

On the Increasing Centralization of U.S. Money and Credit

Amar Bhidé

Working Paper 21-022



On the Increasing Centralization of U.S. Money and Credit

Amar Bhidé
Harvard Business School

Working Paper 21-022

Copyright © 2015, 2020 by Amar Bhidé.

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Funding for this research was provided in part by Harvard Business School.

On the Increasing Centralization of U.S. Money and Credit

My current work on the debt market liquidity (Bhidé 2015) and antecedent book (Bhidé 2010) critiques the de facto centralization of credit extension in the U.S. This essay examines how the increased centralization of credit maps into a decline in the decentralized production of money and the dysfunction this entails.

The argument is half-Hayekian in the sense it argues for an important role for the decentralized, private sector creation of the medium of exchange (which Hayek presumably would have approved of) but anchored in a government monopoly for creating base money (that Hayek did not favor).

Recapitulating the basics

Like marriage, employment, and other social arrangements, money is an elusive construct that serves multiple intersecting functions and has no sharp natural boundaries. Thus money, as the “dollar” or the “euro”, can serve as a unit of account the way liters and gallons are used to specify volumes. Or as with dollar bills or euro notes, money can serve as a medium of exchange, to pay for things now, and as a store of value to pay for things in the future. Forms of money that serve as a medium of exchange also naturally serve as stores of value, otherwise they wouldn’t be accepted for payment. Assets that serve as stores of value do not necessarily serve as mediums of exchange however, and the degree of their “moneyness” is typically associated with their liquidity – the ease with which they can be converted to forms that can serve as a medium of exchange.

My argument pertains to money directly used as a medium of exchange, necessitated in Jevons’s famous phrase by the absence of a “double coincidence of wants.” This use of money (and its associated unit of account) derives from a combination of decree and convention. Thus the Russian government may decree rubles as the legal tender, but the black market may demand \$100 bills. And acceptability may vary with circumstance: the seller of a house will require a certified check, whereas a grocery store may take personal checks.

In the US, as in most other parts of the world, the medium of exchange is almost entirely produced in one of two ways. First, the government issues “high-powered” (or “base”) money. This money is, by decree, legal tender. Second, banks create “on-demand” claims when they make loans. These claims, through a complex combination of rules, conventions, and clearing mechanisms, are, for all practical purposes, also legal tender. And just as the government can “print” new base money out of thin air, as it were, a system of “fractional banking” gives banks a similar privilege. The demand deposits that banks create when they extend credit are only partially backed by the government’s base money, although technically everyone who has a demand deposit has the right to convert to base money.

Other forms of private debt issuance involve the transfer of base money or demand deposit balances rather than the creation of new money that can be used for payment. For instance, when IBM issues commercial paper, it receives, dollar for dollar, the value of the paper issued. The commercial paper itself may then be traded and therefore be regarded as having the monetary property of a store of

This working paper was prepared for the "Money Design in Global Perspective" conference held on June 1-2, 2015 at the Harvard Law School. It draws heavily on the ideas in my working paper “The Hidden Costs and Underpinnings of Debt Market Liquidity”

value. But to actually use the paper to buy something, the owner of paper has to sell it first for someone else's base money or demand deposit balance. Or, to take a more complex example, someone may sell her property for a note, thus creating debt "out of thin air," with no money changing hands. But here too, the note holder cannot spend the note: She may sell it, receiving someone else's base money or deposit balance; or, she can borrow against the note from a bank, receiving a spendable demand deposit. But again, the money creation is through the bank's loan against the note – a loan which could equally have been made against the property she had sold.*

A (Very) Brief History

The two-tiered, government plus fractional banking, system of money creation in the US dates back to the early days of the Republic, although the specific forms, convention, and rules have changed significantly over time. From independence until the Civil War, government money issuance comprised coinage ("specie") by the Mint. Bank-issued money mainly comprised notes (in the bank's name) given to borrowers as a loan. Bank notes were only nominally convertible to government minted specie, because banks issued many times more notes than they had specie in their vaults and in normal times the notes were passed from hand to hand as a medium of exchange. This system had significant drawbacks: because of variations in confidence about the specie backing the notes of different banks, different notes traded at varying discounts to their nominal conversion value. And in times of financial stress, virtually all notes would cease to be accepted for payments leading to a collapse in the medium of exchange.

The state – or more properly States, since most banks did not have Federal charters[†] – undertook a variety of measures to sustain confidence in banks and the notes they issued. These included scrutiny of applications to start new banks, periodic bank examination, note insurance (a predecessor to deposit insurance), and requirements for minimum holdings of specie (a predecessor to reserve and liquidity requirements). By most accounts, these measures were ineffectual.[‡]

Banking legislation passed in the Civil War radically changed the composition of base and bank-issued money. The Federal government abandoned the stricture against the government issuance of paper currency while ending issuance of notes by banks. Coinage continued, but base money increasingly comprised the government's paper money. And when banks could not extend credit in the form of notes, they did so by crediting the checking accounts of borrowers (a practice that had started before the Civil War).

* The case of rising market values of assets that can serve as collateral for bank borrowing muddies the issue of what leads to creation of money a bit. For instance higher home or stock prices can support more mortgage or margin debt – and thus the creation of spendable money. Even here though, the step of a bank extending credit is necessary.

[†] The Bank of North America, chartered by the Congress of the Confederation in 1782, the First Bank of the United States (chartered in 1791), and the Second National Bank (chartered in 1816) were exceptions to the state chartering of banks.

[‡] Some do regard the "free banking" rules that emerged in the last decades of the antebellum period to have been a success.

The uniformity of government-issued paper currency reduced the costs of hand-to-hand payment but the new arrangements did not end collapses in the medium of exchange. Now when depositors lost confidence in banks they would “run” to convert their account balances into government-issued paper instead of trying to convert notes into specie. Just as banks previously issued more notes than they held specie, they now created more balances in demand deposits than they had government-issued currency. A run would therefore force banks to suspend the conversion of deposits into paper currency – making checks drawn against the deposits unacceptable for payment.

To prevent these runs the Federal government increased the resources devoted to bank examination, created the Federal Reserve system in 1913 to serve as a lender of last resort (and alleviate seasonal shortages of paper currency that were thought to trigger runs), and, through New Deal banking legislation in the 1930s, created a broad nationwide system of deposit insurance (by then, all prior statewide systems had ceased to operate). These measures seemed, at least till 2008, to put an end to collapses in the medium of exchange.

Tradeoffs

The creation of money “out of thin air” by banks has long been regarded as dangerous. In the antebellum era, the hostility to bank-issued notes reflected a broader suspicion of paper currency (which state governments and the federal government were not allowed to issue). Thomas Jefferson referred to note-issuing banks as, “more dangerous than standing armies” (Klebaner 1974 p. 49). Some states, including Arkansas, California, Iowa, Oregon, and Texas refused to incorporate banks at various times before the Civil War.

Opposition to fractional banking has continued even though banks no longer issue their own notes. Various “narrow banking” alternatives have been proposed that would require 100% reserve requirements on demand deposits, making banks pure credit intermediaries. As in the antebellum era, the opposition to fractional banking reflects concerns about relying on highly-leveraged and therefore “runnable” financial institutions to produce much of the medium of exchange. Opposition may also derive from a sense of inequity: why should the public give private businesses the profitable privilege of seniorage?

Yet, since the passage of the New Deal banking acts, there have been no systemic “runs” on banks covered by federal deposit insurance. Even in 2008, we should note, there was no run on the “regular” banking system. Rather, it was primarily funds held outside FDIC-covered accounts that were immobilized. These included money market funds that had previously been declared to be “unrunnable.”

The persistence of fractional banking even in times when it was prone to systemic collapse is especially noteworthy. Even as antebellum lawmakers sought to prohibit note-producing banks, such banks proved irrepressible. States that refused to charter banks could not escape the circulation of paper money, however. Banks from Illinois and Wisconsin, for instance, flooded Iowa, which did not charter banks, with “some of the worst money in the Union” in the mid-1850s (Merritt 1900 p. 137 cited in Klebaner 1974 p. 20).

One explanation for private note production in the first half of the nineteenth century is that it provided a medium of exchange that could keep up with a rapidly growing economy. If the medium of exchange had been restricted to minted specie, its growth would have been tethered to the overall balance of trade (with the supply of specie increasing when trade was in surplus and shrinking when the nation's trade fell into a net deficit) or the discovery of new sources (as happened with gold in California). Of course, the commitment to maintain convertibility meant that there had to be some long-term relationship between the volume of bank notes in circulation and the specie backing them, but it did not have to be rigid.

But can the bank production of money serve the public interest when with the end of the gold standard gives governments considerable flexibility in the quantity of base money they can issue?

In a word, yes.

It isn't just the absence of mutuality (A produces what B wants but B doesn't produce what A wants) or synchronicity (A produces what B wants, B produces what C wants and C produces what A wants – but not at the same time) that necessitates a medium of exchange; it is also uncertainty. In a dynamic economy, we cannot predict what we will want and when – or who will most value what we produce when we produce it. Money allows us to keep our options open and engage in a sequence of complete but unplanned transactions without any need for complicated multi party contracting or settling up.

Yet, as mentioned, for money to serve as a medium of exchange, it must also be a source of value and therefore scarce. Just an undersupply of money can be damaging, an oversupply can render it useless. An effective medium of exchange must be sparingly produced, costly – at least in terms of the opportunity cost – to hold and cannot simply be sprinkled over the land as if by helicopter in the hope that it will fall into the right hands. But the standard “market” solution used to efficiently allocate scarce non-monetary resources cannot be used to make and distribute money. The medium of exchange cannot itself be sold for money. Moreover, its marginal utility for any particular user fluctuates. Yet, renting it when needed to the highest bidders risks placing it with those who have most overestimated their capacity to repay.

Relying on decentralized lending by banks – which take into account the borrower's capacity to pay interest and repay principal – helps solve the problem of efficient, “as needed” placement. And, as with many non-monetary goods and service, particularly ones where there are no significant economies of scale, the decentralized production of money as a by-product of decentralized lending arguably strikes a better balance between too much and too little money overall than would a purely centralized process.

This is not to underestimate the utility of centrally produced base money in setting the value of bank-produced money. Even under the gold-based monetary system of the antebellum, when a plethora of bank notes of circulated, it was the coinage of the mint that provided an anchor. It is difficult to imagine how, in the absence of gold standard, a uniform medium of exchange (or standardized unit of account) could exist without some core of the money supply being provided by the government.

Nor is it the case that all forms of bank lending help accommodate the local and temporal fluctuations in demand for a medium of exchange. In particular, making long-term loans for home mortgages, for one example, simply adds to the long term amount of money in circulation. The monetary purpose of buffering unsynchronized production and consumption is better served when banks conform to the classical “real bills” doctrine of making “short-term, self-liquidating loans to finance the conversion of raw materials into goods and their transportation to market” (Litan 1987 p. 14). In other words, while some maturity transformation is vital to the function that justifies the privilege of seniorage this is not the case for more extended maturity transformations.

Policies promoting centralization

In my current and previous work I have argued that public policy has directly and indirectly encouraged the extension of credit based on centrally developed and administered statistical models while tilting against decentralized lending, especially of the “real bills” variety.

An important indirect path to centralization has been through rules-governing retirement funds. Tax policies have encouraged a shift from pay-as-you-go intergenerational transfers (within families or through Social Security) to a savings based approach to funding retirement. Employers and individuals can make pretax contributions to retirement plans and returns earned by the plans are tax-exempt. At the same time, ERISA rules intended to control the dishonest, incompetent, or reckless management of pension assets have encouraged investment in diversified portfolios of tradable assets while discouraging the buying of illiquid assets. And increased pension demand for tradable securities has encouraged the “securitization” of loans that banks once held to maturity.

Rising institutional ownership of debt has reduced the proportion of debt held by banks, not the total amount of bank-owned debt (since total indebtedness has also increased). But the form in which banks own debt has changed. Capital and liquidity rules after the 1980s have encouraged banks to replace loans held to maturity with tradable “securitized” loans of the sort also favored by pension funds and other institutional investors.

Additionally, I suggest securitization, especially of small loans, has fundamentally altered how credit is extended. Using parsimonious statistical models to screen loan applicants assuages the concerns of arm’s length investors about information asymmetries and the quality of loan originators by making the lending process transparent. Nearly everything can be disclosed, including the structure of the model and all the data inputs. Other rules have provided a further impetus to model-based lending. Housing policies guarantee “conforming” mortgages made in accordance with a specified model. And using parsimonious statistical models protects loan originators from the risk of violating anti-discrimination laws.

Using parsimonious models in turn effectively centralizes credit extension. The judgments of one modeler (or small modeling team) rather than of dispersed lending officers and credit committees govern who gets credit. Parsimonious models are also structured to include only those variables that predict overall default rates, while excluding variables that matter in specific places for specific borrowers.

The statistical approach also happens to appeal to many bank CEOs. CEOs of larger firms have more leeway to pursue private benefits and secure higher compensation than the CEOs of small firms, where stockholders can exercise more control. The control problem is especially acute in US banks because regulators discourage unfriendly takeovers and prohibit acquisitions by non-banking entities. Bank CEOs may therefore push securitization – and loans that are easy to securitize – where the capital and staff constraints on growth are lower than in traditional lending. In particular, CEOs will tend to favor mortgage and consumer loans over loans to small- and medium-sized businesses (that happen to closely match the lending prescribed by the real bills doctrine).*

Well-intentioned efforts to revive the economy and reduce banking risks after 2008 have magnified these centralizing tendencies. Tougher capital and liquidity requirements have further encouraged banks to favor securitized holdings and avoid loans that cannot be securitized. Meanwhile, the US Federal Reserve, the Bank of England, and, most recently, the ECB have purchased huge amounts of securitized debt but have not purchased traditional loans. At the same time, small banks that are more likely to make and hold traditional loans have been disproportionately disadvantaged by new banking rules than large banks (that can more easily spread the fixed costs of compliance).

Unintended consequences of the policy bias

The greater centralization of credit and money creation I argue has significant drawbacks. If the problem with traditional lending practice is unwarranted bias, then the problem with statistical discrimination is that it doesn't adequately discriminate between actual differences in creditworthiness. Borrowers who shouldn't get credit, do, while those who should, do not. The increase in lending mistakes that arise from ignoring idiosyncratic, local information falls particularly heavily on good borrowers; even if they do get credit, their rates reflect the greater number of bad loans made.

Policies that favor easily securitized consumer loans over credit supplied to small- and medium-sized businesses may impair economic dynamism. It is noteworthy that the extended decline in startups and fast-growing small businesses (Decker, Haltiwanger, Jarmin, and Miranda (2014)) has coincided with a huge increase in securitized consumer credit.

Centralization of money and credit increases systemic risks. When identical or similar models are widely used to underwrite or provide credit ratings for a very large number of loans, misspecifications of the model or misestimates of parameters – because of omitted variables, structural changes in the economy, or simple mistakes – can lead to poor lending choices on a large scale. In contrast, whereas mistakes, malfeasance, or bad luck are unavoidable in case-by-case lending, in the absence of a collective mania, decentralized credit decisions protect against mass contemporaneous failure. More broadly, because greater centralization entails more concentration of the banking sector, the credit system becomes more vulnerable to the mismanagement or misfortune of a few mega-banks.

* Moreover, banks cannot easily recreate their capacity to make traditional loans once they have shut it down, and regulatory barriers to entrants seeking to fill gaps left by incumbents are formidable.

And perhaps most ironically, greater centralization actually reduces the ability of central bankers to counter economic downturns by impairing the “transmission mechanisms” of monetary policy. Thus zero interest rate policies and quantitative easing have apparently enabled large companies to issue long-term bonds at miniscule rates and in amounts that outstrip their investment opportunities. Meanwhile, small- to medium-sized businesses that cannot issue bonds continue to struggle for credit because banks have shed their capacity to make loans to such businesses. Money money everywhere, but for the credit thirsty, not many drops to drink.

References:

- Bhidé, Amar. 2010. *A Call for Judgment: Sensible Finance for a Dynamic Economy* New York: Oxford University Press
- Bhide, Amar. 2015. The Hidden Costs and Underpinnings of Debt Market Liquidity. Columbia University Center on Capitalism and Society Working Paper No. 79. Available at SSRN: <http://ssrn.com/abstract=2206996> or <http://dx.doi.org/10.2139/ssrn.2206996>
- Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives*, 28(3): 3-24.
- Klebaner, Benjamin J. 1974 *Commercial Banking in the United States: A History*. Hinsdale, Ill.: The Dryden Press
- Litan Robert E., 1987 *What should banks do?* Washington D.C.: The Brookings Institution