule of law and a free market in ideas, there is little incentive to make the fundamental choice in favor of capital (and personal) freedom. That is why Eswar Prasad (2017) does not think the renminbi will become a “safe haven” currency for some time.

In sum, “historical evidence illustrates . . . the logically inexorable need to sacrifice one of the three objectives forthrightly or two or three of them in part” (Yeager 1976: 652).

Conclusion

Leland Yeager never lost sight of the logic of the market price system, the benefits of limited government and private property rights, the dignity of the individual, and the beauty of the market process and its spontaneous order. Like Buchanan, he understood that the order of the market emerges from the process of voluntary exchange, and that such a result depends crucially on the institutions supporting and framing a free-market system.6

One institution that is of the utmost importance is sound money—that is, money with a predictable and stable purchasing power. Yeager’s emphasis on how monetary disorder (monetary disequilibrium) can disrupt relative prices and their coordinating function, and make economic calculation more difficult, made him deeply interested in examining alternatives to discretionary government fiat money, including privatization. He was searching for a “monetary constitution” that would anchor expectations about the future value of money and thus help facilitate exchange—widening the scope of markets and thus the range of options open to individuals.

6According to Buchanan (1982), “The ‘order’ of the market emerges only from the process of voluntary exchange among the participating individuals. The ‘order’ is, itself, defined as the outcome of the process that generates it.”
References


_________ (1979) Keynote Address at the meeting of the Chesapeake Association of Economic Educators, University of Baltimore, December 7; from Leland Yeager’s unpublished note cards.


BOOK REVIEWS

American Default: The Untold Story of FDR, the Supreme Court, and the Battle over Gold
Sebastian Edwards

Sebastian Edwards has written a very important book on a monumental episode in U.S. history, the great debt default of 1933–35, which was a true turning point in American political and economic history. The episode belies myths that Americans construct about their history: America has strong protection of private property rights, is bound by the rule of law, and pays its debts.

Between 1933 and 1935, a series of events and actions culminated in default on U.S. sovereign debt and corresponding forced defaults on private debts. In recounting that story, Edwards proceeds at a fast pace, as the events themselves did. He writes well, and, at times, the book reads like a novel.

Edwards reprises the period 1929–32 and then moves into the presidential campaign in full swing by March 1932. It’s mostly about FDR and campaign staff and advisers because, incredibly, President Hoover did not campaign until very late. At the core of FDR’s advisers were professors constituting the “Brain Trust,” mainly Raymond Moley, Rexford Guy Tugwell, and Adolfe Berle. FDR leaned on them heavily. As Edwards makes clear, however, they knew little about the issues they would be facing (such as gold and foreign exchange).

It was not at all clear in early 1932 that the gold standard was a problem for America. Jacob Viner had pointed out that the United States...
States had the largest stock of gold in the world. The percentage of monetary liabilities backed by gold holdings at the Federal Reserve exceeded the statutory requirements of 40 percent. Between then and Election Day (November 8, 1932), all that would change.

What changed everything, of course, was that the Federal Reserve never acted to stem the crushing price deflation and tumbling economic activity. Nor did the Fed do anything effective to address the banking crisis. The crumbling banking system would force political action. And the political forces of the day, conservative, populist, and just plain self-serving, would shape the course those actions took.

The “do nothing” tag on Herbert Hoover is incorrect. He conceived and even had a blueprint drawn up for a bank holiday to stop bank closures. However, he let his attorney general and the Fed board of governors dissuade him from the plan. FDR adopted the idea and even used Hoover’s blueprint.

When a new wave of bank failures hit in October 1932, neither the outgoing nor the incoming administration had a plan to save the banks. Compounding this, the interregnum between Election Day and the presidential inauguration was much longer than now. Inauguration Day was March 4, 1933. Hoover was paralyzed by events, and FDR was not ready to act. Edwards effectively describes the situation on the eve of Inauguration Day: “the largest democracy in the world had no banking system to speak of. It appeared that both New York state and Illinois would have to declare banking holidays to stem the outflow of metal.”

Roosevelt used Hoover’s Treasury team (along with his own) to come up with a plan over the weekend of March 4–5. By presidential proclamation, all banks were temporarily closed. As the plan evolved during the following week, banks were classified into three categories: A (opened immediately after the bank holiday); B (in need of recapitalization); and C (to be closed permanently). By March 29, banks with an estimated 90 percent of deposits had reopened. Critically, the Federal Reserve acted as lender of last resort in this period, while the Reconstruction Finance Corporation supplied capital in the form of preferred stock purchases. The banking crisis ended.

In the meantime, Congress quickly passed the Emergency Banking Act to give legislative backing to FDR’s proclamation, which was made under the Trading with the Enemy Act of 1917. The Banking Act contained a provision giving the government the
authority to decree a gold embargo. It was intended to spur depositors to return gold to the banking system. It did not work.

A series of domestic political and international monetary events began forcing FDR’s hand and set in motion a move off the gold standard. A call to remonetize silver galvanized the administration into action. Even those calling for “moderate inflation” were horrified by the prospect of remonetizing silver. Support for inflation was growing in Congress, however, so the administration sought a way to end the deflation, allow for moderate inflation, but avoid unleashing uncontrolled inflation. Debtor relief was the motivation.

The reader may wonder where the Fed fit into this story. The startling answer is almost nowhere. With all the calls for inflation, almost no one thought the central bank might play a role. As I noted earlier, the “Brain Trust” had no background in such matters. People like Irving Fisher were kept on the fringe of the debates. The stalking horse for the inflationists was the agriculture bill. In truth, most of the political forces asking for inflation were seeking a rise in commodity prices, especially agricultural prices. That was certainly true of FDR. In general, monetary policy debates of the time were dominated by monetary cranks and the otherwise uninformed. Politics ruled over economics.

Executive Order 6111 banned gold exports and called into question gold clauses in debt contracts. Gold clauses went back to the Civil War period when the dollar’s link to specie was broken. Gold clauses began to be inserted into debt contracts to ensure that creditors would be paid in undepreciated currency. The clauses did not require that creditors actually be paid in gold but in dollar equivalents of a stated amount of gold. No one expected that most debt would in fact be paid in gold. The purpose of the clauses was to curb government’s propensity to inflate. And, indeed, the clauses worked. All of this would be important in the *Gold Clause Cases* the Supreme Court heard in 1935.

I’ll skip over an extended discussion of the chapter on the London Monetary and Economic Conference of 1933—a critically important conference that might have restored the international monetary order. In what some may see as an anticipation of current events, FDR was not inclined to multilateralism at the time. He preferred bilateral negotiations. FDR would pull the rug out from under his trusted adviser and delegate to the conference, Raymond Moley, who later resigned.
The cases challenging the invalidation of the gold clause were making their way through the court system and would eventually be combined in the *Gold Clause Cases*. They involved different legal issues and presented different fact patterns, including different questions of legal standing and computation of damages. There were American holders of U.S. Treasury debt; foreign holders of U.S. Treasury debt; American holders of domestic, private debt; and foreign holders of U.S. private debt. All had gold clauses.

Creditors and debtors alike were in limbo. There was growing uncertainty in debt markets. Everyone had a reason for the legal issues to be resolved. The resulting Supreme Court cases occupy the heart of the book. I commend Edwards for his command of the legal issues and his ability to render them intelligible to the reader.

There is no question that the U.S. government defaulted on its sovereign debt and that it compelled private issuers to default by voiding the gold clauses. The violation of a material covenant in a debt instrument constitutes a default. Creditors can take legal action, including forcing bankruptcy, after default in private debt cases. The gold clauses were certainly material to the debt instruments that contained them. So there was only one central legal issue in these cases: Did Congress have the legal authority to annul contracts?

The government’s case rested on Article I, Section 8 of the U.S. Constitution, which gives Congress the power to “coin money [and] regulate the value thereof.” However, invoking that clause begged the question of what to do about existing contracts. Even the dissenting justices acknowledged that the government could change the gold content of the dollar. Again, the issue was whether Congress could alter existing contracts.

The Supreme Court found, by a 5–4 vote, that abrogation of the gold clauses in private debt contracts was constitutional. By a vote of 8 to 1, however, the Court found that the congressional joint resolution abrogating the gold clause in public debt was unconstitutional, meaning the government lost the public debt cases on its primary argument. Then, in a 5–4 vote, the Court ruled in favor of the government’s secondary argument: holders of U.S. debt had suffered no losses because they were paid in dollars whose value had increased by 1933. Deflation saved the government’s day.

In my view, the Supreme Court decided on pragmatic grounds. As Edwards indicates, it did so to avoid a head-on clash between two
branches of government. That was only delayed, not avoided. The Court left no clear legal precedent on the primary constitutional question.

Edwards deals ably with the economic consequences. He makes a case that monetary easing in 1934 occurred because of the gold clause decision. But easing had always been in the Fed’s power, even with the gold standard. There is no question that the United States had joined the long list of defaulting governments. Near-term recovery, which was not permanent, came at a long-run cost.

Edwards’s interest in this important U.S. episode stemmed from his experience with the Argentine debt default of 2002. The facts differ, but the legal and economic issues are similar. The U.S. government of 1933 had borrowed in gold dollars and wanted to pay back in paper dollars (to simplify). The Argentine government had borrowed in dollars and wanted to pay back in pesos. Both governments argued economic necessity. The United States litigated in its own court system and won as a practical matter. Argentina litigated in international arbitration panels and lost. One could easily conclude that there are different rules for large and small countries.

Referencing Reinhart and Rogoff’s *This Time Is Different*, Edwards observes that “sovereign debt restructurings constitute a never-ending story, which repeats itself with an astonishing degree of circularity.” I highly recommend *American Default*. It is more than compelling history; it is a tract for our times.

Gerald P. O’Driscoll Jr.
Cato Institute

**The Virtue of Nationalism**
Yoram Hazony

In his eloquent and ambitious defense of the virtues of nationalism, Yoram Hazony has a clear political target: the “emerging liberal construction” that opposes the idea that nations should have their “unique laws, traditions, and policies.” Hazony wants to make nationalism great again, so to speak, and he considers the global elite unduly biased against it. By his account, the self-absorbed cosmopolitan advocates of effacing national diversities and specificities are
disdainful of the habits and affections of localism, and thus lack understanding of history and peoples.

While he is right to be infuriated by the hypocrisy of contemporary elites, he distorts European history to persuade the reader of his concept of nationalism. And that concept seems to be a particular, narrowly defined one that we might call pluralism. His argument, essentially, is that a world of nations is a world where nations accept each one’s right to be different. Alas, we know that this was not always the case—when the principle of nationality occupies center stage, it tends to have little tolerance for any other ideal. Nationalism tends to become an obsession for its followers and to claim the monopoly of good in the public sphere.

An Israeli, Hazony is a man of his circumstances, and sometimes proudly so. He is understandably frustrated by “the international efforts to smear Israel, to corner Israel, to delegitimize Israel and drive it from the family of nations.” His national loyalty pervades the book and, although his sense of belonging is laudable, it leads him to make arguments that are tendentious to the point of being illogical.

Hazony sees liberalism as an international approach to politics, one that plays with gimmicks such as the idea of a universal human nature to foster supra-national government. This latter can be seen as anti-democratic: democracy requires people who speak one language, a lasting legacy of nationalism. It is not by chance that universal franchise is alien to international agencies—they are at best technocracies, whose actions reflect the ideology of a tiny minority rather than the wisdom of the people at large. Hazony thus defends nationalism as the broad political worldview to which Israel, England, and the United States now adhere, in contrast to the United Nations or the European Union.

Could such an understanding of the great game of power in the contemporary world, with advocates of international bodies opposing champions of nations, be a good narrative for European history at large? The Virtue of Nationalism is basically predicated on such an attempt.

For Hazony, the advent of Protestantism in Europe in the 16th century allowed the creation of “national laboratories for developing and testing the institutions and freedoms we now associate with the West.” He sees institutional pluralism and the lack of a strong, leveling central power as an offspring of the Treaty of Westphalia rather
than as a feature of European history that goes back at least to the fall of the Roman Empire.

By so doing, he is blind to the extent to which Europe was pluralistic during the Middle Ages. One of the faults of Hazony’s book is that he considers ideas only at their face value. It is true that the Catholic Church aimed to be universal. But it was precisely this claim to universality that made it oppose political power in the Investiture Controversy, and such struggles between popes and monarchs helped make liberty and pluralism possible. By weakening each other’s hegemonic pretenses, conflicts between the Church and empires helped produce self-governing cities throughout Europe.

But in Hazony’s understanding, the history of Europe becomes “the long struggle of nations such as England, the Netherlands, and France to liberate themselves from the pretensions to universal empire of the German and Spanish Habsburgs (that is, the ‘Holy Roman Empire’).” This is hardly a persuasive story. Let’s forget for a moment that both England and the Netherlands were, for more than four centuries, large colonial empires themselves. Can European history really be seen as a struggle between England and France, on the one hand, and Germany and Austria on the other?

England was intermittently at war with France from the 13th to the 19th century. On the other hand, the current English dynasty once bore the name of “Hanover.” The personal union of the Kingdom of Hanover and the United Kingdom ended only in 1837, with Queen Victoria’s accession to the throne. And Victoria’s husband was German, Prince Albert of Saxe-Coburg and Gotha. Britain and Austria-Hungary were allies in the first half of the 18th century and later in their resistance against France and Napoleon.

European history didn’t start with World War I. Even if it did, that would hardly say much in favor of nationalism, let alone its virtues as a growing medium for tolerance and peaceful coexistence.

But Hazony has a different view. One can’t but admire his intellectual acrobatics. He dissociates nationalism from “national socialism.” For him, “Hitler was no advocate of nationalism,” because the Führer saw the national state “as an effete contrivance of the English and French.” Instead, Hitler longed for his own version of the German Holy Roman Empire. “The Nazi extermination of the Jews in Poland, Russia, and the rest of Europe and North Africa was not a national policy but a global one. . . . It could not have been conceived
or attempted outside the context of Hitler’s effort to revive and perfect long-standing German aspirations for universal empire.”

To see how unfortunate this reasoning is, one needs to remember nothing further than the infamous Nazi motto: “Ein Volk, ein Reich, ein Führer.” The Volk was not to be bounded by the borders of post-World War I Germany but should claim its proper place on the world stage. The nation, then, was to make the state. This is nothing else but the core of nationalism, albeit with distinct racial overtones.

Logically and historically, the alternative to national states is not a national empire (or super-state) but rather a government in which power doesn’t see its legitimacy as a byproduct of ruling over a single nation. This was indeed the case with the old order of multinational empires. Whatever Hitler wanted to build, it was clearly not a multinational, or supra-national, state.

The Volk molds the new political order; it commands its homogeneity. The adoption of one single point of view, that of the “nation,” is what defines nationalism. The tendency toward leveling and homogenizing is thus far stronger in the case of Germany than in the case of multinational empires such as Austria-Hungary, with its 11 official languages. If only because of size, such an empire needs a certain tolerance of different ethnicities, religions, and cultures. The quest for a homogeneous community is precisely what made nationalism so appealing in democratic polities—and it is perhaps the driving force behind its revival in our multicultural societies.

Hazony’s book should be contrasted with Elie Kedourie’s superb work on the same subject (Nationalism, 1960). Hazony offers a classic definition of a nation, “a number of tribes with a common language or religion, and a past history of acting as a body for the common defense and other large-scale enterprises.” Kedourie pointed out that “in nationalist doctrine, language, race, culture, and sometimes even religion, constitute different aspects of the same primordial entity, the nation.”

The national idea is never content with ancillary status once it gets into politics. It drives a movement for political independence as the ultimate political end. It becomes the terminus ad quem of public policy and, for many, the end which justifies all means. “National welfare,” how many have died in your name?

For a great stretch of human history, different ethnicities lived together in supra-national political bodies. They didn’t necessarily
live happily together, but neither did they think the secret of political serenity lay in sharing a homogeneous “nation.”

For Hazony, liberalism is _ipso facto_ internationalism. But is it? For “modern” liberals, of the John Rawls kind, the national state is a necessary evil, as no other entity has proved so efficient in redistributing wealth. For “classical” liberals, of the Milton Friedman kind, the confidence in international trade often goes together with a preference for smaller, local government. Some, like the German Ordoliberalists, try to “encase” national government in supra-national, constitutional projects such as the European Union (or at least a version of it). But liberals are hardly a homogeneous bloc, including on this issue.

Indeed, many liberals thought that a national state with a certain degree of homogeneity was the proper seat for free, democratic institutions. Indeed, liberals hoped national states might eventually be more peaceful than big empires—as republics, they would be freed from the dynastic links that often led to European wars. Ludwig von Mises said something similar in sketching the contours of a “liberal or pacific nationalism” in his _Nation, State and the Economy_ (1919). “The princely state has no natural boundaries. . . . To keep on acquiring new possessions until one encounters an equally strong or stronger adversary—that is the striving of kings.”

It is then particularly ironic that, as a champion of the dangers liberalism may produce “when detached from its biblical and Protestant origins,” Hazony thinks that Mises advocated world government in his 1927 manifesto _Liberalism_. Mises and Hayek, Hazony maintains, “argued that a consistent application of the liberal point of view leads to an international federal state without significant boundaries between nations.”

Mises is typically so stubbornly clear that it should be impossible to misunderstand him. Yet Hazony carefully searches for a quotation that allows him to make the libertarian economist a champion of an allegedly aggressive neoliberal order. He picks a section in _Liberalism_ where Mises talks of “nothing less than unqualified, unconditional acceptance of liberalism” pervading the world and thus breeding the conditions of world peace. The great Austrian was actually criticizing the international body of the time (the League of Nations), expressing hope for “a frame of mind” that looks to see individual rights protected, not just within one’s country but also abroad. While he used the word “superstate,” what he was actually discussing is a liberal sensibility that may traverse national boundaries.
This comes from a Misesian attempt to sketch out “a liberal foreign policy.” That section of *Liberalism* is a remarkable collection of caveats against allegedly peace-fostering policies that could backfire (from “standardized” education to the creation of “economic areas”). Indeed, Mises thinks that “a world order must be established in which nations and national groups are so satisfied with living conditions that they will not feel impelled to resort to the desperate expedient of war.” Such a humanitarian attitude, which is indeed part of the classical liberal legacy, was all the more cogent after the disastrous experience of World War I.

“The unqualified, unconditional acceptance of liberalism” was, for Mises, a “frame of mind” to be accomplished through development and cultural persuasion, not a strategy to be accomplished at gunpoint. Global fraternity was not to result from coercively leveling cultures, nor by forcing national governments into legal straitjackets, but by the prosperity bred by people trading with each other—so thought Cobden, Bright, and Spencer, none of whom was a fan of the British Empire. They all were wary, before Mises, of emergent nationalism, colonialism, and imperialism.

Hazony considers this idea of global fraternity “dogmatic and utopian” as “it assumes that the final truths concerning mankind’s fate have long since been discovered, and that all that remains is to find a way to impose them.” Such a statement has some vague resonance with Mises’s point, but only if we consider the idea that being richer and living longer is the “final truth concerning mankind” and “imposing” as a synonym for allowing people to find their own way out of misery.

One can agree with Hazony that it is naïve to assume that “political life is governed largely or exclusively on the basis of the calculations of consenting individuals.” But to assume that governments are just bigger families is the oldest trick of the apologists for interventionism. “Paternalism” never goes with limited government.

Hazony’s entire profession of faith in nationalism is predicated upon the idea that “mutual loyalty of individuals to one another is the most powerful force operative in the political realm.” But does mutual loyalty have to reverberate in loyalty to a national state? Isn’t it the case that some people indeed feel loyal to their church, their family, their town, their football club far more than, or at least as much as, to an abstract notion called “the nation”?
In his sympathy for England and the United States, Hazony’s nationalism has stronger liberal nuances than he himself would admit. It is hard not to sympathize with his criticism of the hypocrisy of the global elite. But even though he concedes that there is such a thing as a “neo-nationalism” that follows Rousseau and the French revolutionaries and “is known for its tendency toward absolutism,” his rosy portrait of nationalism is not convincing. He envisions an “international order of national states” based on national independence and what he calls “the biblical moral minimum for legitimate government.” In the Bible, “the king or ruler, in order to rule by right” needs to “devote himself to the protection of his people in their life, family, and property, to justice in the courts, to the maintenance of the Sabbath, and to the public recognition of the one God.” In modern jargon, the ruler needs to assure “the minimum requirements for a life of personal freedom and dignity for all.”

Alas, this “all” is limited by the boundaries of what the nation is: a bigger tribe, which is very often ruthless toward those who do not belong to it.

Alberto Mingardi
Istituto Bruno Leoni and Cato Institute

Hans Gersbach

In his speech in the House of Commons on November 11, 1947, Winston Churchill famously said, “Democracy is the worst form of government, except for all those other forms that have been tried from time to time.”

Churchill was and is still right. Even the best democracies have fundamental problems. The Hobbesian problem remains: a monopoly of government power, even if necessary to maintain the rule of law, can lead to the deterioration of democratic systems and the inadequate provision of public goods. Such a development can be currently seen in countries like Venezuela and Turkey.

But even if no fundamental threats to the rule of law occur as they have in Venezuela and Turkey, there are other hidden influences that can slowly erode individual liberty and extend the power of
politicians in otherwise well-functioning democracies. And the electorate may not even perceive what is happening. The inadequate control of representatives by voters, particularly after they have been in office four or five years, is an important factor that contributes to eroding democracies.

Hans Gersbach has been trying to design better democratic systems since the 1990s. His new book, *Redesigning Democracy: More Ideas for Better Rules*, suggests political mechanisms to address some of the inherent problems in democracy. It comprises two different parts. The first part (Contractual Democracy) argues for contracts between political candidates and the citizens, thus allowing for post-election assessment of the candidates’ promises. Gersbach focuses on the fact that inferior short-term projects that show results in the first election period may be selected over superior long-term projects that only come to fruition in the second election period. The second part of the book (Rules for Decision-Making and Agenda-Setting) argues that better political rules for agenda-setting can lead to more optimal outcomes in terms of divisible public goods.

In Part I, Gersbach proposes using contracts with political candidates to improve the democratic system. That produces obvious difficulties: it is impossible for millions of voters to enter into and assess contracts offered by political candidates. Gersbach suggests an independent authority that assesses the performance of the politicians in fulfilling the contract. The candidate who succeeds is rewarded or penalized according to how well he implements the contract. The possible conflicting interests of the “independent” authority, however, are not discussed.

Gersbach shows that, with only elections rather than incentive contracts, the public cannot motivate a politician to undertake long-term projects if the politician’s discount factor is below a certain threshold. Moreover, if reelection prospects are sufficiently uncertain, a politician may not initiate long-term projects even if his discount rate is zero. If politicians can offer contracts stipulating future transfers, however, the situation changes. There, in a two-person election, the candidate with the lower discount factor will become indifferent toward choosing the long-term or the short-term project. The transfers proposed for the contracts can be monetary, or special honors can be awarded. The politician with the larger discount factor will be elected and his reelection prospects are certain. Under this
system, even if the public does not know the discount factors of the political candidates—that is, in the case of asymmetric information—there is an equilibrium in which all politicians will undertake beneficial long-term projects.

In Part 2, chapter 8, Gersbach discusses how a variable level of a public good financed by taxes can be provided within the framework of a democratic constitution, even when aggregate shocks are absent. He shows that an efficient provision occurs if the following four rules are met: (1) a supermajority rule requiring a prespecified vote-share for the adoption of a particular level of the public good; (2) a rule levying the same tax rate on all individuals except to the proposal maker (who is exempt from taxation); (3) no subsidies; and (4) the agenda setter has to pay a fixed amount for agenda setting. Note that all citizens have the right to propose an agenda and to vote on the selected one. The agenda setter is chosen among all agenda proposers by fair randomization.

Gersbach offers a second insight in chapter 8: in the case of aggregate shocks to benefits and costs, the efficiency of democratic constitutions can be preserved if supermajority rules are replaced by tax-sensitive rules, where the required majority is a strictly monotonically increasing function of aggregate tax revenues. A third insight is that it is always possible to find a democratic constitution that leads to a Pareto-improvement compared with the status quo.

In chapter 9, Gersbach takes a different approach to compensate the minority who have lost the first vote on whether the public good should be provided. In the second round, only members of the outvoted minority participate under a unanimity rule to determine the financing of the public good. Therefore, the majority, having succeeded in the earlier vote, can be burdened with higher taxes than the minority, although there is an upper limit for individual tax burdens. This approach allows for some protection of the minority outvoted in the first voting round. Gersbach shows that this scheme, on balance, outperforms simple majority voting.

The final chapter attacks a different problem in providing public goods that may occur in democracies: when a minority with strong preferences for a public good is outvoted by a majority with weak preferences for the good, and everyone shares the same tax burden to finance it. Gersbach again recommends a two-stage approach to solve this problem. First, an initiative group is formed that can propose higher taxes for its own members than for the rest of the
electorate. Second, a general vote is taken on whether that proposal will be accepted. 

*Redesigning Democracy* suggests creative new procedures for enabling democratic systems to achieve better results than are presently available. Gersbach’s proposals admirably do not alter the rights of all citizens to participate in elections and to set the agenda. Hopefully, some existing political systems will consider, if not introduce, the proposed reforms. But scholars working in the field of public choice are always faced with the difficult task of finding new ways to motivate politicians, interest groups, and the electorate to introduce reforms. The innovations proposed in this book are worthy of consideration.

Peter Bernholz
University of Basel

**Capitalism without Capital: The Rise of the Intangible Economy**
Jonathan Haskel and Stian Westlake

In the introduction to *Capitalism without Capital*, economists Haskel and Westlake focus on the concept of “investment.” They argue that “investment is what builds up capital, which together with labor, constitutes the two measured inputs to production that power the economy, the sinews and joints that make the economy work.” Traditionally, when economists measured investment, they were measuring investment in physical goods, plants, and machinery. However, with the advent of the internet in the 1990s, the idea of a new “knowledge economy” emerged, based on what economists began to recognize as the results of research and development (R&D) and the largely nonphysical ideas resulting from it. If this new economy were to be measured by economists, the valuation of these intangible assets would need to be incorporated into their models of economic growth.

In his seminal research into Microsoft Corporation’s financial accounts, economist Charles Hulten found that the traditional assets of plant and equipment were only $3 billion, equivalent to 4 percent of Microsoft’s assets and 1 percent of its market value. This was a stark example of “capitalism without capital”—namely, a 21st-century corporation that develops “specific products or processes,” or invests
in “organizational capabilities” that create or strengthen “product platforms that position a firm to compete in certain markets.”

Haskel and Westlake’s central argument is that there is “something fundamentally different about intangible investment, and that understanding the steady move to intangible investment helps us understand some of the key issues facing us today: innovation and growth, inequality, the role of management, and financial and policy reform.” There are, however, two major differences between intangible assets and tangible assets. First, most measurement conventions ignore intangible assets. As they become more important, however, we are now trying to measure capitalism without capital. For example, conventional accounting practices do not measure an intangible investment, such as developing better software, because making such a measurement is a difficult, arduous process and accountants (being cautious types) prefer not to do so except in limited circumstances (such as after the asset has been successfully developed and sold). Second, the basic economic properties of intangibles make an intangible-rich economy behave differently from a tangible-rich one. Why? First, intangible investments tend to represent sunk costs; second, they generate spillovers; third, they are likely to be scalable; and fourth, intangible investments tend to possess synergies, or complementarities, whereby they are more valuable together in the right combinations.

Haskel and Westlake present economic data for Europe and the United States that show intangible investment overtook tangible investment at the time of the Great Recession in 2008. It did so for five reasons:

- First, labor-intensive services become more expensive relative to manufactured goods, and intangible investments (such as design, R&D, and software development) depend much more on labor. Thus, over time, one would expect intangible investment spending to rise relative to tangible investment.
- Second, new technology also seems to be increasing the opportunities for businesses to invest in intangibles, and, as many intangibles involve information and communications, they can be made more efficient with better information technology (IT).
- Third, while the structure of the economy will affect the relative importance of intangibles, that effect will change over time. For example, while the service sector in the 1990s was more tangible-intensive, the manufacturing sector is now more
intangible-intensive than tangible-intensive and has grown more so.

• Fourth, there is some evidence that looser regulation of product markets and labor markets encourages intangible investment.

• Fifth, market size is critical, because intangibles (such as Starbucks’ brand or Facebook’s software) can be “scaled up” more or less indefinitely.

The authors go through a detailed history of how economists have measured intangible investment. The first real attempts took place in the 1980s, when the U.S. Bureau of Economic Analysis (BEA), in conjunction with the IBM Corporation, began to produce indexes of computer prices that were quality adjusted. Subsequently, those computer price indexes made a major difference in allowing the BEA to measure how much investment U.S. businesses were allocating to computer hardware. By the early 2000s, there was a growing belief among U.S. business economists that firms were spending significant funds on assets that had no physical presence but were valuable and durable. In 2005, Carol Corrado, Charles Hulten, and Dan Sichel initiated their foundational research to measure intangible investment in the United States (and gradually other countries). The result was an operational definition of intangible investment divided into three broad categories each producing a different type of capital asset: (1) computerized investment, e.g., software and databases; (2) innovative property, e.g., R&D and product and service development; and (3) economic competencies, e.g., brand names, business models, and organization-level training.

Haskell and Westlake then turn their attention to the “unusual economic characteristics of intangibles”—namely, scalability, sunkenness, spillovers, and synergies (the “Four S’s”). Scalability means intangible assets can be used repeatedly and in multiple places at the same time, unlike tangible assets. The scalability of knowledge derives from a key feature of ideas—namely, that non-rivalry, and powerful network effects fuel scalability. Scalability is important in the modern economy, as it has been crucial to the success of companies like Google, Facebook, and Microsoft, as well as to creating barriers to potential competitors of these firms. Sunkenness involves irrecoverable costs, and when it comes to intangible assets, unless a company has valuable intellectual
property (IP), such “knowledge” becomes almost impossible to liq-
uidate. This creates a problem because investments with high
irrecoverable costs can be difficult to finance, especially with debt,
in which there is a lack of collateral available to a lender liquidation.

*Spillovers* are often created by intangible investments, and it is rel-
atively easy for other businesses (including competitors) to take
advantage of these intangible investments. While companies may
keep trade secrets, some ideas are simply nonexcludable. Such
spillovers often emanate from company R&D activities, but they are
also found in branding and marketing, copying of organizational
innovations, and employee training. Spillovers matter, argue Haskell
and Westlake, for three reasons:

- First, if companies are unsure if they will obtain the benefits of
  their investments, they will likely invest less.
- Second, there is a company premium in making the most of
  their own intangible investments (or from other companies’
  intangible spillovers), and those firms will have competitive
  success.
- Third, spillovers affect the geography of modern economies,
  such as creative people locating to urban areas to encourage
  connectedness.

*Synergies* are important in intangible-based economies, as ideas
and other ideas (especially in technology) often work well together.
The concept of “open innovation,” defined as when a firm directly
engages with and benefits from new knowledge gathered from out-
side the company, is a major driver of intangible synergy at both the
company and industry levels, as well as for the national and local
economies.

The authors also argue that the “Four S’s” produce two general
characteristics: *uncertainty* and *contestedness*. Haskell and Westlake
believe that in an intangible-rich economy firms would naturally be
expected to exhibit greater uncertainty. In addition, intangibles also
tend to be contested because of the ambiguity of intangible invest-
ment ownership rules. Patent disputes over ownership of intangible
property tend to be less well established than those over tangible
property.

Haskell and Westlake see intangible spending playing an impor-
tant role in the recent trend in secular stagnation—that is, the fact
that business investment is consistently low, despite every
economic indication showing it should be to the contrary. The authors argue that this stagnation is partly due to the shift from tangible to intangible investments occurring since the 1990s. Because of the characteristics of intangibles, the authors believe that, since intangibles can be scaled up, the largest and most profitable firms in industries break away from their less successful competitors. Moreover, because they are unmeasured, the measured productivity and profitability appear high. However, as a result of reduced intangible investment (as happened after the Great Recession), intangible capital building slowed, which generated fewer spillovers, thus causing firms to scale up less, resulting in a slowing of total factor productivity.

The authors further document that the rise of synergies and spillovers from intangibles might be expected to increase wealth and income inequality, as well as inequality between competing companies and increasing differences in employee pay. In addition, intangible-intensive firms will need managers with particular skills and education and will pay them more handsomely. Since cities are places where synergies and spillovers occur, the rise of intangibles makes cities increasingly attractive places to live, thus driving up the price of prime housing. Haskell and Westlake speculate that there is an inequality of esteem, whereby the cultural characteristics required to succeed in the intangible economy are at odds with people having traditional views who are supporters of populist movements in Europe and the United States.

The economy’s greater reliance on intangible investment changes the investment debate, prevalent in the United States and the United Kingdom, that there is inadequate infrastructure investment. Haskell and Westlake note that the importance of spillovers and synergies has increased the importance of places where people meet to share ideas, as well as the importance of public transportation and social spaces that make the urban economy work. Information technologies, however, are reducing the need for some aspects of face-to-face interaction. In addition, the authors see a need for an evolving intangible infrastructure, one built upon standards and norms and a foundation of trust and social capital.

Haskell and Westlake recognize the common critiques of the financial system: it is unsuited to the critical task of business investment, financial markets are oriented to the short term, there is
misunderstood risk, and it places perverse incentives on managers. In an intangible economy, the authors expect to see a movement away from bank lending as the primary source of business financing. Some of this financing shortfall will be replaced by the creation of new debt products secured against IP, but the major driver would be a shift toward the use of equity as a means of financing small- and medium-sized businesses. They also see new efforts at tax reform, in the form of ending the favorable treatment of debt and introducing increased tax advantages for business start-ups, coupled with the emergence of new financial institutions enabling small-scale equity investment and facilitating due diligence.

In addition, the authors also expect to see public equity investment dominated by institutions, many of which will commit to taking larger equity stakes in intangible-based firms, therefore enabling greater overall investment in the economy. The largest of these institutional investors may adopt a different strategy to invest widely across an industry-based ecosystem. Because they have a stake in the industry as a whole, large investors will benefit from such strategic investment (in spite of large spillovers) even if a different firm takes advantage. Moreover, Haskell and Westlake forecast an expansion of venture capital, although the development of serious venture capital sectors in many locales or in entirely new industry sectors is less certain. Finally, if public subsidies to private-sector institutions cannot generate enough public spillovers, the authors recommend further financial support for what they term “truly knowledge-generating institutions,” perhaps including research institutes rather than universities.

What will successful companies look like in an intangible economy, and how can managers and investors create and invest in them? Companies that produce intangible assets will want to maximize synergies, create opportunities to learn from others’ ideas (by appropriating their spillovers), and retain the company’s best talent. Skilled leadership will be increasingly valued, as those managers will guide firms to coordinate intangible investments in different product/service areas and exploit their synergies. Financial investors who place a premium on quality equity research and on insight into firm management will also do well. This will be a challenge, however, as funding equity analysis is becoming more difficult for many institutional investors, because of the inherent tension
between diversification, which allows shareholders to gain from spillover effects, and concentrated ownership, which reduces the costs of analysis.

Haskell and Westlake note five fundamental issues facing policymakers in an intangible economy:

- First, since intangibles tend to be contested, the authors expect an economy increasingly dependent on intangibles to place a premium on solid IP frameworks. But working out a “good” and “effective” IP framework will be a major undertaking.
- Second, creating the conditions for ideas (“synergy”) should be an important objective for policymakers. This challenge includes encouraging effective urban development and research into new forms of collaboration and communication.
- Third, the challenge of financial markets and their underinvestment in scalable, sunk intangible investments needs addressing. The authors expect that a thriving intangible economy will make significant changes to financial architecture to make company investment easier, along with cultural changes in the business environment.
- Fourth, all other things being equal, it is likely that it will be harder for most businesses to appropriate the benefits of capital investment in an intangible economy. Haskell and Westlake expect such intangible-rich economies to increase public investment in intangibles, including, but not limited to, scientific R&D. Thus, a greater proportion of the economy’s investment will be publicly funded, which will create major demands both on the effectiveness of government (as to its competence and impartiality), and on its popular legitimacy.
- Fifth, governments must find a solution for dealing with increasing inequality and social division, two things that intangibles seem to encourage. The authors believe that to make the most of the spillovers and synergies of intangibles requires effective social institutions and trust.

Haskell and Westlake offer the reader much intellectual fodder about a topic of increasing importance in developed (and developing) economies. They have taken the issue of “capitalism without capital” straight on, with a carefully laid out research approach that addresses both macroeconomic issues (such as inequality and financial stability) and microeconomic issues (such as management and IP).
Their approach largely works. The emerging data over the last quarter century convincingly document the shift from tangible to intangible spending in developed economies, such as in the United States and OECD countries. While there is controversy about accounting measurements and standards concerning intangible investments, there is enough evidence to verify the economy-level shift to these types of investments. Nevertheless, the authors make a strong case that much of the intangible spending by firms goes unmeasured.

The authors’ conceptual framework (the “Four S’s”) is useful for identifying and defining the characteristics of intangibles. One constructive criticism, however, is that, while “network effects” are mentioned under “scalability,” they have greater importance for intangible assets than tangible assets because of the overwhelming influence of digitalization on the intangible economy. The authors, however, argue that the shift to intangibles is likely to be accelerating at least in part because of the changing work-related skill sets required in an intangible economy. They opine that the economic results of this shift from tangibles may be responsible for the rise in populist movements in the United States and Europe. There may be some truth to that opinion, because the evolution to advanced manufacturing (robotics and digital controls) has reduced the demand for labor, hence reducing job opportunities for those with low skills. Manufacturing continues to become increasingly capital intensive and less labor intensive. However, this may simply be another example of “creative destruction,” resulting in new types of employment that complement those “creatives” directly working with intangible assets.

Issues that Haskell and Westlake raise regarding the financial sector, while familiar to entrepreneurs and small business owners, are exacerbated by the unique nature of intangible assets. The greater use of equity to finance growth has a long history with initial public offerings (IPOs), but such offerings are not made in the early stages of business growth (and certainly only then for a minority of business enterprises). Before any movement toward the greater use of equity can take place, viable intangible accounting measurements and standards need to be developed in conjunction with improved valuation methods for IP to provide sufficient collateral for newly developed debt instruments.

The authors are not convinced that existing company IP protections can reduce spillovers flowing to competing firms. While
intangible assets are more difficult to clarify for their scope of patent protection, it is not an impossible undertaking by patent examiners. In some cases, this clarification is accomplished administratively; in other cases, legislatively. Yet spillovers from IP are often intentional; network effects may allow for nonenforcement of patent-protected technology to accelerate network effects and subsequently allow the firm to compete with complementary products (whose patents are enforced) based on the original technological innovation. In addition, in the case of IT, the rapid obsolescence of many inventions means business-method patents may not be cost-effective, and firms depend on trade secrets.

Perhaps the authors’ most controversial policy recommendation is to increase public investment in intangibles, largely based on the spillover problem. Haskell and Westlake present this recommendation as a subtle paean to industrial policy. It is not clear, however, whether this policy recommendation is “research” in general, “basic research” (which is predominately funded by government), or “applied research” (where government supplies a minority of such funding). This may be a moot point, however, because of the murky transition between basic and applied research on intangible products. The question of “which industries to publicly invest in” always arises. Yet a firm’s “knowledge appropriation issue” with intangibles appears greater than with tangible assets. Until that problem is resolved, reliance on research institutes focused on developing new intangible assets (eventually available to companies planning on developing commercial products) may be more effective than working solely with universities.

In sum, once the authors present the necessary, semi-dry opening chapters on economic data trends among developed economies, they move through the remaining chapters in an easily digestible fashion. While business management scholars would present much of the material on the intangible economy’s effects on the firm in greater depth and acuity, as well as offer suggestions for its future competitive success, the overarching theme of Capitalism without Capital is clearly presented throughout the book. For those uninitiated to the seismic foundational changes that have occured, and continue to occur, in the world’s developed economies, this book will be an illuminating experience.

Thomas A. Hemphill
University of Michigan-Flint
The "crypto world," as the budding industry based on the application of blockchain technology is often called, is a very tribal place. Perhaps the potential for vast riches in “Web 3.0” causes the splintering. Or perhaps it is an example of self-selection that first-movers in a new industry tend to be aggressive and opinionated. Either way, one need only spend a short time in the right corners of Twitter and Reddit to witness the sort of factional competition usually reserved for politics, religion, and sports.

Just as in those other contests, leaders of each crypto-tribe tend to emerge. One such faction is “Bitcoin maximalists,” a term coined by Ethereum cofounder Vitalik Buterin, of which Saifedean Ammous has become a leader. He and others in this faction see Bitcoin as the only cryptocurrency with a future. In particular, Ammous expects the “Bitcoin standard” to become the new gold standard and the anchor of the international monetary system. Bitcoin maximalists thus resemble old-fashioned gold bugs in insisting that there is only one truly “sound money”: alternative cryptocurrencies are to them what silver coinage was to Grover Cleveland Republicans.

In The Bitcoin Standard, Ammous presents something of a Bitcoin Maximalist Bible: a guide, from an Austrian school of economics point of view, to the historical context of the now famous white paper written by the mysterious Satoshi Nakamoto and to the economic characteristics of Bitcoin that make it so endearing to those who view it as sound money.

The Bitcoin Standard attempts to make the case that Bitcoin is a digital form of money that can provide a viable alternative to central bank fiat currencies. The first four of the book’s 10 chapters explain what money is, how it has evolved over time, and how and why various monies throughout history have succeeded or failed to serve their purpose—namely, to store value and exchange it. The next three chapters dive deeper into the implications for society of sound versus unsound money as defined by Ammous. Finally, the last three chapters explain why the author considers Bitcoin “the sound money of the digital age,” while addressing misconceptions and lingering questions about the still young technology.
The *Bitcoin Standard* certainly has its virtues. To Ammous’s credit, the book does a fine job describing historical examples of money and such concepts as salability, hard money, stock-to-flow ratios, and money as a medium of indirect exchange. One of the best things about Bitcoin entering public consciousness is that it has made people stop and think about money itself—what it is, what makes a good or bad money, and why the government has a monopoly on its issuance. *The Bitcoin Standard* should be commended for pushing the narrative forward that free-market money would, to borrow a phrase from those in Silicon Valley, “make the world a better place” by limiting the power of the state over the individual. When central banks can increase the money supply to pay off government debt accumulated by irresponsible spending, the state is less responsive to the classical liberal conception of government, in which operations are restricted to only what its subjects deem tolerable. Ammous deserves kudos for helping to convey these concepts to a wider audience.

However, whatever praise he deserves for disseminating these ideas to a broader audience is diminished by his packaging of them alongside an exaggerated view of the wonders of “sound money.” To him, everything from the fall of the Roman Empire and the “breakdown of the modern family” to the apparently disappointing state of modern art has “unsound money” to blame. In Ammous’s view, it is no coincidence that “Florentine and Venetian artists were the leaders of the Renaissance, as these were the two cities which led Europe in the adoption of sound money,” and “it was hard money that financed Bach’s Brandenburg Concertos while easy money financed Miley Cyrus’s twerks.” Surely such assertions are better suited to Twitter posts than to a work purporting to convey serious lessons about monetary economics, Austrian or otherwise.

When *The Bitcoin Standard* isn’t merely glib, it tends to suffer from its author’s narrow perspective. For example, his survey of money’s history hardly mentions the large part played by credit and the different forms of money based upon it, including most bank deposits. Ammous seems to believe that the only “sound money” throughout history, until the advent of Bitcoin, has been gold. Yet his analysis of the gold standard leaves much to be desired and fails to provide readers with historically accurate facts that are relevant to the book’s supposed subject—the future of the global monetary order. For example, his assertion that the rise of Hitler and Mussolini, along
with World War II, was triggered by European nations’ abandonment of the gold standard and the embrace of Keynesian economics is historically inaccurate. Likewise, he lets his imagination run wild in blaming what he calls the “Keynesian deluge” for the murder of classical liberal (free-market) economists in Russia, Italy, Germany, and Austria.

The most offensive of his misrepresentations is the book’s strawman treatment of John Maynard Keynes. Ammous accuses him of causing the Great Depression, which broke out several years before Keynes’s General Theory was published in 1936. He also asserts that Keynes “never studied economics or researched it professionally,” an absurd claim to level at a man who almost literally learned economics at his father’s knees and who was, among other things, the author of A Treatise on Money (1930) and a long-time editor of the Economic Journal. Finally, Ammous dwells upon Keynes’s notorious pedophilia, presumably to discourage readers from seeing any merit in his economic theory and policies. How Ammous thought such an ad hominen attack on Keynes would help his argument that Bitcoin can provide a decentralized alternative to central banking is unclear.

One of the more interesting arguments made in the book is the reference to “zero-to-one” events—a term he borrows from technologist Peter Thiel—in which visionaries pioneer the first successful example of a new technology. According to Ammous, such events occurred more frequently in eras of “sound money”—including the “belle époque” of 1871 to 1914, when the gold standard prevailed throughout Europe and the Americas—than in the subsequent era of government-produced fiat monies. That outcome, argues Ammous, occurred because sound money, which holds its value over time, enables individuals to “lower their time preference” such that they can take more time to produce and invest in ever-more complicated goods to “satisfy ever-more remote needs,” and, in turn, advance civilization.

Ammous claims that innovations are mostly “one-to-many”—matters of scaling, marketing, and optimization—under the current fiat money system. Although there is nothing wrong with one-to-many innovations, says Ammous, we should think about why there aren’t as many zero-to-one innovations nowadays. Bitcoin itself is an example of a zero-to-one innovation: it is the first example of blockchain technology, and it created a verifiably scarce digital cash. Ammous, as a Bitcoin maximalist, has little
confidence in other applications of blockchain technology like stablecoins, land-title registration, privacy coins, and uncensorable political discussion. In downplaying those applications, he misses Bitcoin’s one-to-many opportunity: to scale and optimize free-market money and government-circumvention.

It is disappointing that Ammous spends so little time discussing Bitcoin and its potential as a replacement for central bank money. Only the last three chapters of his book deal with Bitcoin. He therefore passes up the chance to have a fair discussion of several unresolved arguments. To his credit, on the debate over fractional reserve banking in a Bitcoin standard world, Ammous lends credence to the vision outlined by the recipient of the first-ever Bitcoin transaction from Satoshi Nakamoto, the late Hal Finney, on the Bitcoin forum in 2010. There, Finney cited theories of competitive free banking to describe a scenario in which Bitcoin is the “high-powered money” that serves as a reserve currency for banks that issue their own digital cash. . . . Some would be fractional reserve while others may be 100% Bitcoin backed.”

However, Ammous doesn’t spend any time addressing the concerns held by many economists that Bitcoin’s limited supply—an attribute lauded by Ammous but exacerbated by the unknown amount of Bitcoin lost forever—would eventually result in undesirable deflation. Nor does he summarize and confront the more technical arguments about mining power centralization or whether the incentive system can survive once the 21 millionth Bitcoin is released and miners are only rewarded by transaction fees. These are important questions if a Bitcoin standard is ever going to supplant central banks.

The result is a missed opportunity to provide a foundation from which others can advance the literature on the important question of how cryptocurrencies will interact with the current monetary system. Ammous’s choice to cater to “internet Austrians” and his “Bitcoin maximalist” Twitter followers by offering a one-sided view of monetary history may help with book sales, which he can convert to Bitcoin and hoard, but it does little to advance our collective understanding of the topic the book professes to address.

Tyler Whirty
Cato Institute
The Fed gained unprecedented power and unconventional monetary tools during the crisis. What have we learned?

At this year’s conference, leading scholars and policymakers will examine changes in the Fed’s operating framework since the financial crisis, consider the effect of Fed policy on interest rates and asset prices, and discuss the future path of monetary policy.
From the Cato Institute’s Center for Monetary and Financial Alternatives

In *Less Than Zero*, George Selgin cuts through the confusions that govern most thinking about monetary policy. It changed my thinking about deflation and inflation when it was first published, and it may do the same for yours.

—RAMESH PONNURU, Senior Editor, *National Review*

AVAILABLE EXCLUSIVELY AT AMAZON.COM
Ten years after the 2008 financial crisis we are again facing the possibility of economic turmoil as the Fed and other central banks exit their unconventional monetary policies. Although central banks will move gradually, unforeseen circumstances could trigger a flight to safety and a collapse of asset prices.

Contributors to *Monetary Policy in an Uncertain World* draw lessons from the decade of unconventional monetary policies and offer proposals for reducing monetary uncertainty, including adopting a rules-based monetary regime.

PAPERBACK: $19.95 • EBOOK: $9.99
AVAILABLE EXCLUSIVELY AT AMAZON.COM
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
www.cato.org