

WORKING PAPER 2117

Bank Money Creation and the Payments System

Biagio Bossone

November 2021



BANK MONEY CREATION AND THE PAYMENTS SYSTEM

BIAGIO BOSSONE

International financial consultant

Senior adviser, Finance, Competitiveness & Innovation, The World Bank

email: Biagio.Bossone@gmail.com and BBossone@WorldBank.org

ABSTRACT

This article investigates how the role that banks play in the payment system space affects their money creation power and process. In particular, the article analyzes how the payments market share of each bank affects its money creation power and how payment settlement technologies and rules determine the banks' demand for funding and, hence, their money creation power. Also, as the power to create money enables money creators to extract extra-profits or rents ("seigniorage") from the economy, the article evaluates analytically how banks extract seigniorage through money creation and how bank seigniorage differs from profits from pure financial intermediation. By showing the central role that payment systems play in the context of such an important economics topic as money creation, the article seeks to emphasize the relevance of payment system analysis for macroeconomic theory and practice and points to the need for achieving better integration of the two disciplines.

Key words: Bank; Bank money creation; Central bank policy; Demand deposits; Financial intermediaries; Funding; Lending; Payment and settlement systems.

JEL Code: E51; E58; G21.

OCTOBER 2021

BANKS' MONEY CREATION AND THE PAYMENTS SYSTEM

1. INTRODUCTION¹

Today, the largest share of the monies circulating worldwide – even including the fast-growing incidence of digital currencies – are bank deposits. These are claims on commercial banks, issued by commercial banks to their clients, which are transferable to third parties as means of payments and redeemable in cash at par by their holders on demand.

This article reconsiders the role of commercial banks as money creators and investigates how their role in the payment system space affects their money creation power and process. In particular, the article analyzes how commercial banks' payment obligations and the rules governing payments settlement determine (and constrain) the extent to which each bank can exercise the power to create money. For clarity, commercial banks (henceforth "banks") are here conventionally defined as financial institutions that accept demand deposits from the public *and* lend money to borrowers in the form of demand deposits issued to them. All other functions, activities and services that can be associated with banking are not considered in this article, as they are not relevant to its purpose.

The article briefly reviews the mainstream theory of banks as financial intermediaries and considers recent contributions to the idea that, unlike other intermediaries, banks hold the power to create money (Section 2). The article notices that even these recent contributions ignore the constraints that derive for bank money creation from the banks' payment obligations and analyzes such constraining factor and how it operates. The article also evaluates how the payments market share of each bank affects its own money creation power and how payment settlement technologies and rules determine

¹ International financial consultant, senior advisor to the Payment Systems Development Group, the World Bank, and external expert, Monetary and Capital Markets Department, International Monetary Fund. The opinions expressed in this article do not necessarily coincide with those of the institutions within which I am associated, and the responsibility for any errors therein contained is obviously mine only. I am grateful to my wife Ornella for her constant support.

the banks' demand for funding and, hence, their money creations power (Section 3). Furthermore, as the power to create money enables money creators to extract extra-profits or rents ("seigniorage") from the economy, the article evaluates analytically i) how banks extract seigniorage through money creation and how bank seigniorage differs from the profits derived from pure financial intermediation, ii) how seigniorage originates from bank money creation, and iii) how seigniorage relates to profits from pure financial intermediation (Section 4). Finally, by showing the central role that payment systems play in the context of such an important economics topic as money creation, the article concludes by emphasizing the relevance of payment system analysis for macroeconomic theory and practice and points to the need for achieving better integration of the two disciplines (Section 5).

2. DO BANKS CREATE MONEY? A BIRD'S EYE VIEW OF THE LITERATURE

For a century or more, starting in the early 1900s, economists recognized that banks have the power to create money. Some held that they can actually do so *ex nihilo* (out of nothing), while others thought this option would be feasible only for the banks acting collectively, as a system, and would not be possible for individual banks acting alone. The theory remained prominent until the early post-war years, when it disappeared altogether from the research literature and textbooks and adherents continued to thrive only within heterodox schools.²

In most contemporary academic literature (see below), banks are represented as pure financial intermediaries. Accordingly, they have no power to create credit and money; they can only buy assets with funds they have previously acquired in the form of deposits or through issuance of shares or bonds. This view is the counterpart – on the finance side – to what loanable funds theory is on the macroeconomics side, whereby, simply put, new investments can only be funded with accumulated

² For a detailed review of the theories of banks as money creators and their historic evolution, see Werner (2016).

(pre-existing) savings; accordingly, financial intermediaries (banks amongst them) help the two to match each other at the "right" price (i.e., the equilibrium interest rate).

The view of banks as mere financial intermediaries is at least as old as Von Mises (1912), which noted that the activity of the banks as negotiators of credit is characterized by the lending of other people's (that is, borrowed) money. Several decades later, the view was articulated in the seminal works by Gurley and Shaw (1955, 1960), which rejected the idea that banks hold the power to create money and explained instead that banks engage in multiple creation of special liabilities, in relation to the class of assets they hold, which enables them to collect deposits and lend them out to clients. Tobin (1963), too, refused the idea that banks create money and credit, and still later Sealey and Lindley (1977) modelled banks as financial firms specialized in a transformation process that involves the borrowing of funds from surplus spending units and the lending of those funds to deficit spending units. Such process consists of a multistage activity whereby loanable funds are borrowed from depositors and lent to enterprises that use them to purchase capital, labor and material inputs and employed them in the production of earning productive assets.

Since then this theory has prevailed, with authoritative contributions spanning from Fama (1980), to Diamond and Dybvig (1983), Bernanke and Blinder (1988), Riordan (1993) Bernanke and Gertler (1995), Kashyap et al. (2002), and more recently Brunnermeier and Sannikov (2016) and IMF's Adrian and Mancini-Griffoli (2019), just to cite some of the most relevant works. The theory of financial intermediation encompasses also the so called "credit view," which has become a workhorse hold in macroeconomics. The theory proposes a banking credit channel of monetary transmission, whereby banks are financial intermediaries that transform liabilities into assets with different

characteristics (in terms of maturity and liquidity) and monitor debtors' repayment capacity and actual repayments.³

Widely read books on money and banking (e.g., Cecchetti, 2008; Casu and Girardone, 2006; Matthew and Thompson, 2005), and the latest edition of one of the most popular textbooks in macroeconomics (Mankiw, 2019), feature banks as mere conduits for the transfer of funds from savers to investors.

In recent years, the issue has been revisited and the role of banks as money creators has been revived, with great resonance also in the specialized blogosphere.⁴ Among the most cited works are McLeay et al. (2014a, b) and Jakab and Kumhof (2015). The former clarifies that money creation in practice differs from some popular misconceptions: banks do not act simply as intermediaries, lending out deposits that savers place with them, and nor do they "multiply up" central bank money to create new loans and deposits, rather, whenever a bank makes a loan, it simultaneously creates a matching deposit in the borrower's bank account, thereby creating new money. The amount of money created in the economy ultimately depends on the monetary policy of the central bank, which, in normal times, is carried out by setting interest rates. Similarly, Jakab and Kumhof, cit., recognize that banks create new funds in the act of lending, through matching loan and deposit entries, both in the name

³ As a leading proponent of the credit view, Bernanke (1993) frequently refers to "credit creation," clarifying that: "...By credit creation process I mean the process by which, in exchange for paper claims, the savings of specific individuals or firms are made available for the use of other individuals or firms (for example to make capital investments or simply to consume). Note that I am drawing a strong distinction between credit creation, which is the process by which saving is channeled to alternative uses, and the act of saving itself... In my broad conception of the credit creation process I include most of the value-added of the financial industry, including the information-gathering, screening, and monitoring activities required to make sound loans or investments, as well as much of the risk-sharing, maturity transformation, and liquidity provision services that attract savers and thus support the basic lending and investment functions. I also want to include in my definition of the credit creation process activities undertaken by potential borrowers to transmit information about themselves to lenders: for example, for firms, these activities include provision of data to the public, internal or external auditing, capital structure decisions, and some aspects of corporate governance." (p. 50).

⁴ Just Googling "banks create money" at the time of writing (20 October 2021) gets 541.000.000 results.

of the same customer, on their balance sheets. They view banks as fundamentally monetary institutions unconstrained by a deposit multiplier mechanism and constrained only by profitability and solvency considerations.

On the other hand, money creation by commercial bank has been a tenet of post-Keynesian economics since at least the 1970s (e.g., Moore 1979, 1983), and is one of the founding elements of the circuit theory of money, which studies the role of money in monetary production economies with historical time (Realfonzo, 2012). The literature on the circuit theory of money is too vast for trying to provide here a list of references that would do justice to the many contributors to this brand of non-mainstream monetary economics. One contribution that offers a good illustration of the theory's fundamentals is by Augusto Graziani (cit.), one of the theory originators and most inspiring "circuitist." Another excellent reference is Rossi and Rochon (2003).

A decisive contribution to the *banks-create-money* theory is Werner (2014), as further elaborated in Werner (2016), whose analysis goes at the very heart of what uniquely enables banks to create money and whose empirical tests reject the financial intermediation and fractional reserve theories. According to Werner, banks create gross credit just like nonbanks, but, unlike in the case of nonbanks, this credit is not counter-balanced by an equal reduction in credit balances elsewhere, leaving a net positive addition to credit and deposit – hence money – balances: net credit creation. Such credit creation is visible in the permanent expansion in the bank's balance sheet. As Werner explains, this is possible because banks are typically exempted from regulations that require financial firms to hold client deposits in segregated accounts with banks or money market funds. Thus, while nonbank financial intermediaries may not mix their clients' deposit accounts with their other liabilities towards the clients, banks are permitted to keep customer deposits on their own balance sheet. This means that banks are able to access the records of the customer deposits held with it and add a new "customer deposits" as payable liabilities of the banks arising from loan contracts.

Finally, two non-academic contributions deserve mentioning: one is Fullwiler (2012), which provides a rigorous reprisal to Krugman's rearguard defense of banks as pure intermediaries, arguing that banks are not limited by their deposits either individually or in the aggregate; the other is Sheard (2013), which explains analytically that banks do not lend out deposits or reserves but lend by creating deposits.

3. BANKS' MONEY CREATION AND PAYMENTS ACTIVITY

The theories of banks as money creators, including the most recent versions just cited – ignore a crucial aspect that is strictly inherent in banking: the role that banks play in the payment system and the constraints that derive for their money creation power from their payment obligations and from the structures and rules of the payment system(s) where they operate.

Indeed, banks create the money they lend by fiat, in the form of newly issued demand deposits, and do not need central bank reserves to do that. They simply create their own demand deposit liabilities by crediting borrowers' accounts with the funds loaned to them; by such an act of creation, new liabilities and assets are simultaneously recorded on the bank's balance sheet. In fact, banks create money not only when they lend but also when they sell deposits, for instance, in securities (re-)purchasing operations. Deposit lending features very close analogies to deposit selling: as banks issue deposits to customers in exchange for money, banks become owners of the money received and acquire the rights to use it as they wish, subject to existing laws and regulations (Bossone, 2021a).

However, when banks create money by extending new loans, they need to avail themselves of the resources required (by interbank conventions, payment system rules, or central bank regulations) to settle the mutual obligations that are triggered when the newly created deposits are mobilized by borrowers to make payments or when they are redeemed by their holders against cash; otherwise they would default on their obligations and incur the related costs and penalties, including market punishments and regulatory sanctions. Indeed, contemporary payment systems in most jurisdictions

across the world require that settlement takes place in central bank money since this is the safest and most liquid asset that an economy can use to settle transactions (CPSS-IOSCO, 2012). The other key reason for so requiring is that, as long as banks demand reserve balances to pay each other, the central bank can set the marginal price of reserves, thus determining the opportunity cost for banks to lend funds to each other and thereby influencing interest rates in the money markets and further on across all assets markets.

Notice that the principle of settling obligations using a liquid and safe asset has been adopted by payment systems even before central banks were created and is adopted today in payment systems where the use of central bank reserves is not a feasible option. Where this is the case, settlement happens using the liability issued by one of the participating banks, typically the one bank that is most highly reputed, financially solid, and sufficiently large, which acts as a central institution providing clearinghouse and settlement services to all other banks participating in the system. All these other banks must hold enough balances of such liability for them to be able to fulfill their settlement obligations in due time. Loans may be extended by the central institution to participants that need liquidity to complete settlement, but eventually all such loans must be retired and settled in the liability issued by the central institution and which no other bank can issue (as it would not be acceptable to the others as settlement asset).⁵

Where settlement takes place in central bank money (reserves), the reserves that banks need to raise ("funding") for settlement purposes consist of:

- i. Cash reserves and reserves deposited with the central bank (held in response to regulatory requirements, for payment and cash-out purposes, or as part of general pool of liquid resources);

⁵ This solution was the norm in the clearinghouse arrangements adopted in the US and UK in the past centuries (Norman et al., 2011), and is still today the case for several payment systems especially for cross-border transactions (CPSS, 2013).

- ii. Reserves from settlement of incoming payments or transfers from other banks;
- iii. Borrowings from the interbank market;
- iv. Borrowings from the central bank;
- v. Unencumbered assets in the balance sheet that can be liquidated immediately; and
- vi. New deposits of cash from (old and new) customers (new noncash deposits may consist only of deposits transferred from other banks, which as such fall under item ii).

Thus, while banks may in principle create all the money the economy is willing to absorb, in practice their money creation power (or, more precisely, their ability to support the payments activity associated with their money creation) is constrained by their own ability to raise the required funding and the cost of doing so.

True, as many authors correctly remark (see Section 2), reserves are not a binding constraint on lending in so far as central banks do not choose to determine a specific level of reserves that bring about the desired short-term interest rate, which then ration bank demand for reserves at that rate. Instead, central banks set a policy interest rate and at that rate stand ready to supply all the reserves that banks demand – the supply of reserves is thus endogenous. However, although the supply of reserves accommodates fully the demand from banks, banks do not have an infinitely elastic capacity to acquire reserves.

The capacity of each bank to raise funding (reserves) is constrained by two scale factors: i) the size of the bank itself, as proxied by its share of the payment market(s) where it operates and ii) the bank's level of capital (i.e., equity plus accumulated reserves). A bank's share of the payment market(s) where it operates reflects its capacity to receive funding through incoming payments from other banks, which should on balance be adequate for settling outgoing payments. The level of capital of a bank, on the other hand, reflects (*ceteris paribus*) its capacity to access (borrow) external funding to ensure settlement of its obligations under adverse contingencies, such as temporary liquidity shortages.

The capacity of a bank to satisfy its funding needs acts as a constraint on the bank's power to create money by lending or selling deposit claims. At the margin, the cost of funding for each bank may exceed the expected return from creating money or may even become prohibitive. A bank's idiosyncratic decision to, say, double at once its total lending by issuing new deposits would run against the bank's capacity to support the new lending level with adequate funding. The bank would always be able – in principle – to create money by signing off new loans with a stroke of a pen (or, in our times, with a click of a mouse); yet, the new volume of money created would not be sustainable under the bank's normal funding structure. As most of the payments generated by the new loans would go to payees holding accounts with other banks, the bank would be unable to settle its obligations if it were unable to raise the needed funds. The bank would not be able to expand at once its deposit base as needed or to capture a larger payments market share overnight, it would not have all the collateral needed to borrow from the central bank and borrowing so much extra funds from the market might be perceived as too risky by depositors and other lenders or investors, vis-à-vis its capital level. Even a deep money market and a fully accommodating central bank's monetary policy stance would not help, if the bank's funding structure and capital were unchanged: the marginal cost of the extra funding could actually become infinite for the bank.

The conclusion is that a bank can create money by lending or selling deposits only within an overall resource envelope that is determined by the bank's capacity to raise funds (reserves) at costs that are consistent with the sustainability of its business as a going concern. Banks do create money but they can do so within the limits of their funding plan, which must take into account their in- and out-payments (as adjusted for growth and risk factors), their borrowing capacity, and the cost of raising and holding funds vis-à-vis the prospective return from lending or investing the money created.⁶ It

⁶ Setting aside ethical considerations, not even gimmicks used by banks to raise capital fictitiously (see Werner, 2016, for the case of Credit Suisse) would do the trick, precisely because of the fictitious nature of the extra capital raised. Gimmicks would work only if a bank used the extra (fictitious) capital to (falsely) project the image of a larger and financially healthier institution and, hence, expand its depositor base. The additional

doesn't matter whether the funding is in place prior or after the act of money creation: creating deposit claims, *per se*, does not require any funding, and central bank reserves (and even less deposits) are not what banks lend out when they extend credit to their clients. What matters, though, is that i) funding is in place when the bank's new financial obligations to other banks (or other payment system participants) fall due and are to be settled, and that ii) funding can be obtained without impairing the bank's financial equilibrium and business objectives.

Within the planned funding envelop and its associated costs, the bank extends its loans until the marginal returns on lending and investing, account taken of the risks and administrative costs involved, does not exceed the marginal cost to the bank of raising and holding funds (Tobin, 1963). On this very price mechanism rests ultimately the power of central banks to govern monetary conditions in contemporary economies, where sophisticated markets make the use of direct control instruments (e.g., required reserve ratios or credit ceilings) inefficient and ineffective. Central banks operate indirectly by changing the policy interest rates and by publicly communicating their policy strategy in view of achieving their low-inflation objective. These indirect instruments are intended to affect the current and expected costs of funding to the banks, thus influencing their risk-return prospects and ultimately their money creation decisions. Other regulatory tools (such as capital, liquidity, and leverage ratios) may act as constraints on the overall power of banks to create money; they do not intervene directly on deposit issuances but influence the cost factors that banks consider when extending new credit. Importantly, also, they link the money creation power of each bank to the size of the bank, as discussed next.

It must be noted, however, that in a context where most or all banks were to decide conjointly to expand their lending activity, the funding constrain of each bank would be simultaneously relieved

funding only, not the larger capital *per se* (which is fictitious), would enable the bank to support a larger payments activity and, thus, to expand its money creation.

as payments in the system would (be expected to) increase and, hence, generate larger funding flows from incoming payments triggered by new loans: this would by itself enhance the banks' money creation power (all else equal). Such circumstance would be typical of expansionary phases of the business or financial cycle, where positive anticipations of future higher activity levels, supported by an accommodative stance of monetary policy, would encourage and coordinate individual bank decisions toward more lending. Obviously, the opposite would happen during contractionary phases and in the presence of central bank signals of monetary policy tightening.

Importantly, the power of each bank to issue money varies with its relative size in the payments market, as shown next.

3.1 Bank money creation and bank's size

A growing share of the payments market allows a bank to economize on the funding necessary to support money creation. As the bank's size grows, the bank needs less additional funding as a proportion of the new money created.

In a fractional reserve regime – broadly defined as any reserve regime where banks are not required to fully back their demand deposit liabilities with central bank reserves – banks typically hold only a fraction of reserves against their total demand deposit liabilities out of which payments take place, and the volume of reserves they use for settling interbank obligations and cash withdrawals from customers are only a fraction of the total transactions settled. Since raising reserves bear a cost, banks seek to optimize the balance of reserves (as a fraction of their deposits liabilities) to be held for settlement purposes. As a bank's share of the payments market(s) grows, the bank can reduce its optimal reserve balances. For instance, as the bank's share of incoming payment grows, the bank economizes on the cost of raising funds, and as the bank expands and diversifies its depositor base, it can increase the volume and value of payments that it settles on its own books ("on us" payments) and can better stabilize its funding due the higher asynchrony of deposit withdrawals (at least in

normal times). Consolidation generates increasing returns to banks, enabling them to create money (by lending or selling deposits) with lower reserve margins needed for coverage (*ceteris paribus*).⁷

The following stylized and simplified model shows how a bank's payments activity affects its money creation power as reflected by the extra profit that the bank can extract from the economy through it.⁸ The model is kept deliberately simple for the purpose of identifying the root factors originating the seigniorage associated with bank money creation.

Bank *B* operates as a profit-maximizing firm in a fractional reserve regime. It is authorized by regulation both to offer its customers demand deposit (DD_B) accounts earning interest i_{DD} and to lend money to borrowers by crediting their accounts with funds L_B (and by issuing to them equivalent deposit claims) at interest i_L , where for simplicity but with no loss of generality, the bank faces no costs other than deposit remuneration and the cost of raising central bank reserves RES needed to settle its payment obligations.

Bank *B*'s payments market share is proxied by $\sigma_B = \frac{DD_B}{DD}$, the ratio of the bank's demand deposit liabilities over total demand deposits outstanding in the system, DD .⁹ Scale effects are such that at zero-dimensional share, $\sigma_B = 0$, bank *B* operates in a perfectly competitive environment and enjoys

⁷ For evidence on economies of scale in banking, see Anolli et al. (2015) and Wheelock and Wilson (2015), and the ample literature therein referred, which show that banks (especially large ones) operate under increasing returns to scale in production. Also, value-maximizing banks benefit from "systemic scale economies" effects, whereby intermediaries operating in large systems are expected to have lower costs of production, risk absorption, and reputation signaling than intermediaries operating in small systems (Bossone and Lee, 2004).

⁸ While, in reality, banks provide a variety of services in addition to lending (e.g., account management, payments, investments, financial advisory, wealth management, etc.), typically against the charge of fees and commissions from customers, in the simplified model used in this section, banks only accept deposits from customers and lend money to borrowers in the form of newly issued deposits.

⁹ The payments market share could obviously be proxied by alternative and more accurate indicators; however, this issue does not need to detain us here.

no economies of scale in liabilities management, whereas at maximum share, $\sigma_B = 1$, the bank enjoys the largest economies of scale in liabilities management possible, is a monopolist in the lending market and a monopsonist in the deposit market.

Bank B 's funding cost structure, thus, is given by

$$[(1 - \mu_B)i_{DD} + \mu_B i_{RES}]DD_B = \phi_{DD}DD_B,$$

Bank B 's cost structure, thus, is given by

$$[(1 - \mu_B)i_{DD} + \mu_B i_{RES}]DD_B = \phi_{DD}DD_B,$$

where $\mu_B = \frac{RES_B}{DD_B}$, with $0 \leq \mu_B \leq 1$, where i_{RES} is the interest rate at which the central bank lends reserves and is a policy instrument controlled by the central bank, and where $\mu_B = 0$ for $\sigma_B = 1$ and $\mu_B = 1$ for $\sigma_B = 0$, that is, in principle parameter μ_B may vary from 0 to 1, taking the extreme value of zero in the (hypothetical) case a cashless economy with only one banks, where all payments are all "on us" and their settlement does not require central reserves, and taking the extreme value of 1 in a perfectly competitive economy with atomistic banks. Parameter μ_B could also be set by central bank regulation or else it could be set by each bank as part of its liquidity management policy. In this last case, and assuming a fractional reserve regime, the parameter would also be a function of the structure, rules, and procedures of the payment system(s) where each bank operates.

Note that $\phi_{DD} = \phi_{DD}(i_{DD}, \mu_B, i_{RES})$ is the cost of raising and holding reserves RES as a fraction of demand deposit liabilities and includes both the remuneration of bank B 's demand deposit liabilities and the cost of accessing reserves through alternative sources. At i_{RES} , the central bank supplies all the reserves that banks demand.

One may therefore posit:

$$i. \quad i_L = i_L(\sigma_B), \quad \text{with } \frac{\partial i_L}{\partial \sigma_B} > 0 \text{ and } i_L = i_L^* \text{ for } \sigma_B = 0,$$

where i_L^* is the equilibrium loan interest rate under perfect competition,

$$\text{ii. } i_{DD} = i_{DD}(\sigma_B), \quad \text{with } \frac{\partial i_{DD}}{\partial \sigma_B} < 0 \text{ and } i_{DD} = i_{DD}^* \text{ for } \sigma_B = 0,$$

where i_{DD}^* is the equilibrium deposit interest rate under perfect competition, and

$$\text{iii. } \mu_B = \mu(\sigma_B), \quad \text{with } \mu' > 0 \text{ and } \begin{cases} \lim_{\sigma \rightarrow 0} \mu_B = 1 \\ \lim_{\sigma \rightarrow 1} \mu_B = 0 \end{cases}$$

Item iii. implies that, as the bank's payments market share grows, the bank economizes on the reserves held at the central bank.

Bank B's loan supply function is

$$L_B^S = L_B^S(i_L - \phi_{DD}) = L_B^S(i_L, \phi_{DD}(i_{DD})) \leq L_B^S(\rho) \text{ for given } \mu_B \text{ and } i_{RES},$$

$$\text{with } i_L > \phi_{DD} \text{ and } \begin{cases} L_B^{S'} > 0 \text{ for } i_L \leq \rho \\ L_B^{S'} = 0 \text{ for } i_L > \rho \end{cases}$$

where ρ is the critical credit risk threshold, beyond which bank B no longer lends (i.e., the bank rations the supply of loans), and where the demand functions for bank B 's loans is

$$L_B^D = L_B^D(i_L) \quad \text{with } L_B^{D'} < 0,$$

On the deposits market side, while the supply is determined by the supply of loans, the demand for bank B ' deposits is

$$DD_B^D = DD_B^D(i_{DD}) \quad \text{with } DD_B^{D'} > 0$$

Under the above conditions, at zero-dimensional market share ($\sigma_B = 0$), that is, with perfect competition, equilibrium attains at

$$L_B^* = \min_L [L_B^*(i_L^*, i_{DD}^*), L_B^S(\rho)]$$

where bank B 's maximum profit is normal profit

$$Max \Pi_B |_{\sigma_B=0} = i_L^* L_B^*(i_L^*, i_{DD}^*) - \phi_{DD}(i_{DD}^*) DD_B^*(i_{DD}^*)$$

for given μ_B and i_{RES} and $i_L^* \leq \rho$. Call π_B^* the unit normal profit rate, at which bank B extracts no seigniorage. As the market share of bank B 's increases, the bank enjoys larger scale economies and greater monopolistic/monopsonistic powers; as a result, its profit grows above normal (that is, zero extra profit),

$$\text{Max } \Pi_B |_{\sigma_B=1} > \text{Max } \Pi_B |_{1 > \sigma_B > 0} > \text{Max } \Pi_B |_{\sigma_B=0},$$

and its unit seigniorage rent increases $s_B \equiv \pi_B - \pi_B^* > 0$. Notice that s does not reflect any value added created for the economy by bank B ; it amounts to pure extraction of net resources from the economy by the bank and, unlike the seigniorage earned by the central bank (which is partly returned to the state budget after covering for the central bank's expenses and profit), commercial bank seigniorage is owned by the bank (net of any tax imposed by the state on the banks' income) and retained by it and or paid out to its shareholders.

Notice that, in the real world, commercial banks provide various types of services in addition to lending (e.g., for account management and for payments), typically against the charge of fees and commissions from customers. In this article, for reasons of simplicity, but without loss of generality for the results derived in it, it is assumed that banks exclusively accept deposits from customers and lend money to borrowers in the form of newly issued deposits; no other activity is involved.

3.2 Payment settlement rules and technologies

While the payments market share of individual banks affects, as discussed, each bank's optimal demand for reserves at the margin, the rules and technologies for settling payments determine the structural demand for reserves of the whole system. Payment system rules determine such demand via two channels: the settlement modality (typically netting or gross) and the technology adopted for transferring value at settlement. Each of the two modalities has drastically different effects on the banks' demand for reserves, with netting requiring much less reserves than the gross modality (at the cost of higher settlement risk). Similarly, by re-introducing elements of netting into the gross modality, contemporary payment systems' automated technologies (such as liquidity-saving mechanisms in hybrid real-time gross settlement systems) allow banks to economize on their use of reserves for any given value of payments to be settled and increase the velocity of reserves circulating

in the systems. Both factors interact with payment system scale economies and affect the money creation power of funds.¹⁰

As an exemplification of the above factors, let us take the hypothetical case of a cashless economy with a fully consolidated banking system where all agents hold accounts with only one bank. In this case, all payments and money transfers would be "on us" for the single bank, that is, they would all take place on the books of the single bank. In such case, the bank would need no reserves for settling transactions; it would be under no debt obligation towards its customers, and it might thus create all the money the economy would be willing to absorb, at a given interest rate, without having to hold (costly) reserves. The single bank's deposit liabilities would become the accepted instrument to settle debts within the economy and would *de facto* be irredeemable, much as legal tender is in any economy.

Conversely, in the opposite extreme case of an atomistic banking sector (with zero-dimensional banks), every single bank could still issue new deposits as loans to borrowers; yet, they would have to cover all new deposits with central bank reserves and their money creation capacity would *de facto* be nihil. In such a system, the fractional reserve regime would *de facto* become a 100% reserve regime – or a “narrow banking” regime – where every deposit unit created and loaned out by a bank would have to be funded with an equivalent unit of central bank reserves. This is because, from the moment the loan proceeds are withdrawn by the borrower from the atomistic bank and used to make payments, funds would have to be transferred to payees holding accounts with other atomistic banks and would have to be settled in central bank reserves.

In the real world, systems are populated by a plurality of banks; yet, in highly consolidated systems, the largest banks (and the whole systems) feature economies of scale in the use of reserves, as

¹⁰ Significant scale economies in payment systems are found by Beijnen and Bolt (2007) and Humphrey (2009).

discussed. Consolidation enhances the banks' power to expand their balance sheet by deposit creation while economizing on the funding (reserves) needed to back the deposits and to support the interbank payments activity associated with the mobilization or redemption of the newly deposits created.

It is important to understand bank money creation in the context of different types of payment settlement system technologies and rules. Appendix 1 develops a simple example. It assumes that settlement of interbank payments takes place in central bank money (reserves). Since holding reserves carries an opportunity cost for banks (unless such reserves are remunerated at the same level as their best possible alternative option), the cost from holding reserves affects banks' money creation power. As different payment settlement system rules require banks to mobilize different volumes of reserves (all else being equal), each type of system rules bears a different impact on bank money creation.

4. BANK MONEY CREATION AND "SEIGNIORAGE"

When a nonbank financial intermediary intermediates funds, say, by accepting funds from savers against the issuance of term deposits or securities and by onlending such funds as, say, mortgages to homebuyers, the intermediary typically receive the funds in an account that it holds with a bank and transfers those funds from that account to the bank accounts of the homebuyers. In such a case of pure financial intermediation, there is no money creation and the funds onlent to the homebuyers are the same funds that were originally saved by the intermediary's customers. As said, both the intermediary and its borrowers hold deposit accounts with banks and all fund transfers in the example (i.e., from the savers to the intermediary, and from the intermediary to the homebuyers) take place across deposit accounts held with banks.

The crucial difference between banks and pure financial intermediaries, as explained by Werner (2014) and discussed above, is that only banks (not the intermediaries) may – by regulation – credit their borrowers' bank deposit accounts with newly issued deposit claims. It is this special account structure, permitted by regulation, that confers on the banks the power to "create" money by allowing them to originate new assets (loans) and equivalent new liabilities (deposits).

Thus, money creation by lending is unique to banking, in as much as banks only are allowed by regulation to i) hold customers' demand deposits and issue loans to customers in the form of demand deposits and ii) carry both operations through customer accounts held in their own books. On the other hand, these features are not shared by pure financial intermediaries, which may a) intermediate only pre-existing funds and b) perform lending functions only by ordering fund transfers to take place across accounts held with banks. Unlike banks, pure intermediaries *may not, and cannot*, create money by issuing new deposit liabilities and originating corresponding loan assets.

This being the case, whenever money is created by fiat and lent or sold to the economy in exchange for real resources, a special type of rent – seigniorage – is extracted from the economy by the money creator. This is the unique power that the money creator holds and exercises, in that it can generate exchange value *ex nihilo* and can thus exert a degree of command over real resources that it has not contributed to produce.

It is of interest to analyze how seigniorage originates from bank money creation and how it relates to profits from pure financial intermediation.

4.1 Seigniorage from bank money creation and profit from financial intermediation

A stripped-to-the-bone model is used in this section to address this issue. The model builds on two financial institutions, a bank and a pure intermediary that differ from each other precisely, and only, for the money-creation power that characterizes the bank only. The two intermediaries are otherwise identical to each other.

Call the two financial institutions bank B and pure intermediary I . Bank B is the same as in Section 3.1. Pure intermediary I is authorized to offer its customers (non-demand) term deposit (TD_I) accounts bearing interest i_{TD} and to lend long-term funds L_I at interest i_I . Both intermediaries operate in identical quasi-monopolistic market environments and at the same scale, $1 > \sigma_B = \sigma_I > 0$.

For purposes of comparison, the following additional assumptions apply. The two entities feature identical administrative and operational cost structures (including, inter alia, such items as the costs for screening, selecting, and monitoring borrowers, and those to enforce contract terms), which for simplicity are here set to zero. They charge on loans the same net-of-risk interest rate ($i_L = i_I$). Credit rationing is assumed away.

As in Section 3.1, bank B maximizes profits

$$\Pi_B|_{1 \geq \sigma_B \geq 0} = i_L L_B(i_L, i_{DD}) - \phi_{DD}(i_{DD}) D D_B(i_{DD}),$$

and pure intermediary I maximizes profits

$$\Pi_I|_{1 \geq \sigma_I \geq 0} = i_{TD} L_I(i_I, i_{TD}) - i_{TD} T D_I(i_{TD})$$

where $\Pi_{B,I}|_{\sigma_I=0}$ sets the normal (zero excess) profit of the two intermediaries under perfect competition.

Now, looking at bank B 's funding cost structure, notice that RES consists of reserves from incoming payments ($RINP$) and borrowed reserves (BR), which include reserves borrowed (wholesale) from other banks (RW) and those borrowed from the central bank (RCB). Bank B 's funding cost reflects also the cost of holding collateral (safe and liquid) assets against potential cash borrowing needs.

Knowing that $i_{DD} < i_{TD}$, and noting that

$$i_{RES} = \frac{i_{RINP} RINP + i_{BR} BR}{RES}$$

where $i_{RINP} = 0 < i_{TD}$, and $i_{BR} = \frac{i_{RW} RW + i_{CB} RCB}{BR} < i_{TD}$, since the cost of borrowing reserves from the interbank market or from the central bank is lower than the interest rate on term deposits, the following strict inequality holds:

$$\phi_{DD} = (1 - \mu_B) i_{DD} + \mu_B i_{RES} \ll i_{TD},$$

that is, all else equal, bank B 's cost of funding is lower than the cost of managing liabilities for pure intermediary I . Moreover, since both institutions are profit maximizers and the marginal cost of

lending is lower for bank B than for pure intermediary I , the former will always be able to create money and expand equilibrium lending beyond what is feasible for the pure intermediary, that is, $L_B^* \gg L_I^*$, where L_B^* is the volume of loans that the economy is willing to borrow at i_L^* , where $i_L^* < i_I^*$ are equilibrium values. Thus, it will always be the case that

$$\Pi_B|_{1>\sigma_B=\sigma_I>0} - \Pi_I|_{1>\sigma_I=\sigma_B>0} \gg 0.$$

It follows that, all else equal, money creation allows banks to extract from the economy larger rents than is possible for pure financial intermediaries, under (hypothetically) identical market conditions. Importantly, in the above hypothetical case, monopolistic/monopsonistic powers affect both intermediaries identically; thus, the extra-profit extracted by the bank (which goes into seigniorage) is due exclusively to its money creation power, which by construction is precluded to the pure financial intermediary.¹¹ This would be apparent considering, for instance, the (ceteris paribus) effect of a decrease in scale σ_B on bank B's cost structure ϕ_{DD} through parameter μ_B , and hence on interest margin $i_L - \phi_{DD}$, given i_L and ϕ_{DD} , and thus ruling out monopolistic/monopsonistic effects.

One way of setting apart the effects of a bank's economies of scale and market power over seigniorage is to consider that the former – economies of scale – affect quantities (i.e., they enhance the bank's power to create money while economizing on reserves) while the latter – market power – affects prices (i.e., it enables the bank to increase the loan-deposit interest rate spread either by acting on the

¹¹ While seigniorage (as any forms of rent) relates to the existence of (quasi) monopolistic power, it is not quite the same as rent from monopoly: the latter derives from the exclusive control that a monopolist exercises over a scarce or irreproducible resource or its price; seigniorage, on the other hand, derives from the money creator's exclusive power to generate exchange value in excess of the cost to produce and distribute it. This power does require exclusivity, as (quasi) monopoly does; but it is the exclusivity of generating new positive exchange value, not that of controlling the supply of, or access to, a pre-existing resource, or its price.

quantity of credit supplied on the supply price of credit). The two effects are necessarily intertwined, however, and seigniorage is whatever extra rents can be extracted through money creation.¹²

5. CONCLUSION

This article has reconsidered the role of banks as money creators and has investigated how the role that banks play in the payment system space affects their money creation process. The article has analyzed how banks' payment obligations and the rules governing their settlement determine (and constrain) the extent to which each bank can exercise the power to create money. In particular the article has studied how the payments market share of each bank affects its money creation power and how payment settlement technologies and rules determine banks' demand for funding and, hence, their money creation power. Also, as the power to create money enables money creators to extract extra-profits or rents ("seigniorage") from the economy, the article has evaluated analytically how banks extract seigniorage through money creation and how bank seigniorage differs from profits from pure financial intermediation.

By showing the central role that payment systems play in the context of such an important economics topic as money creation, the article aims to emphasize the relevance of payment system analysis for macroeconomic theory and practice, and points to the need for achieving better integration of the two disciplines. In future research, especially in light of the fast evolution of technology as it applies to finance, and notably to payment systems and services, it will be important to analyze how the introduction of new digital currencies (including from central banks) and the entry into the payments market of new digital currency issuers might affect the funding structure of banks and, hence, their money creation power and credit supply to the economy.

¹² On the effects of bank seigniorage on the aggregate output and the resource distribution, see Bossone (2021b, c), available from the author on request.

REFERENCES

- Adrian, T., and T. Mancini-Griffoli (2019), “The rise of digital money.” FinTech Notes, Note/19/01, Washington, D.C.: International Monetary Fund.
- Anolli, M., E. Beccalli, and G. Borello (2015), “Are European banks too big? Evidence on economies of scale,” *Journal of Banking & Finance*, 58, 232-246.
- Beijnen, C., and W. Bolt (2007), “Size matters: economies of scale in European payments processing,” DNB Working Paper, No. 155, November.
- Bernanke, B. S. (1993), “Credit in the macroeconomy,” *FRBNY Quarterly Review*, 50–70, Spring 1992–93.
- Bernanke, B. S., and A. S. Blinder (1988), “Credit, Money, and Aggregate Demand,” *The American Economic Review*, Vol. 78, No. 2, Papers and Proceedings of the One- Hundredth Annual Meeting of the American Economic Association (May), 435-439.
- Bernanke, B. S., and M. Gertler (1995), “Inside the Black Box: The Credit Channel of Monetary Policy Transmission,” *Journal of Economic Perspectives*, American Economic Association, vol. 9(4), 27-48, Fall.
- Bossone, B. (2021a), “Money and customer funds in the world of digital finance: Who really owns what?,” *Journal of Payment Strategy & Systems*, Volume 15, Number 1, 37-53, Spring.
- Bossone, B. (2021b), “Commercial bank seigniorage and the macroeconomy,” *International Review of Financial Analysis*, 76, 1-15.
- Bossone, B. (2021c), “
- Bossone, B., and J. K. Lee (2004), “In Finance, Size Matters,” *IMF Staff Papers*, Volume 51, Number 1, April.
- Brunnermeier, M., and Y. Sannikov (2016), "The i theory of money," mimeo.
- Casu, B., C. Girardone, and P. Molineux (2006), *Introduction to Banking*, Harlow: FT Prentice Hall, Pearson Education Ltd.
- Cecchetti, S. G. (2008), *Money, Banking and Financial Markets* (2nd ed.). New York: McGraw-Hill Irwin.
- CPSS (2013), *The role of central bank money in payment systems. report by the Committee on Payment and Settlement Systems*, Bank for International Settlements, Basel, August.
- CPSS-IOSCO (2012). *Principles for financial market infrastructures*. joint report by the Committee on Payment and Settlement Systems Technical Committee of the International Organization of Securities Commissions, Bank for International Settlements, Basel, April 2012.
- Diamond, D. W., and P. H., Dybvig (1983). “Bank Runs, Deposit Insurance, and Liquidity.” *The Journal of Political Economy*, Vol. 91, No. 3., June, 401-419.
- Fama, E. F. (1980), “Banking in Theory of Finance,” *Journal of Monetary Economics* 6(1):39-57 · January.

- Fullwiler, S. (2012), "Scott Fullwiler: Krugman's Flashing Neon Sign," *Naked Capitalism*, 2 April
- Gurley, J. G., and E. S. Shaw (1960), *Money in a Theory of Finance*, Washington, D.C.: The Brookings Institution.
- Gurley, J. G., and E. S. Shaw (1955), "Financial aspects of economic development," *American Economic Review*, 45, 515-528.
- Humphrey, D. B. (2009), "Retail Payments: Integration and Innovation Payment Scale Economies, Competition, and Pricing." ECB Working Paper Series, No. 1136, December.
- Jakab, Z., and M. Kumhof (2015). "Banks are not intermediaries of loanable funds — and why this matters," Working Paper No. 529, Bank of England, May.
- Kashyap, A., R. Rajan, and J. Stein, J. (2002), "Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking," *Journal of Finance*, 57, 33-73.
- McLeay, M., A. Radia, and R. Thomas (2014a), "Money in the Modern Economy: An Introduction." *Bank of England Quarterly Bulletin*, 54(1), 4-13.
- McLeay, M., A. Radia, and R. Thomas (2014b), "Money Creation in the Modern Economy," *Bank of England Quarterly Bulletin*, 54(1), 14-27.
- Mankiw, N. Gregory (2019), *Macroeconomics*, 10th edition. New York: Worth Publishers.
- Matthews, K., and J. Thompson (2005), *The Economics of Banking*, Chichester: John Wiley and Sons.
- Moore, B. (1979), The Endogenous Money Stock, *Journal of Post Keynesian Economics*, 2(1), 49-70.
- Moore, B. (1983), Unpacking the Post Keynesian Black Box: Bank Lending and the Money Supply, *Journal of Post Keynesian Economics*, 5(4), 537-556.
- Norman, B., R. Shaw, and G. Speight (2011), "The history of interbank settlement arrangements: exploring central banks' role in the payment system," Working Paper No. 412, Bank of England, June.
- Riordan, M. H. (1993). Chapter 11: Competition and bank performance: a theoretical perspective, in C. Mayer, and X. Vives (eds.), *Capital markets and financial intermediation*, Cambridge: Cambridge University Press.
- Realfonzo, R. (2012), Circuit Theory, in J. E. King, *The Elgar Companion to Post Keynesian Economics*, Edward Elgar, 87-92.
- Rochon, L.P., and M. Setterfield (2012). A Kaleckian model of growth and distribution with conflict-inflation and Post Keynesian nominal interest rate rules. *Journal of Post Keynesian Economics*, 34(3), 497-520.
- Rossi, S, and L. P. Rochon (ed. by) (2003) *Modern Theories of Money. The Nature and Role of Money in Capitalist Economies*, Edgar, Edward Publishing, Inc.

Rossi, R., and L. P. Rochon (eds.) (2017), *Post-Keynesian Economics*, The International Library of Critical Writings in Economics series, Edward Elgar Publishing, Cheltenham, UK.

Sealey, C., and Lindley, J. T. (1977), "Inputs, outputs and a theory of production and cost at depositary financial institutions," *Journal of Finance*, 32, 1251–1266.

Sheard, P. (2013), "Repeat After Me: Banks Cannot and Do Not "Lend Out" Reserves," Standards & Poor's Rating Services, RatingsDirect, Economic Research, 13 August.

Tobin, J. (1963), "Commercial banks as creators of 'money'," In D. Carson (ed.), *Cowles foundation paper 205, Banking and Monetary Studies*, Homewood: Irwin.

Von Mises, L. (1912), *Theorie des Geldes und der Umlaufsmittel*, Leipzig: Duncker und Humblot.

Wheelock, D. C., and P. W. Wilson (2015), "The Evolution of Scale Economies in U.S. Banking," Economic Research, Federal Reserve Bank of Saint Louis, Working Paper 2015-021C.

Werner, R. A. (2016), "A lost century in economics: Three theories of banking and the conclusive evidence," *International Review of Financial Analysis*, Volume 46, July.

Werner, R. A. (2014), "How do banks create money, and why can other firms not do the same? An explanation for the coexistence of lending and deposit-taking," *International Review of Financial Analysis*, 36 (2014), 71-77.

APPENDIX 1. BANK MONEY CREATION AND PAYMENT SYSTEM STRUCTURES AND RULES

It is important to understand bank money creation in the context of different types of payment settlement systems. In this appendix, following current practice worldwide, it is assumed that settlement of interbank payments takes place in central bank money (reserves). Since holding reserves carries an opportunity cost for banks (unless such reserves are remunerated at the same level as their best possible alternative option), the cost from holding reserves detracts from commercial bank extra profit (discussed in Appendix 1). As different payment settlement systems require banks (*ceteris paribus*) to mobilize different volumes of reserves, each type of system has a different impact on bank extra profits.

A simple numerical example makes the point. Take an economy with a central bank (CB), two banks (A, B), and four customers (a1, a2, b1, b2), where a1 and a2 are customers of bank A and b1 and b2 are customers of bank B. Banks A and B hold no reserves with CB initially. The case is analyzed where the same transactions are executed and settled in two different systems, respectively: i) a real-time gross settlement (RTGS) system and ii) a deferred netting settlement (DNS) system.

RTGS and DNS systems

In RTGS systems transfer of money or securities takes place from one bank to another on a "real time" and on a "gross" basis. Settlement in "real time" means payment transaction is not subject to any waiting period: the transactions are settled as soon as they are received, accepted and processed. Transactions are executed only if they are covered by sufficient liquidity. "Gross" settlement means that transactions are settled on a one to one basis, without bundling or netting them against any other transactions. Once executed, payments are settled with "finality" (funds transfers are irrevocable and unconditional). In DNS systems, credit and debit transactions are recorded on the clearinghouse books throughout the business day, and at the end of the operating day (in case of one netting cycle) or at the end of each netting cycle (in case of multiple cycles a day), the clearinghouse calculates the total "net" multilateral balances of all transactions and determines what each bank owes to/is due from all others. Settlement of such balances then takes place with finality.

Reserve management in RTGS systems

The following sequence takes place:

- A lends \$100 to a1 (and creates an equal amount of deposits)
- a1 pays \$100 to b1
- A borrows \$100 from CB and transfers \$100 to B: interbank settlement is done
- B lends \$150 to b2 (and creates an equal amount of deposits)
- b2 pays \$150 to a2
- B receives \$100 from A (item 3 above) and borrows \$50 from CB
- B pays \$150 to A: interbank settlement is done
- A repays \$100 to CB
- B borrows \$50 from A and repays \$50 to CB, or
- B renews its loan from CB
- At the end of the sequence, CB has issued \$150 intraday.

Under item 9 all \$150 reserves go to zero by end-of-day as A and B repay their loans to CB: no net central bank money (reserves) creation. Yet \$250 commercial bank money (demand deposits) were created, which supported \$250 worth of transactions.

Under alternative item 10, the \$50 intraday loan converts into an overnight loan and becomes money supply: \$50 net reserves creation. But, as above, \$250 demand deposits were created by banks, which supported \$250 worth of transactions.

Reserve management in DNS systems

Assuming the same transactions take place in a NSS, it can be immediately seen here that the end-of-day balance to be settled between A and B is \$50, which B borrows from CB. Although this resembles item 10 above, netting in fact saves borrowing costs overall, since under RTGS total borrowing from CB is \$150 while under netting it is only \$50. Deposit creation and transactions are as above.

In both cases, commercial bank money has been created in excess of central bank money and has supported a correspondingly larger volume of (real and financial) transactions. As the demands for converting deposits into cash are limited, the greatest share of deposits created in excess of reserves remain in the system as such: this large mass of deposits outstanding (net of all conversions, destructions and new deposit creations) earn an ongoing flow of net revenues. These net revenues originate from the banks' exclusive power to create money (in excess of central bank money) and represent their extra profits.

Money creation, reserve requirements, and extra profits

Bank extra profits range between a maximum value when money issuance is not constrained by the demand for reserves and a minimum of zero when money issuance is subjected to full reserves requirements (e.g., currency board, 100% reserve or narrow banking). In all intermediate cases, profits are above zero.