

Banking: A New Monetarist Approach*

Chao Gu

University of Missouri-Columbia

Fabrizio Mattesini

University of Rome Tor-Vergata

Cyril Monnet

University of Bern and
Study Center Gerzensee

Randall Wright

University of Wisconsin - Madison and
Federal Reserve Bank of Minneapolis

April 30, 2012

Abstract

We develop a theory of banking with two features: (i) banks take deposits and make investments on behalf of depositors; (ii) bank liabilities (claims on deposits) facilitate third-party transactions. Existing models have (i) or (ii), not both, even though they are intimately connected: both originate from limited commitment. We describe an environment, characterize feasible and efficient allocations, and interpret the outcomes as banking arrangements. Banks are essential: without them, the set of feasible allocations is inferior. We show that it can be efficient to sacrifice investment returns in favor of more trustworthy demand deposits. We identify characteristics making for good bankers, and confront these predictions with economic history.

*We thank many colleagues for comments and discussions, especially Steve Williamson, Erwan Quintin, Chris Phelan, V.V. Chari, Ken Burdett and the referees. We also thank the Editor, Bruno Biais, for his generous input, including an outline for the Introduction that we followed closely (although, of course, we remain responsible for the execution). Wright thanks the NSF and the Ray Zemon Chair in Liquid Assets at the Wisconsin School of Business for research support. The views expressed here are our own and do not necessarily reflect those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

This paper develops a theory of banking capturing two features: (i) banks accept deposits and make investments on behalf of depositors; (ii) bank liabilities, claims on said deposits, facilitate exchange with other parties. While some models of banking (see below) address either the first or second, it is desirable to have a framework that incorporates both, because the two activities are connected in a fundamental way: as we show, both originate from limited commitment. Of course, banks may do more, such as providing liquidity insurance or information processing. We downplay these functions, which have been studied extensively elsewhere, and focus instead on banks arising endogenously as a response to commitment problems. Commitment issues are central because banking concerns the intertemporal allocations of resources, which hinges on the incentive to make good on one's obligations. Banks are agents that are trustworthy, in the sense that they have strong incentives to honor their promises, and this allows claims on deposits to serve as a means of payment, or inside money.

Our formal model incorporates the following ingredients. There are two types of infinitely-lived agents. Each period in discrete time is divided in two subperiods. Type 1 wants to consume in the first, while type 2 wants to consume in the second subperiod. Type 1 can produce and invest in the first subperiod, thus generating second-subperiod output valued by type 2. In a first-best world, it would be efficient to have type 2 lend to type 1, enabling him to consume and invest in the first subperiod, with type 2 consuming the product of the investment in the second. In the second subperiod, however, type 1 is tempted to abscond with the proceeds instead of paying off type 2 (as in the cash-diversion models of, e.g. DeMarzo and Fishman 2007 or Biais et al. 2007). This is more of a problem to the extent type 1 is impatient, has better opportunities to divert the proceeds of his investments, or has a smaller probability of getting caught. In general, we need to impose a repayment constraint guaranteeing type 1 does not behave opportunistically, which can hinder the ability to exploit intertemporal gains from trade.

We then introduce a different agent, who is like the first type but may have a lower incentive to behave opportunistically or a higher probability of getting caught. Even if this third agent is less efficient than type 1 at producing second-subperiod output, when the incentive problem is severe the following scheme is efficient: type 1 works in the first subperiod and deposits the output with the third agent, who invests on his behalf. Since type 2 knows the third agent is more inclined to deliver, he is willing to produce more in the first subperiod for type 1. This resembles banking: type 1 deposits resources with, and delegates investment to, his banker; and claims on these deposits facilitate transactions between 1 and 2. These liabilities constitute inside money, a role played historically by banknotes, and later checks and debit cards, backed by demand deposits. This arrangement allows type 1 to get more from type 2, compared to pure credit, because the bank is more trustworthy than the type 1 agent. Again, this can be true even if the bank does not have access to the highest-return investment opportunities. The function of the theory is to formalize the above ideas, and then put the model to use in several applications.¹

Before getting into substantive economic results, we briefly mention method. We proceed with minimal prior assumptions about who banks are or what they do. The agents that become bankers here are not fundamentally different from depositors (e.g., in terms of preferences), although they may have less of a problem dealing with certain frictions, like imperfect monitoring and commitment. Obviously, some frictions are needed, since models like Arrow-Debreu have no role for banks, or any other institution whose purpose is to facilitate the process of exchange. The simplest such institution is money, and a classic challenge in monetary economics is to ask what frictions make money essential in the following sense: it is said to be essential when the set of feasible allocations is bigger, or better, with money than without

¹The applications are previewed below, but we mention here that they focus on issues other than ones that have been studied extensively elsewhere, e.g. banks' tendency to borrow short and lend long, or to make deposits available on demand except in unusual circumstances like suspensions. Those *could* be analyzed in our framework, too, but we prefer to concentrate on more novel results concerning the use of claims on deposits to facilitate exchange in situations where direct credit is imperfect.

it (Wallace 2010). We similarly want to know when banks are essential. Here the planner, or mechanism, instructs some agents to perform certain functions resembling salient elements of banking, and this activity is essential: if it were ruled out the set of incentive-feasible allocations would be inferior.²

We formalize the idea that agents are better suited to banking (accepting and investing deposits) when they have a good combination of the following characteristics that make them more trustworthy (less inclined to renege on obligations):

- they are relatively patient;
- they are more visible, by which we mean more easily monitored;
- they have a greater stake in, or connection to, the economic system;
- they have access to better investment opportunities;
- they derive lower payoffs from diverting resources for strategic reasons.

Some of these are obvious, e.g. patience is good for incentives. Others seem less so, e.g. it may be good to delegate investments to those with a greater stake in the system, even if they have less lucrative investment opportunities, since this can facilitate transactions with other parties.

In terms of the literature, Gorton-Winton (2002) and Freixas-Rochet (2008) provide surveys. Much of this work is based on information frictions, including adverse selection, moral hazard and costly state verification. One strand, originating with

²In the working paper (Mattesini et al. 2009), we go into more detail concerning the way our approach relates to the one advocated by Townsend (1987,1988). This method first lays out an environment, including frictions (e.g. information or commitment problems), and then tries to interpret outcomes (e.g. incentive-feasible or efficient allocations) in terms of institutions one observes in actual economies. We want to know which frictions lead to a role for banking. For this question, one cannot *assume* missing markets, incomplete contracts etc., although something like that may *emerge*: “the theory should explain why markets sometimes exist and sometimes do not, so that economic organisation falls out in the solution to the mechanism design problem” (Townsend 1988). Relatedly, we adhere to a generalization of Wallace’s (1996) dictum: “money should not be a primitive in monetary theory – in the same way that a firm should not be a primitive in industrial organization theory or a bond a primitive in finance.” By extension, banks should not be a primitive in banking theory; they should arise endogenously. See also Araujo and Minetti (2011).

Diamond and Dybvig (1983), interprets banks as coalitions providing liquidity insurance.³ Another approach, pioneered by Leland and Pyle (1977) and developed by Boyd and Prescott (1986), interprets banks as information-sharing coalitions. Yet another, following Diamond (1984) and Williamson (1986,1987), interprets banks as delegated monitors taking advantage of returns to scale. Relative to these theories, although monitoring is also part of the story, we focus more on commitment problems. Rajan (1998) previously criticized banking theory for not concentrating more on incomplete contracts, or markets, based on limited enforcement.⁴ We agree that commitment/enforcement issues are central, but we think this needs to be endogenous. In this regard, we build on Kehoe and Levine (1993) and Alvarez and Jermann (2000), although the application is new. We also highlight Cavalcanti and Wallace (1999a,1999b), where inside money also facilitates trade.⁵ However, that model does not have deposits, delegated investments, loans or endogenous monitoring. We provide an integrated approach, capturing these features as well as the role of bank liabilities in the exchange process.

As a preview and summary, the rest of the paper is organized as follows. Section 2 describes the basic environment. Section 3 discusses incentive-feasible and efficient allocations when there is a single group, consisting of two types, as described above. This provides a simple model of credit with imperfect commitment, monitoring and collateral, but no banks. In Section 4 we consider two groups, and show it can be efficient for some agents in one to act as bankers for the other group. Section 5 shows how to implement efficient allocations using the liabilities of bankers as money. Section 6 takes up various extensions of the baseline model, to study in more detail which agents should become bankers, how many we should have and how big they should be, how we should monitor them when it is costly, what is the tradeoff between trustworthiness and rate of return, and why intermediated lending can be useful. Most

³See Ennis-Keister (2009) for a recent contribution with references.

⁴See also Calomiris-Kahn (1991), Myers-Rajan (1998) and Diamond-Rajan (2001).

⁵See also Sanches-Williamson (2010), Koepl et al. (2008), Andolfatto-Nosal (2009), Huangfu-Sun (2008), He et al. (2005,2008), Wallace (2005) and Mills (2008).

of the analysis focuses on stationary allocations, which can be a binding restriction. Hence, Section 7 relaxes this by considering efficient dynamic allocations, and shows the key economic results survive. Section 8 briefly compares the model to some facts from banking history. Section 9 concludes.

2 The Environment

Time is discrete and continues forever. There are two *groups*, a and b , each with a $[0, 1]$ measure of agents. In a group, agents can be one of two *types*, $j = 1, 2$.⁶ We refer to agents of type j in group i as j^i (e.g., 1^a is a type 1 agent in group a). Each period, each agent can be active or inactive. Inactive agents do not consume or produce, and get utility normalized to 0, in that period. Agents in group i are active with probability γ^i and inactive with probability $1 - \gamma^i$, where γ^i can differ across groups, so that they can have different degrees of connection to the economic system. In each period there are two group-specific goods, 1^i and 2^i , $i = a, b$. What defines a group is that agents have utility functions with only goods produced in their group as arguments. Active agents 1^i consume good 1^i and produce good 2^i , while active agents 2^i consume good 2^i and produce good 1^i . Letting x_j and y_j denote consumption and production by agents j , utility $U^j(x_j, y_j)$ is increasing in x_j , decreasing in y_j and satisfies the usual differentiability and curvature conditions. We assume $U^j(0, 0) = 0$, normal goods, and a discount factor across periods $\beta \in (0, 1)$.

Each period is divided in two subperiods, and good j must be consumed in subperiod j . Thus, type 1^i agents must consume before 2^i , which makes credit necessary. To have a notion of collateral, good 2^i is produced in the first subperiod, and invested by either type 1^a or 1^b , with fixed gross return ρ^a or ρ^b in terms of second-subperiod goods (there is no investment across periods, only across subperiods). This may be as simple as pure storage, perhaps for safekeeping, or any other investment opportunity; merely for ease of presentation do we assume a fixed return. Also, type 2^i

⁶Types are permanent. The main results go through when agents are randomly assigned types each period, but the analysis is messier (Mattesini et. al. 2009).

agents cannot invest for themselves (all we really need is that they cannot invest as efficiently, to generate gains from trade), while a type 1^i agent can invest good goods produced in his own group or the other group. An important friction is this: when type 1^i agents are supposed to deliver the goods, in the second subperiod, they can renege to obtain a payoff λ^i per unit of diverted resources, over and above $U^1(x_1, y_1)$. If $\lambda^i = 0$, investment constitutes perfect collateral, since type 1^i has no gain from renegeing when the production cost is sunk. However, if $\lambda^i > 0$, there is an opportunity cost to delivering the goods.

Formally, diversion can be interpreted as type 1 consuming the investment returns, but it stands in for the more general idea that investors can divert resources opportunistically. To be clear, when we say utility for agents is defined only over goods produced in their own group, we mean U^i is only a function of these goods. Type 1^i also gets a payoff $\lambda^i y$, over and above U^i , from absconding with y units of the proceeds of investments of goods from either group. This is the key incentive issue in our model.⁷ We assume $U^1(x_1, y_1) + \lambda^i \rho^i y_1 \leq U^1(x_1, 0)$ for all x_1 and y_1 , so that *ex ante* it is never efficient for type 1^i to produce and invest for their own consumption; they only consider consuming the proceeds opportunistically *ex post*. In this setup, by design, any trade or other interaction across groups is only interesting for its incentive effects, not for more standard mercantile reasons discussed in (international) trade theory. Also, there is no outside money, so we can concentrate on inside money.

Although Section 7 discusses nonstationarity, for now we focus on stationary, and symmetric, allocations. These are given by vectors $(x_1^i, y_1^i, x_2^i, y_2^i)$ for each group i , and descriptions of cross-group transfers, investment and diversion. If we sometimes proceed as if the planner collects production and allocates it to investors and consumers, this is only to ease the presentation – all a planner really does is make suggestions concerning actions/allocations. When there are no transfers across groups or diversion, allocations are resource feasible if $x_1^i = y_2^i$ and $x_2^i = \rho^i y_1^i$. Therefore we can summarize

⁷The introduction of $\lambda^i > 0$ is motivated by the idea that, although investment acts as collateral in the model, as Ferguson (2008) puts it “Collateral is, after all, only good if a creditor can get his hands on it.”

a feasible allocation by (x_1^i, y_1^i) . When there is no ambiguity we drop the subscript and write (x^i, y^i) . Finally, the planner/mechanism has an imperfect monitoring technology: any deviation from the suggested outcome in group $i = a, b$ is detected with probability π^i , and goes undetected with probability $1 - \pi^i$. There are many ways to rationalize this; a straightforward one is to assume imperfect record keeping: information concerning deviations “gets lost” with probability $1 - \pi^i$ across periods. The random monitoring technology differs across groups to capture the idea that some are more visible than others, and thus, presumably, less inclined to misbehave.

3 A Single Group

With a single group, we drop the group superscript i . Now all a planner/mechanism can do is recommend a resource-feasible allocation (x, y) for agents in the group. This recommendation is incentive feasible, or IF, as long as no one wants to deviate. Although we focus on the case where agents cannot commit to future actions, suppose as a benchmark they can commit to some degree. One notion is full commitment, by which we mean they can commit at the beginning of time, even before they know their type, chosen at random before production, exchange and consumption commence. Then (x, y) is IF as long as the total surplus is positive,

$$S(x, y) \equiv U^1(x, y) + U^2(\rho y, x) \geq 0. \quad (1)$$

Another notion is partial commitment, where agents can commit for the future, but only after knowing types. Then IF allocations entail two participation constraints

$$U^1(x, y) \geq 0 \quad (2)$$

$$U^2(\rho y, x) \geq 0. \quad (3)$$

The case in which we are actually interested involves no commitment. This means that, at the start of *every* period there are two participation conditions

$$U^1(x, y) + \beta V^1(x, y) \geq (1 - \pi) \beta V^1(x, y) \quad (4)$$

$$U^2(\rho y, x) + \beta V^2(x, y) \geq (1 - \pi) \beta V^2(x, y), \quad (5)$$

where $V^j(x, y)$ is the continuation value of agent j . In (4)-(5) the LHS is the payoff from following the recommendation, while the RHS is the deviation payoff.⁸ A deviation is detected with probability π , which results in a punishment to future autarky with payoff 0 (one could consider weaker punishments but this is obviously the most effective). With probability $1 - \pi$ it goes undetected and hence unpunished. Since agents are active with probability γ each period, $V^1(x, y) = \gamma U^1(x, y) / (1 - \beta)$ and $V^2(x, y) = \gamma U^2(\rho y, x) / (1 - \beta)$. From this it is immediate that (4)-(5) hold iff (2)-(3) hold, and so, by design, interesting dynamic considerations for now involve happenings across subperiods within a period.

In particular, when agent 1 invests y , he promises to deliver ρy units of good 2 in the next subperiod, but can always renege for a short-term gain $\lambda \rho y$. If caught he is punished with future autarky. So he delivers the goods only if

$$\beta V^1(x, y) \geq \lambda \rho y + (1 - \pi) \beta V^1(x, y),$$

where the RHS is the payoff to behaving opportunistically, again detected with probability $1 - \pi$. Inserting $V^1(x, y)$ and letting $\delta \equiv \lambda(1 - \beta) / \pi \gamma \beta$, this reduces to

$$U^1(x, y) \geq \delta \rho y. \tag{6}$$

As shown in Figures 1a-1b, the *repayment constraint* (6) is a clockwise rotation of 1's participation constraint. This plays a prominent role in the sequel. A low β , low monitoring probability π , low stake in the system γ , or high diversion value λ all increase δ and the temptation to default. We say an agent is more *trustworthy* when he has smaller δ , which means he can credibly promise more (or, has better credit).

⁸At the suggestion of a referee, we mention that although we do not explicitly define a formal game, we can still use methods from game theory, including the one-shot deviation principle – which, for our purposes, is nothing more than the unimprovability principle of dynamic programming.

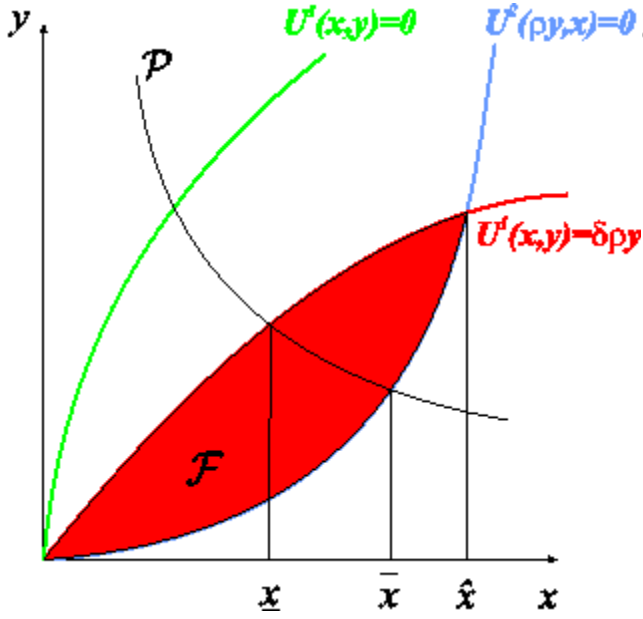


Figure 1a: RC loose.

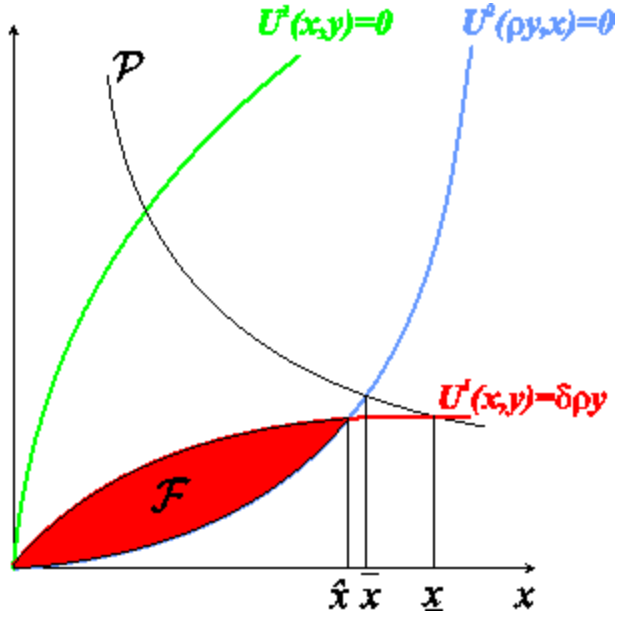


Figure 1b: RC tight

The IF set with no commitment is denoted

$$\mathcal{F} = \{(x, y) \mid (3) \text{ and } (6) \text{ hold}\}.$$

As Figures 1a-1b show, \mathcal{F} is convex, compact and contains points other than $(0, 0)$, so there are gains from trade, under the usual Inada conditions. For comparison, the IF set with partial commitment \mathcal{F}_P satisfies (2)-(3), and with full commitment \mathcal{F}_F only requires (1), where $\mathcal{F} \subset \mathcal{F}_P \subset \mathcal{F}_F$. In Figures 1a-1b (\hat{x}, \hat{y}) is the unique point other than $(0, 0)$ where (3) and (6) intersect, and clearly $\delta^b < \delta^a$ implies (\hat{x}^b, \hat{y}^b) is northeast of (\hat{x}^a, \hat{y}^a) . A more general result, also easy to verify, is:

Lemma 1 *If $\delta^a > \delta^b$ and $\rho^a = \rho^b$ or $\delta^a = \delta^b$ and $\rho^a < \rho^b$ then $\mathcal{F}^a \subset \mathcal{F}^b$.*

One can define various notions of allocations that are Pareto optimal, or PO. The *ex ante* PO allocation is the (x^o, y^o) that maximizes *ex ante* surplus $S(x, y)$. A natural criterion for *ex post* (conditional on type) welfare, which we use below, is

$$W(x, y) = \omega_1 U^1(x, y) + \omega_2 U^2(py, x). \quad (7)$$

As we vary the weights ω_i in (7) we get the Pareto set (contract curve)

$$\mathcal{P} = \left\{ (x, y) \mid \rho \frac{\partial U^1(x, y)}{\partial x} \frac{\partial U^2(\rho y, x)}{\partial y} = \frac{\partial U^2(\rho y, x)}{\partial x} \frac{\partial U^1(x, y)}{\partial y} \right\}. \quad (8)$$

It is possible that $\mathcal{P} \cap \mathcal{F} = \emptyset$ or $\mathcal{P} \cap \mathcal{F} \neq \emptyset$, as seen in Figures 1a-1b. Some other simple results (see Mattesini et al. 2009 for details) are:

Lemma 2 *Given normal goods, \mathcal{P} defines a downward-sloping curve in (x, y) space.*

Lemma 3 *$\arg \max_{(x, y) \in \mathcal{F}} W(x, y) \in \mathcal{P}$ iff (6) is not binding.*

4 Multiple Groups

Consider two groups a and b , where $\rho^a = \rho^b = \rho$, but $\delta^a > \delta^b$ so that 1^a have more of a commitment problem than 1^b . The IF set for the economy as a whole is given by allocations (x^i, y^i) for each group, plus descriptions of interactions across groups, as we now discuss. Consider first a pure transfer t : all 1^b agents produce an extra $t > 0$ units of good 2^b and give it to agents 1^a , who invest it and use the proceeds for their own benefit.⁹ Since there are γ^b/γ^a active 1^b agents for each active 1^a , payoffs are

$$\hat{U}^1(x^a, y^a, t) \equiv U^1(x^a, y^a) + \lambda^a \rho t \gamma^b / \gamma^a \quad (9)$$

$$\hat{U}^1(x^b, y^b, t) \equiv U^1(x^b, y^b + t). \quad (10)$$

We need to analyze transfers for the following reason. We are ultimately interested in a different scheme, where output from one group is transferred to the other group to invest, with the proceeds transferred back. This delegated investment activity can change the IF set, but so can pure transfers. To show that delegated investment can do more, we must first analyze transfers.

With $t > 0$, the participation conditions for 2^i are as before,

$$U^2(\rho y^i, x^i) \geq 0, \quad i = a, b, \quad (11)$$

⁹Transfers in the other direction, from 1^a to 1^b , are given by $t < 0$, and it is never useful to have simultaneous transfers in both directions. Note that t is like a lump sum tax on 1^b , with the proceeds going to 1^a , except it is not compulsory: 1^b can choose to not pay t , at the risk of punishment to future autarky, which happens with probability π^b .

but the repayment constraints for 1^i change to¹⁰

$$\hat{U}^1(x^i, y^i, t) \geq \delta^i \rho y^i, i = a, b. \quad (12)$$

The IF set with $t > 0$ satisfies (11)-(12). Notice t only enters these conditions through $\hat{U}^1(x^i, y^i, t)$. Thus, when it comes time to settle obligations, t affects the continuation values for 1^a and 1^b , but not the short-run temptation to renege. Since agents 1^a are better off and agents 1^b worse off with $t > 0$, this relaxes the repayment constraints in group a and tightens them in group b . So if these constraints are binding in group a but not b , this expands the IF set.

To see just how much we can accomplish with transfers, consider the biggest t satisfying (11)-(12). This maximization problem has a unique solution \tilde{t} , and implies allocations $(\tilde{x}^i, \tilde{y}^i)$. Clearly, \tilde{t} rises as δ^b falls. Suppose, e.g., $U^1(x, y) = x - y$, $U^2(\rho y, x) = u(\rho y) - x$, and to make the case stark $\lambda^b = 0$. Then IF allocations in group b solve

$$u(\rho y^b) - x^b \geq 0 \quad (13)$$

$$x^b - y^b - t \geq 0. \quad (14)$$

The maximum t and the implied allocation for group b are given by $\tilde{y}^b = y^*$, $\tilde{x}^b = u(\rho y^*)$ and $\tilde{t} = u(\rho y^*) - y^*$, where $\rho u'(\rho y^*) = 1$. In this example, production by 1^b is efficient, 2^b give all their surplus to 1^b , and \tilde{t} taxes away the entire surplus of group b . Giving \tilde{t} to 1^a relaxes their repayment constraint as much as possible – any $t > \tilde{t}$ leads to defection in group b .

We now introduce deposits, $d > 0$, production of good 2^a by 1^a transferred to 1^b for investment, then transferred back for consumption by 2^a . Since 1^a is now only obliged to pay out $\rho(y^a - d)$ in the second subperiod, his repayment constraint

¹⁰In case it is not clear, (12) is the condition for 1^i to pay off 2^i (i.e., agents in their own group). For 1^a who is meant to divert the returns from t , this can be written

$$\lambda^a \rho t \gamma^b / \gamma^a + \beta \hat{U}^1(x^a, y^a, t) / (1 - \beta) \geq \lambda^a \rho (t \gamma^b / \gamma^a + y^a) + (1 - \pi) \beta \hat{U}^1(x^a, y^a, t) / (1 - \beta),$$

which simplifies to (12).

becomes¹¹

$$\hat{U}^1(x^a, y^a, t) \geq \delta^a \rho(y^a - d). \quad (15)$$

Similarly, since 1^b is now obliged to pay out $\rho(y^b + d\gamma^a/\gamma^b)$,

$$\hat{U}^1(x^b, y^b, t) \geq \delta^b \rho(y^b + \gamma^a d/\gamma^b). \quad (16)$$

We also face a resource constraint

$$0 \leq d \leq y^a. \quad (17)$$

The IF set with deposits \mathcal{F}_d is given by an allocation (x^i, y^i) for each group i , together with t and d , satisfying (11) and (15)-(17).

We relax the repayment constraint in group a while tightening it in group b with $d > 0$, as we did with $t > 0$, but *deposits and transfers are different* in the way they impact incentives: t only affects continuation values, while d affects directly within-period temptations to renege by changing the obligations of 1^a and 1^b . This implies that deposits are essential, in the technical sense used above: if we start with $d = 0$, and then allow $d > 0$, for some parameters the IF set expands. We are not claiming $\mathcal{F} \subset \mathcal{F}_d$ for all parameters, or for any $d > 0$; the claim is that deposits can be essential for some parameters when we get to choose d . The following states this formally:

Proposition 1 *For all parameters $\exists d$ such that $\mathcal{F} \subset \mathcal{F}_d$; for some parameters $\exists d$ such that $\mathcal{F}_d \setminus \mathcal{F} \neq \emptyset$.*

Proof: Since any allocation in \mathcal{F} can be supported once deposits are allowed by setting $d = 0$, it is trivial that $\mathcal{F} \subset \mathcal{F}_d$. To show more allocations may be feasible with deposits, it suffices to give an example. To make it easy, suppose $\lambda^b = 0$, so that holding deposits does not affect the repayment constraint in group b . Then there are some allocations for group a that are only feasible with $d > 0$. To see this, set $t = \tilde{t}$ to maximize the transfer from b to a , as discussed above. Given $(x^b, y^b, t) = (\tilde{x}^b, \tilde{y}^b, \tilde{t})$,

¹¹These conditions allow transfers as well as deposits, as they use the payoffs defined in (9)-(10).

all incentive constraints are satisfied in group b . In group a , the relevant conditions (11) and (15) are

$$\begin{aligned} U^2(\rho y^a, x^a) &\geq 0 \\ \hat{U}^1(x^a, y^a, \tilde{t}) &\geq \delta^a \rho (y^a - d). \end{aligned}$$

For any allocation such that $\delta^a \rho y^a \geq \hat{U}^1(x^a, y^a, \tilde{t})$, $d > 0$ relaxes the repayment constraint for 1^a , and hence expands the IF set. ■

While the example used in the proof has $\lambda^b = 0$, this is not necessary. Suppose for both groups $U^1(x, y) = u(x) - y$, $U^2(y, x) = y - x$, $\rho = 1$, $\delta^i = \delta$ and $\lambda^i = \lambda \in (0, 1)$, and consider maximizing welfare W , with weights $\omega_1^i = 1$ and $\omega_2^i = 0$. Then we can summarize the set \mathcal{F}_d , when b makes transfers t and accepts deposits d , by pairs (x^a, x^b) satisfying

$$u(x^a) - x^a + \lambda t \geq \delta(x^a - d) \tag{18}$$

$$u(x^b) - x^b - t \geq \delta(x^b + d); \tag{19}$$

and symmetrically for transfers/deposits going the other way. To show deposits expand \mathcal{F} beyond what can be achieved solely with transfers, notice $t > 0$ relaxes (18) by λt and tightens (19) by t , while $d = \lambda t / \delta$ relaxes (18) by the same amount and only tightens (19) by $\lambda t < t$. To obtain the same level of slack in one group, deposits require less tightening in the other. Figure 1 shows feasible outcomes in (x^a, x^b) space for three cases. If $t = d = 0$, \mathcal{F} is given by the square. Using transfers from b to a (a to b) but no deposits, it expands by the dark blue (red) area. Using deposits from a to b (b to a) *further* expands \mathcal{F} to also include the light blue (red) area.

The economic intuition is simple. Suppose that you are 1^a , and want goods from 2^a in exchange for a pledge to pay him back later. When the time comes to pay up, if $\lambda^a > 0$ you are tempted to renege, opportunistically diverting the resources that were earmarked for settlement. This limits your credit. Your temptation is relaxed by depositing $d > 0$ with 1^b to invest on your behalf. Of course, we must also consider 1^b 's temptation, but generally, whenever $\delta^b < \delta^a$, deposits allow you to get more from

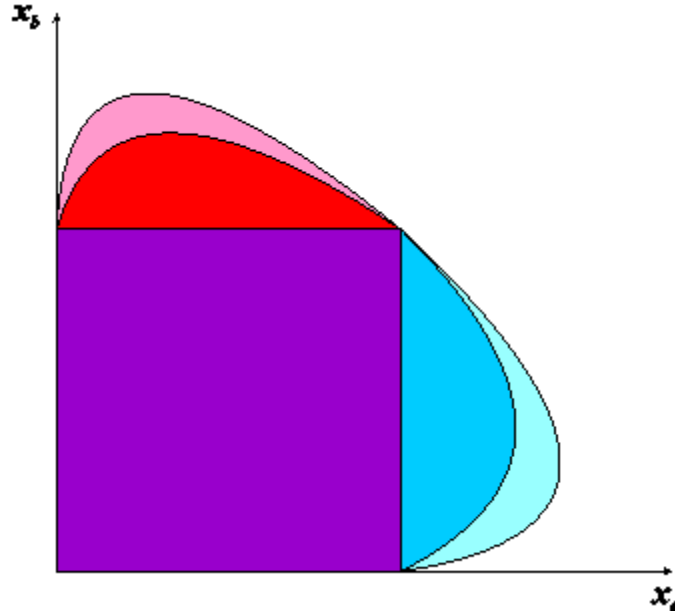


Figure 1: An example where deposits are essential.

2^a than a personal pledge. As a special case, $\lambda^b = 0$ means 1^b is totally credible – or, his investments constitute perfect collateral – so you can deposit all your resources with him. Another case of interest is $\pi^a = 0$, where your personal promise is worth nothing, in which case you cannot trade at all absent deposits. We think it is accurate to call 1^b your banker, as we explain in more detail in the next Section.

5 Inside Money

Having deposits used in payments is imperative for a complete model of banking, as over time various bank-issued instruments have played this role, from notes to checks to debit cards. This is one of the most commonly understood functions of banking, as evidenced by Selgin’s (2006) entry on “Banks” for *Encyclopedia Britannica*:

Genuine banks are distinguished from other kinds of financial intermediaries by the readily transferable or ‘spendable’ nature of their IOUs, which allows those IOUs to serve as a means of exchange, that is, money. Commercial bank money today consists mainly of deposit balances that can be transferred either by means of paper orders known as checks or electronically using plastic ‘debit’ cards.

Most formal banking models fail to speak to this issue.¹² Inside (bank) money does aid in transactions in the Cavalcanti-Wallace model, but that has nothing that resembles deposits or investment. A complete model ought to have both.

We proceed with a heuristic presentation, then give the equations. The question with which we begin is, how can a mechanism keep track of the agents' actions in the arrangement discussed in the previous Section? One way that is especially appealing when record keeping is imperfect or costly is the following: when an agent 1^a wants to consume in the first subperiod, he produces and deposits output y^a with an agent 1^b *in exchange for a receipt*. Think of the receipt as a bearer note for ρy^a . He gives this note to 2^a in exchange for his consumption good x^a . Naturally, 2^a accepts it, since the note is backed by the promise of the trustworthy 1^b . Agent 2^a carries this paper to the second subperiod, when he wants to consume (here he wants to consume with probability 1, but it is not hard to make it random, if one wants the setup to look more like Diamond-Dybvig). At that point 2^a redeems the note for his consumption good. Banker 1^b pays 2^a out of deposits – principle plus return on investments – to clear, or settle, the obligation.¹³

INSERT FIGURE 3 HERE (with key F-3)

The exchange pattern is illustrated in Figure ???. To make the story precise, we need to be specific about how agents meet and what is observed. We now explicitly interpret groups a and b as inhabiting different locations, or *islands*. To ease the exposition, let $\gamma^i = \rho^i = 1$, $U^1 = u(x) - y$ and $U^2 = u(y) - x$. Also, let $\pi^a = 0 < \pi^b$. To discuss circulating paper, we assume that any agent can costlessly produce indivisible, durable, intrinsically worthless objects that could, in principle, function as bearer notes. To avoid technical details, we assume agents can store at most

¹²In Diamond-Dybvig, e.g., agents with a desire to consume withdraw deposits and eat them. Presumably this is not meant to be taken literally, but stands in for the idea that they want to buy something. But why can't they buy it using claims on deposits as a means of payment? We understand that the model is not meant to answer this question. But then, as useful as it may be for some purposes, the model is incomplete.

¹³Another scheme one might consider is this: suppose 1^a gives his output directly to 2^a who then gives it to 1^b to invest. This has delegated investment, but not inside money. One can rule that out, however, by assuming 2^a cannot transport first-subperiod goods, just like they cannot invest them. Then receipts, which anyone can transport, are essential.

one note. (This does not affect substantive results, but it means we do not have to rule out potential deviations where agents accumulate multiple notes over time, and cash them in bundles – which they do not want to do, anyway, but it clutters the presentation to have to prove it.)

Meetings occur as follows. Within each group j , each agent 1^j is randomly matched with one 2^j for the entire period. We know from standard arguments (e.g., Wallace 2010) that some medium of exchange is necessary for trade on island a , given $\pi^a = 0$, but notes issued by 1^a have no value because 1^a has no incentive to redeem them. And these notes cannot have value as fiat objects, since no one would produce to get one when he can print his own for free. This is not the case for notes issued by 1^b . In addition to the above matching structure, before 1^j and 2^j pair off, 1^a agents travel to island b , and meet some 1^b at random. Then, in subperiod 2, agent 2^a can travel to island b and meet anyone they like – i.e., search by 2^a is directed.¹⁴ For completeness, we have to specify what happens if $n > 1$ agents of type 2^a try to match with the same 1^b agent. In this case, we assume that all n have the same probability $1/n$, but only one actually meets (can trade with) him.

Now consider a simple mechanism that suggests a particular set of actions – basically, meetings and trades – but agents can either accept or reject suggestions. If in a meeting both agents accept, they implement a suggested trade; otherwise, there is no trade in the meeting. If someone rejects a suggested trade, as above, with probability π^i they are punished with future autarky. There are four types of trades we need to suggest: (1) when 1^a meets 1^b , the former should produce and deposit $d = y^a$ in exchange for the latter’s note; (2) when 1^a meets 2^a , the latter should produce x^a in exchange for a note if the former has one, and otherwise there is no trade; (3) when 2^a meets the 1^b who issued the note, the latter redeems it for y^a ; and (4) within group b , x^b is produced by 2^b for 1^b in the first subperiod and y^b is delivered to 2^b in the second, as in the previous Sections, without using notes.

¹⁴In equilibrium, 2^a takes the note back to the 1^b agent that issued it; one can alternatively imagine type 1^b agents redeeming notes issued by *any* bank.

To describe payoffs, let $\tilde{v}_1^a(m)$ be the expected utility of 1^a when he meets 1^b and $v_1^a(m)$ his expected utility when he meets 2^a , given he has $m \in \{0, 1\}$ notes. Then

$$\begin{aligned}\tilde{v}_1^a(0) &= v_1^a(1) - y^a \\ \tilde{v}_1^a(1) &= v_1^a(1) \\ v_1^a(1) &= u(x^a) + \beta\tilde{v}_1^a(0) \\ v_1^a(0) &= \beta\tilde{v}_1^a(0).\end{aligned}$$

Thus, if 1^a has $m = 0$ when he meets 1^b , he produces/deposits y^a in exchange for a note, while if $m = 1$ they do not trade. Then, when 1^a meets 2^a , if $m = 1$ he swaps the note for x^a , while if $m = 0$ he leaves without consuming, and in either case starts next period with $m = 0$. Similarly, for 2^a

$$\begin{aligned}v_2^a(0) &= \tilde{v}_2^a(1) - x^a \\ v_2^a(1) &= \tilde{v}_2^a(1) \\ \tilde{v}_2^a(1) &= u(y^a) + \beta v_2^a(0) \\ \tilde{v}_2^a(0) &= \beta v_2^a(0),\end{aligned}$$

where $v_2^a(m)$ is the payoff when 2^a has m notes and meets 1^a , while $\tilde{v}_2^a(m)$ is the payoff when he meets 1^b . Since $\pi^a = 0$, on island a the relevant incentive conditions are $v_1^a(1) - y^a \geq v_1^a(0) \geq 0$ and $\tilde{v}_2^a(1) - x^a \geq \tilde{v}_2^a(0) \geq 0$, which reduce to

$$u(x^a) \geq y^a \tag{20}$$

$$u(y^a) \geq x^a. \tag{21}$$

Let \tilde{v}_1^b be the payoff 1^b , a representative banker, when he meets 1^a , and v_1^b his payoff when he meets 2^b in the first subperiod. Let $\hat{v}_1^b(a)$ be his payoff when he meets 2^a and $\hat{v}_1^b(b)$ his payoff when he meets 2^b in the second subperiod. Then

$$\begin{aligned}\tilde{v}_1^b &= v_1^b = u(x^b) - y^b + \hat{v}_1^b(a) \\ \hat{v}_1^b(a) &= \hat{v}_1^b(b) = \beta\tilde{v}_1^b.\end{aligned}$$

The important decision for 1^b is repayment. If he reneges on either 2^a or 2^b , he is detected and punished with probability π^b . But 2^a only gets y^a if he gives 1^b one of his notes; otherwise, the mechanism says 1^b can use the resources for his own benefit $\lambda^b y^a$. It is this part of the implementation scheme that gives 2^a the incentive to produce for a note in the first place. The payoff for 2^b is $v_2^b = u(y^b) - x^b + \beta v_2^b$, which we need to consider, of course, since we need 2^b to produce for 1^b – this is how we give 1^b an equilibrium payoff that provides him with the incentive to honor his obligations. Thus, incentive conditions for group b reduce to

$$u(y^b) \geq x^b \tag{22}$$

$$u(x^b) \geq y^b \tag{23}$$

$$u(x^b) - y^b \geq \delta^b (y^b + y^a) \tag{24}$$

Given the utility functions used here, (22)-(23) are special cases of the participation constraints, and (24) of the repayment constraint, for group b described earlier. Similarly, for group a (20)-(21) are the participation constraints, and there is no repayment constraint given $\pi^a = 0$. Summarizing, we have:

Proposition 2 *Any (x^a, y^a) and (x^b, y^b) satisfying (20)-(24) can be decentralized using banknotes. Since these same constraints define the IF set, any IF allocation can be decentralized in this way.*

In terms of economics, deposit-backed paper issued by group b is used as a payment instrument by a , which is essential since $\pi^a = 0$ implies there is no trade on island a when $d = 0$. In equilibrium, all transactions on island a are now spot trades of goods for notes – i.e., it has been fully *monetized*. This is related to the use of currency in, say, Kiyotaki-Wright (1993) (or, more accurately, since we use partially directed rather than purely random search, versions in Corbae et al. 2003 or Julien et al. 2008), but here we use bank liabilities rather than fiat objects. The difference from most banking theory, including Diamond-Dybvig, is that our bank liabilities are used in transactions. The difference from Calvo-Wallace is that our banks are more than simply note-issuing agents – they also take deposits and make investments.

6 Extensions and Applications

Having shown that banking is essential, and that IF allocations can be decentralized using deposit-backed notes, we take up other questions. For this, we use welfare criterion (7), with the same weights in each groups: $\omega_j^a = \omega_j^b$.

6.1 Who Should Hold Deposits?

With two groups and $\delta^a > \delta^b$, we claim it may be desirable in a Pareto sense to have group a deposit resources with group b , but not the other way around. Let

$$(x_0^i, y_0^i) = \arg \max_{(x^i, y^i) \in \mathcal{F}^i} W^i(x^i, y^i) \quad (25)$$

be the best IF allocation in group i with no transfers or deposits. At (x_0^i, y_0^i) , we obviously cannot make type 1^i better off without making type 2^i worse off, or vice-versa, given $d = 0$. So, we ask, can we make agents in one group better off without hurting the other group with $d \neq 0$? If deposits can help, in this regard, we say they are *Pareto essential*, or PE.¹⁵

Consider the allocation that, for some d , solves

$$(x_d^i, y_d^i) = \arg \max_{(x^i, y^i) \in \mathcal{F}_d^i} W^i(x^i, y^i), \quad (26)$$

where compared to (25), the constraint set in (26) is now \mathcal{F}_d^i . Deposits are PE if there is d such that $W^i(x_d^i, y_d^i) \geq W^i(x_0^i, y_0^i)$ for both i with one strict inequality.

Proposition 3 *Deposits from a to b are PE iff the repayment constraint binds for a and not b .*

We omit a formal proof (see Mattesini et al. 2009), but the idea is simple: if the repayment constraint binds in one group, but not the other, bankers should be selected from the latter. Suppose, e.g., $\mathcal{F}^a \subset \mathcal{F}^b$. Then, since the welfare weights are the same, if the repayment constraint does not bind in group a at (x_0^a, y_0^a) then

¹⁵Essentiality means the IF set becomes bigger or better. By PE, we mean better, according to (25). Transfers cannot help here, since they make one group worse off, so we ignore them.

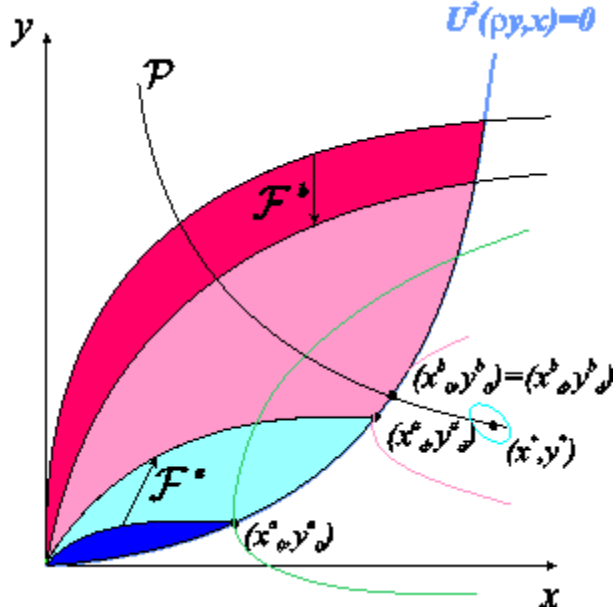


Figure 2: Illustration of Proposition 3.

it cannot bind in b . This is shown in Figure 2 for a case in which (x^*, y^*) is not feasible in either group. When $d = 0$, $(x_0^b, y_0^b) \in \mathcal{P}$ solves (25) for group b , but the commitment problem is so severe in group a that $(x_0^a, y_0^a) \notin \mathcal{P}$. Introducing $d > 0$ shifts in the repayment constraint for group b and shifts out the one for group a . This has no effect on group b , since (x_0^b, y_0^b) is still feasible, but makes group a better off.

6.2 How Should We Monitor?

We now choose monitoring intensity, and thus endogenize δ^i . Assume monitoring in group i with probability π^i implies a utility cost $\pi^i k^i$. Define a new benchmark with $d = 0$ as the solution (x^i, y^i, π^i) to

$$\max_{(x, y, \pi)} W^i(x, y) - \pi k^i \text{ s.t. } (x, y) \in \mathcal{F}^i \text{ and } \pi \in [0, 1]. \quad (27)$$

Notice the repayment constraint must bind, $U^1(x^i, y^i) = \delta^i \rho y^i$, since otherwise we could reduce monitoring costs. Also notice that (x^*, y^*) is generally not efficient when monitoring is endogenous, since reducing π implies a first order gain while moving away from (x^*, y^*) entails only a second order loss.

Suppose for the sake of illustration that we want to minimize total monitoring costs. For now, assume there is exactly active type 1 agent, which means there is a single candidate banker, in each group at each date. Obviously, if agents in one group deposit with the other, we can reduce monitoring cost in the former only by increasing it in the latter. Still, this may be desirable. In the Appendix we prove that if $\gamma^b \geq \gamma^a$, $\lambda^b \geq \lambda^a$ and $k^b \leq k^a$, then $d > 0$ may be desirable but $d < 0$ can never be. Also, we show that when 1^b has a big enough stake in the economy, he should hold all the deposits, so that we can give up monitoring agents 1^a entirely.

Proposition 4 *Fix (x^a, y^a) and (x^b, y^b) . If $\gamma^b \geq \gamma^a$, $\lambda^b \geq \lambda^a$, and $k^b \leq k^a$, then efficient monitoring implies $\delta^b < \delta^a$. Also, if γ^b is above a threshold $\bar{\gamma}$ (defined in the proof) then $\pi^a = 0$.*

One can also show that $d > 0$ may be desirable even if 1^a must compensate 1^b for increased monitoring costs. See Mattesini et al. (2009) for details, but suppose here that we distinguish between the probability of monitoring participation, which is fixed, and of monitoring repayment, which we endogenize. To characterize the efficient outcome, with quasi-linear utility, it can be shown that $d > 0$ is desirable when we compensate agents with transfers for any increase in monitoring costs under certain conditions on k^a and k^b . We can also consider the efficient number of bankers, more generally. Fewer bankers reduce total monitoring cost, but mean more deposits per bank, so that we may need to monitor them more rigorously. In fact, even if there is only one group, if one considers asymmetric allocations, it can be desirable to designate some subset of the type 1^b agents as bankers, and concentrate all monitoring on them. We leave further exploration of these ideas to future work.

6.3 Rate Of Return Dominance

To this point we assumed $\rho^a = \rho^b$, but, clearly, deposits in group b can be PE even if $\rho^a > \rho^b$ (simply by continuity). Heuristically, this explains why individuals keep wealth in demand deposits, despite the existence of alternatives with higher yields:

deposits are better payment instruments – i.e., they are more liquid. Of course, there is an important interaction between liquidity and return. Formally, we state the next result here, relegating details to the Appendix.

Proposition 5 *If $d > 0$ is PE when $\rho^b = \rho^a$, there exists $\varepsilon > 0$ such that $d > 0$ is PE when $\rho^b = \rho^a - \varepsilon$. Deposits in group b are PE if $\delta^a > \bar{\delta}^a$, and either: (a) $\delta^b \leq \bar{\delta}^b$ and $\delta^a \rho > (\rho - 1) u'(\rho \underline{y}^a)$; or (b) $\delta^b > \bar{\delta}^b$ and $\delta^a \rho > \delta^b + (\rho - 1) u'(\rho \underline{y}^a)$, with the thresholds $\bar{\delta}^a$, $\bar{\delta}^b$ and \underline{y}^a defined in the proof.*

6.4 Intermediated Lending

To this point banks undertake investments directly. In reality, although banks do invest some deposits directly, they also lend to borrowers who then make investments. The reason this is worth mentioning is that once we introduce borrowers explicitly, one may wonder how they can credibly commit to repay the bank but not to depositors. What is the use of banks as intermediaries if depositors can lend directly to investors? To address this, we now suppose there is a third group c , as well as a and b . For illustration, all parameters are the same across groups, except $\pi^b > \pi^c > \pi^a = 0$. Also, $\rho^a = \rho^b = 1$, and in all groups, $U^1(x, y) = u(x) - y$ and $U^2(y, x) = u(y) - x$. To incorporate lending, agents in group c have a special technology $f(I)$ that requires at least \bar{I} units of good y^a . Precisely, for $I < \bar{I}$ we have $f(I) = 0$, and for $I \geq \bar{I}$, we have $f(I) = \alpha I$ with $\alpha > 1$.

All is well if the minimum investment \bar{I} is small. But when it is large, \bar{I} may be too expensive for a single group a agent to lend to a group c investor. Absent other frictions, the solution is to have many 1^a agents lend to a given group c investor. But suppose we add an additional friction, that agents in group c can meet at most n other agents each period.¹⁶ Now direct lending may fail, as 1^a would have to produce enough y^a to meet the minimum investment threshold, which may not be worthwhile. And since $\pi^a = 0$, it is impossible for group a lenders to pool their resources, and

¹⁶We do not regard this as particularly deep; it is simply an example to show how other frictions can interact with limited commitment.

have one agent 1^a lend it all to some agent 1^c , since the designated 1^a agent would certainly run off with the proceeds. Here is where intermediated lending can help: a trustworthy agent 1^b collects resources from many agents 1^a , and lends them to a 1^c investor. Delegating lending through a bank allows us to meet the minimum investment level, despite commitment problems within group 1.

To fill in the details, first, note that $\pi^a = 0$ implies group a cannot consume at all without deposits. In principle, 1^a could deposit resources with 1^b for direct investment, but it may be desirable for 1^b to lend the deposits to 1^c , if $f(\cdot)$ constitutes a better opportunity. When group a lends $d = y^a$ directly to group c , the relevant incentive constraints for group a are

$$u(x^a) - y^a \geq 0 \text{ and } u(\alpha y^a) - x^a \geq 0.$$

Similarly, for b and c ,

$$\begin{aligned} u(y^b) - x^b &\geq 0 \text{ and } u(x^b) - y^b \geq \delta^b y^b \\ u(y^c) - x^c &\geq 0 \text{ and } u(x^c) - y^c \geq \delta^c (\alpha \tilde{n} y^a + y^c), \end{aligned}$$

where $\tilde{n} \leq n$ is the number of agents 1 in group a pooling their resources for lending, with $\tilde{n} y^a \geq \bar{I}$. Since an agent 1^c can meet at most n agents, assuming symmetry, the minimum resources that each 1^a must commit is \bar{I}/n . If $u(\alpha \bar{I}/n) < \bar{I}/n$, this is too large for direct lending group a to be viable.

With intermediated lending, the relevant constraints are the same for group a , but now for group b they are

$$u(y^b) - x^b \geq 0 \text{ and } u(x^b) - y^b \geq \delta^b (y^b + \hat{n} y^a - d + \alpha d).$$

The total amount received by banker 1^b from \hat{n} agents in group a is $\hat{n} y^a$, of which he lends $d \leq \hat{n} y^a$ to agents in group c . In the second subperiod, these loans return αd . Given he also invests y^b for agents in his own group, his repayment constraint is as given above. For group c the relevant constraints are

$$u(y^c) - x^c \geq 0 \text{ and } u(x^c) - y^c \geq \delta^c (\alpha \tilde{n} d + y^c),$$

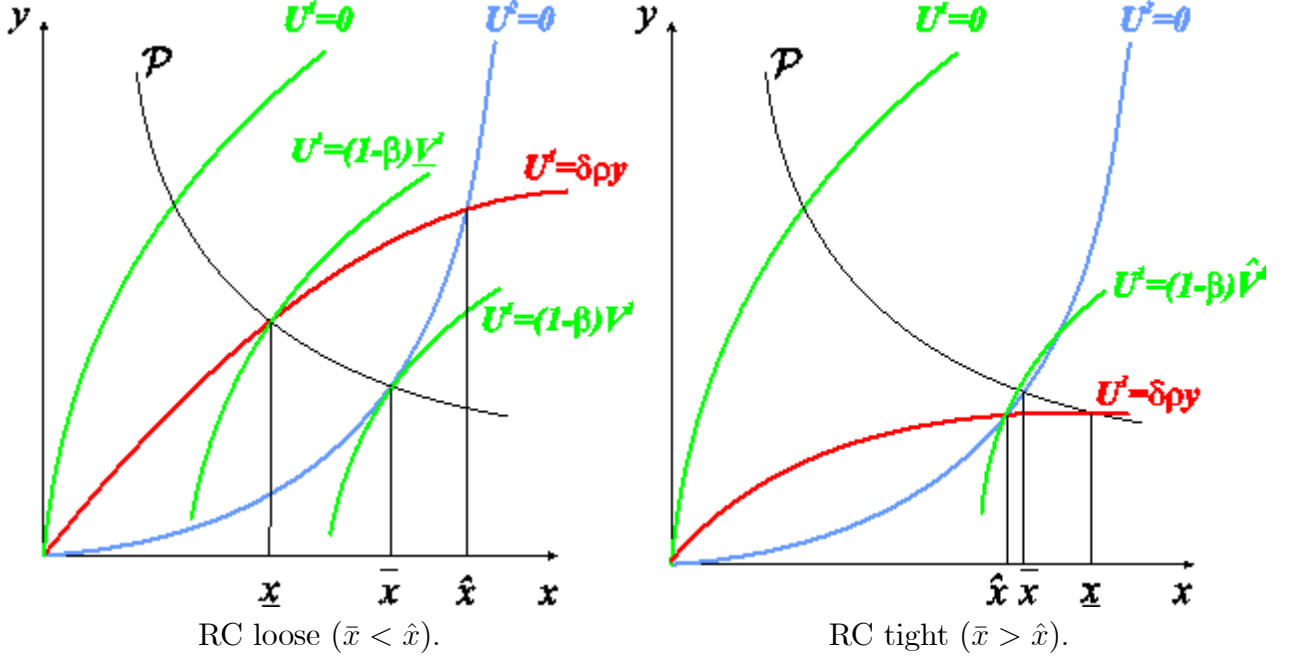
where $\tilde{n}d \geq \bar{I}$, and $\tilde{n} \leq n$ is the number of bankers lending d to an investor. If $\tilde{n} > 1$, the minimum investment is $\bar{I}/\tilde{n}n < \bar{I}/n$.

For \tilde{n} large, $u(\alpha\bar{I}/\tilde{n}n) > \bar{I}/\tilde{n}n$, and the IF set for group a contains points other than autarky. The smaller is δ^b , the larger we can set \tilde{n} . We conclude that intermediated lending can be PE, if we add additional frictions, and in particular if there is a large fixed investment \bar{I} . This is reasonable, since firms often need funds beyond what a single lender can provide. Moreover, a single lender may not want the risk exposure implied by single large investment (although we do not model this explicitly). The point is that we can extend the framework to explain how banks usefully intermediate between depositors and investors, based in part on limited commitment, and in part on other frictions.

7 Nonstationary Allocations

So far, we have restricted attention to stationary allocations. One might suspect that relaxing the stationary restriction could be good for incentives, and if this works *too* well, deposits may no longer be PE. Clearly, this is *not* true in general – e.g., if $\pi^a = 0$ then the only IF allocation in group a with $d = 0$ is autarky, and $d > 0$ may be PE.

But it would be good to know how the results are affected more generally.



We assume $\rho = \lambda = \gamma = 1$ to ease notation. We start with a single group. To begin, we identify three points of reference. Let $(\underline{x}, \underline{y})$ be the stationary allocation given by the intersection of the stationary Pareto set \mathcal{P} with 1's stationary repayment constraint, and let $\underline{V}^1 = U^1(\underline{x}, \underline{y})/(1 - \beta)$ be the associated payoff (lifetime utility) for 1. Let (\bar{x}, \bar{y}) be the intersection of \mathcal{P} with 2's stationary participation constraint and $\bar{V}^1 = U^1(\bar{x}, \bar{y})/(1 - \beta)$. Let $(\hat{x}, \hat{y}) \neq (0, 0)$ be the intersection of 1's stationary repayment constraint with 2's participation constraint and $\hat{V}^1 = U^1(\hat{x}, \hat{y})/(1 - \beta)$. Figure 7 shows these reference points for two cases, with the same preferences, but in the left panel the stationary repayment constraint is loose, which implies $\bar{x} < \hat{x}$, and in the right it is tight, which implies $\bar{x} > \hat{x}$. On the left, we show 1's indifference curves associated with \underline{V}^1 and \bar{V}^1 , but not \hat{V}^1 , since the latter does not play a role when the repayment constraint is loose; and on the right, we only show 1's indifference curves associated with \hat{V}^1 , which is relevant when repayment is tight.

To find efficient dynamic allocations, for a single group, consider the recursive

Pareto problem

$$V^2(V^1) = \max_{x,y,V_{+1}^1} U^2(y,x) + \beta V^2(V_{+1}^1) \quad (28)$$

$$\text{s.t. } V^1 = U^1(x,y) + \beta V_{+1}^1 \quad (29)$$

$$\beta \pi V_{+1}^1 \geq y \quad (30)$$

$$V^2(V_{+1}^1) \geq 0, \quad (31)$$

where the subscript on V_{+1}^1 indicates next period. The objective is to maximize type 2's (lifetime) payoff V^2 taking as given type 1's payoff V^1 , where of course we only consider $V^1 \geq 0$, since we have to satisfy 1's participation constraint at the initial date. Constraint (29), often called "promise keeping" in the literature, is the law of motion for V^1 ; (30) is the dynamic repayment constraint for 1; and (31) is the dynamic participation constraint for 2, guaranteeing he does not defect next period.

Standard methods allow us to replace (31) with $V_{+1}^1 \leq \tilde{V}^1$, where \tilde{V}^1 is the largest payoff we can give to type 1 such that the problem has a solution with $V^2 \geq 0$ – i.e., $V^2(\tilde{V}^1) = 0$. Given this, one can determine the upper bound \tilde{V}^1 explicitly, and characterize the dynamic outcome as a function of the initial condition V_0^1 , in each of the two cases shown in Figure 7.

Proposition 6 *Suppose $\bar{x} < \hat{x}$ (repayment loose). Then the upper bound is $\tilde{V}^1 = \bar{V}^1$. If $V_0^1 \in [\underline{V}^1, \bar{V}^1]$, the efficient allocation is stationary at the intersection of \mathcal{P} and the indifference curve $U^1(x,y) = (1 - \beta)V_0^1$. If $V_0^1 \in [0, \underline{V}^1)$ the efficient allocation is nonstationary, V_t^1 is strictly increasing in t until it converges to some $V_\infty^1 \in [\underline{V}^1, \bar{V}^1]$, and (30) is binding during the transition.*

Proposition 7 *Suppose $\bar{x} > \hat{x}$ (repayment tight). Then the upper bound is $\tilde{V}^1 = \hat{V}^1$. If $V_0^1 = \hat{V}^1$ the efficient allocation is stationary at (\hat{x}, \hat{y}) . If $V_0^1 \in [0, \hat{V}^1)$ the efficient allocation is nonstationary, V_t^1 is strictly increasing in t and converges to $V_\infty^1 = \hat{V}^1$. In either case, $V_0^1 = \hat{V}^1$ or $V_0^1 < \hat{V}^1$, (30) is always binding.*

In terms of economics, first suppose the stationary repayment constraint is loose. Then, if we want to treat 1 well at the initial date, in the sense that $V_0^1 \geq \underline{V}^1$, the

efficient outcome is stationary; and if we want to treat him less well, in the sense that $V_0^1 < \underline{V}^1$, then V_t^1 increases with t . In the latter case, where we do not treat 1 so well, the efficient way to encourage repayment is to backload his rewards. Intuitively, giving 1 a stationary payoff high enough to discourage misbehavior is wasteful, since over time past rewards are sunk and no longer affect incentives. This does not matter if we want to treat 1 well anyway, but if we do not, then it is better to give him an increasing sequence of utilities, since future rewards are operative now and later. But V_t^1 cannot increase indefinitely, since we also have to satisfy 2's participation constraints, so it ultimately converges to V_∞^1 . In the other case, when the stationary repayment constraint is tight, except for the extreme case $V_0^1 = \hat{V}^1$ the outcome is always nonstationary, and involves backloading 1's utility for the same reason. Also note that, in either case, the dynamic repayment constraint always binds along the transition path.

In the case of two groups, there are several possibilities. First, suppose repayment is loose in both groups, as shown in the left panel of Figure 7, and that in group a the initial condition is $V_0^{1a} < \underline{V}^{1a}$ while in group b it is $V_0^{1b} > \underline{V}^{1b}$. Ignoring deposits, from Proposition 6, the efficient allocation is stationary and unconstrained by repayment in b , while it is nonstationary and constrained in a . It should be obvious that in this case $d > 0$ is PE, since it relaxes a binding constraint for a without affecting b . Indeed, we can set $d = \bar{d}$, so that the repayment constraint just binds for b , and get the maximum slack for a without hurting b . The resulting allocation for a may now be stationary. To see this, notice that $d > 0$ rotates the repayment constraint in Figure 7, moving $(\underline{x}^a, \underline{y}^a)$ to the northwest along \mathcal{P}^a , and lowering $\underline{V}^{1a} = U^1(\underline{x}^a, \underline{y}^a)/(1 - \beta)$. This may reverse the inequality $V_0^{1a} < \underline{V}^{1a}$, meaning the efficient allocation in a may be stationary. Hence, if we have enough slack in group b , by letting type 1^b be the banker for 1^a we can relax the latter's constraints to the extent that *we no longer need backloading*. Even if \bar{d} is not big enough to reverse $V_0^{1a} < \underline{V}^{1a}$, and hence the allocation in a remains nonstationary, deposits are still PE because they slacken

repayment for a along the transition.¹⁷

So far we assumed $V_0^{1b} > \underline{V}^{1b}$. We can also have $V_0^{1i} < \underline{V}^{1i}$ for $i = a$ and $i = b$, with a nontrivial transition for both groups. Now the problem is more intricate, but it seems clear that $d_t > 0$ cannot be PE for all t since this only slackens repayment in one group at the expense of the other. However, $d_t > 0$ may be PE at some t in the future. Thus, over time $V_t^{1b} \rightarrow V_\infty^{1b}$, and suppose we get there in finite time, before V_t^{1a} converges. Then repayment will no longer bind in group b , while it is still binding in a , and it seems clear we can set $d > 0$ without hurting group b . In general, even if we know we are in the case where the repayment constraint is loose, as in the left panel of Figure 7, dynamically efficient allocations can be complicated. There is also the case where the repayment constraint is tight for both groups, or tight for one and loose for the other. In general, there are nontrivial transitions, and again dynamically efficient allocations can be complicated. But our goal here is to show that $d > 0$ can be PE, not that $d > 0$ is always PE. This has been established.

To sum up, the leading case is this: in group b , repayment is loose and $V_0^{1b} > \underline{V}^{1b}$, so the efficient outcome is stationary; in group a , repayment is either loose and $V_0^{1a} < \underline{V}^{1a}$, or repayment is tight, both of which imply a nontrivial transition involving backloading for 1^a . Then we can set $d = \bar{d} > 0$, so repayment just binds in group b , and we relax the constraint for a . In the case where in group a repayment is loose and $V_0^{1a} < \underline{V}^{1a}$, $d = \bar{d}$ may reverse the inequality since it lowers \underline{V}^{1a} , rendering the outcome stationary. Again, this means that banking can replace backloading. In any case, $d = \bar{d}$ relaxes constraints for group a . Also, notice that we can do no better if we use a nonstationary banking scheme: setting $d_t = \bar{d}$ for all t maximizes slack in group a .¹⁸ We summarize some of these results formally as follows.

¹⁷One has to be careful in this case. Proposition 6 says the transition in group a takes us to $V_\infty^{1a} \in [\underline{V}^{1a}, \bar{V}^{1a}]$, and it is important to note that we can get there in finite time, and may end up in the interior the interval (since in discrete time we jump a finite distance at each step). Once we get to V_∞^{1a} , if it is interior, then from Proposition 6 one might think repayment is no longer binding and hence $d > 0$ is no longer PE. But is not right, since repayment is no longer binding *given* $d > 0$.

¹⁸We are not saying the optimal d_t *must* be constant; if it were, say, costly to use deposits, one might want d_t to decline over time. But absent such *ad hoc* reasons, there is nothing to gain from nonstationary d_t . We also mention that efficient allocations here do *not* lead to immiseration, as they do in some private-information models (e.g., Atkeson-Lucas 1993).

Proposition 8 *Consider optimal allocations with $d_t = 0$, and suppose that at $t = 0$ the repayment constraint does (does not) bind for group a (group b). Then $d_t = \bar{d} > 0$ for all t is PE, and we can do no better by having d_t vary with t . Given $d_t = \bar{d}$, the efficient dynamic allocation is stationary for b, and can either be stationary or nonstationary for a, depending on whether \bar{d} lowers \underline{V}^{1a} enough to reverse the inequality $V_0^{1a} > \underline{V}^{1a}$.*

8 A Brief History of Banking

Here we compare our theory with some facts from banking history. First, as regards abstracting from outside money, this seems reasonable from the historical perspective, since institutions that accepted commodity deposits were operating long before the invention of coinage (let alone fiat currency). As Davies (2002) describes the situation, in ancient Mesopotamia and Egypt, goods were often deposited in palaces or temples, and later, private houses.

Grain was the main form of deposits at first, but in the process of time other deposits were commonly taken: other crops, fruit, cattle and agricultural implements, leading eventually and more importantly to deposits of the precious metals. Receipts testifying to these deposits gradually led to transfers to the order not only of depositors but also to a third party. In the course of time private houses also began to carry on such deposit business ... The banking operations of the temple and palace based banks preceded coinage by well over a thousand years and so did private banking houses by some hundreds of years.

Importantly, deposit receipts were transferrable, and hence facilitated transactions and payments, as in the model.¹⁹

In his discussion of medieval Venetian bankers, Mueller (1997) describes two types of deposits: regular, which were actual goods that bankers had to deliver on demand; and irregular deposits, involving specie or coins that only had to be repaid with the

¹⁹In ancient Babylon, also, as Ferguson (2008) says: “Debts were transferable, hence ‘pay to the bearer’ rather than a named creditor. Clay receipts or drafts were issued to those who deposited grain or other commodities at royal palaces or temples.” And, as in the model, “the foundation on which all of this rested was the underlying credibility of a borrower’s promise to repay.”

same value, not the same specie or coins. The former were like modern-day safety-deposit boxes; the latter were more like standard bank deposits, and involved a tacit agreement that the banker would invest the resources. When one puts money in a modern account, one usually does not expect to withdraw the same money, only something of appropriate value. This is true in the model, too: a bank's liability is not the deposit *per se*, but the returns on investments. Because they are making investments, bankers are more than mere storage facilities, although certainly safekeeping has something to do with it. The English goldsmiths, who many regard as the first modern bankers, originally offered their depositors little more than security, but early in the 17th century their deposit receipts began circulating in place of cash for payments, the first incarnation of banknotes; shortly after, deposits began be transferred by "drawn note" (check).²⁰

Institutions of the type modeled here – i.e., acceptors of commodities on deposit that end up facilitating transactions – were common well after the emergence of modern banking. In colonial Virginia, e.g., tobacco was commonly used in transactions because of the scarcity of precious metals, and the practice of depositing tobacco in public warehouses and then exchanging certificates, attesting to its quality and quantity, survived for over 200 years (Galbraith 1975). Similarly, in the 19th century, to facilitate transactions and credit arrangements between cocoon producers and silk weavers, warehouses commonly stored dried cocoons or silk and issued warrants that could be used to pledge for credit – the first of these warehouses being funded by a group of entrepreneurs in Lyons in 1859, later imitated by a series of Italian banks (Federico 1997). What we take away from these examples is that a very important early, and sometimes later, feature of deposit banking is that claims on deposits were used to facilitate exchange, as captured by the model.

²⁰See Joslin (1954), Quinn (1997) or Selgin (2010). Although many say goldsmiths were the first modern bankers, others mention the Templars (Weatherford 1997), who during the crusades specialized in moving and protecting money and other valuables. But it is not clear if their liabilities *circulated* as a means of payment, the way goldsmiths' receipts did. Some say that, before the goldsmiths, transferring funds from one account to another "generally required the presence at the bank of both payer and payee" (Kohn 1999; see also Quinn 2002). Even so, deposits facilitated payments. On checks, Spufford (1988) says the Florentines were using these in 1368.

In Venice, Mueller (1997) says deposit banking came to serve “a function comparable to that of checking accounts today ... not intended primarily for safekeeping or for earning interest but rather as a means of payment which facilitated the clearance of debts incurred in the process of doing business. In short, the current account constituted ‘bank money,’ money based on the banker’s promise to pay.” Of course, this only works if bankers are trustworthy. The medieval Rialto banks offer evidence consistent with this: “Little capital was needed to institute a bank, perhaps only enough to convince the guarantors to pledge their limited backing and clients to deposit their money, for it was deposits rather than funds invested by partners which provided bankers with investable capital. In the final analysis, it was the visible patrimony of the banker – alone or as part of a fraternal compagna – and *his reputation as an operator on the market place in general* which were placed on the balance to offset risk and win trust” (Mueller, p.97, emphasis added). They also were subject to monitoring, like in the model.²¹

Many bankers historically started as merchants, who almost by definition have a big connection to the market. The great banking families in Renaissance Italy and Southern Germany in the 16th century were originally merchants who began lending their own capital, and then started collecting deposits from other merchants, nobles, clerics and small investors. They were not the richest individuals: wealth was then concentrated in the hands of landowners, who controlled agriculture, forests and mineral rights. But merchants arguably had the most to lose in terms of reputation from renegeing on obligations: “because commerce involved the constant giving and receiving of credit, much of a merchant’s effort was devoted to ensuring that he could fulfill his own obligations and that others would fulfill theirs.” (Kohn 1999). Further

²¹ “[T]o maintain ‘public faith,’ the Senate in 1467 reminded bankers of their obligation to show their account books to depositors upon request.” (Mueller, p.45). If caught cheating, the punishment was indeed lifetime banishment from Venetian banking, but this apparently happened rarely in history (like in the model). Going back to Roman times, Orsingher (1967) observes: “One of the most important techniques used by Roman bankers was the use of account books analogous to those which all citizens kept with scrupulous care. This account-book was called a Codex and was indispensable in drawing up contracts A procedure peculiar to bankers deserves to be noted: the ‘editio rationum’ or production of accounts. Anyone running a bank could be compelled at a moment’s notice to produce his accounts for his clients’, or even for a third party’s, inspection.”

evidence on the bankers having a big connection to the market is given by Pressnell (1956) in its study of early English country banks during the Industrial Revolution: almost all of these emerged as a by-product of some other economic activity, often some kind of manufacturing.

Again returning again to Venice:

In the period from about 1330 to 1370, eight to ten bankers operated on the Rialto at a given time. They seem to have been relatively small operators on average... Around 1370, however, the situation changed [and] Venetian noble families began to dominate the marketplace. After the banking crisis of the 1370s and the War of Chioggia, the number of *banchi di scritta* operating at any given time on the Rialto dropped to about four, sometimes as few as three. These banks tended, therefore, to be larger and more important than before. (Mueller, p. 82)

This is something else we tried to formalize – issues concerning the efficient number of bankers, revolving around greater credibility/commitment vs. more deposits per bank. While we are not experts on history, all of this evidence taken together illustrates how the model is broadly consistent with the facts.

Finally, in terms of more recent history, what does the theory say about banking panics and the recent financial crisis? Gorton (2009) argues the crisis was a wholesale panic, where some financial firms ran on others by not renewing repurchase agreements, similar to commercial bank customers withdrawing deposits. The location of subprime risks was unknown, depositors were confused about what was at risk, and consequently ran. Our approach is too stylized to capture all these intricacies, but we can use it to think about certain issues. Suppose the probability of being active each period γ is subject to shocks, and uncertainty surrounding these shocks can induce agents to reduce deposits to re-establish incentives. Such shocks depend on the nature of business – γ could be affected, say, by housing markets if the business involves originating mortgage loans. Generally, as our δ parameter goes up, banking works less well, but this is efficient: when δ increases, not only can credit dry up, it *should* dry up. We do not claim recent events were not problematic; only that it may be interesting to look at them through the lens of models like the one developed here.

9 Conclusion

We began by specifying preferences, technologies and frictions like imperfect commitment and monitoring. We then examined incentive-feasible or efficient outcomes, and tried to interpret them in terms of actual arrangements. The model illustrates how it can be desirable for some agents to perform certain functions resembling banking: they accept deposits, they invest or make loans, and their liabilities facilitate transactions among other parties. It is good if banks offer high interest rates, but it can also be efficient to sacrifice return for trust. Banking is essential: without it, the set of feasible allocations is inferior. The framework can be generalized to analyze other phenomena, e.g., banks' tendency to borrow short and lend long. We concentrated on different issues, like deriving the set of characteristics that make for good bankers.

On that note, we close with a question Ken Burdett asked about the paper: Should we be surprised by a theory that predicts it is efficient to put agents into occupations they are good at? If the occupation is, say, singing, it is obvious that people with good range, timbre etc. are right for the job. But that is too easy, since we all agree that music and hence musicians give people direct utility. No one likes bankers, just like no one likes dollar bills, for their own sake. People like money for what it does, and we think it is better to *explain* what makes a particular object an efficient or an equilibrium medium of exchange, not just *assume* it enters in a particular way into utility functions or constraints. Similarly, we think it is useful to have a theory that is explicit about the frictions that give rise to a role for banks in the first place if we want to study their behavior, what characteristics they should have, how many we need etc. This project is a step in that direction.

Appendix

Proof of Proposition 4: Since $\gamma^b > \gamma^a$, it must be that $U^1(x^b, y^b) \geq U^1(x^a, y^a)$.

With deposits d , and since there is one candidate banker in each group, the repayment constraint in group b becomes $-\lambda^b \rho (y^b + d) + \pi^b \gamma^b \frac{\beta}{1-\beta} U^1(x^b, y^b) = 0$. Therefore, we obtain

$$\frac{\partial \pi^b}{\partial d} = \frac{1-\beta}{\beta} \frac{\lambda^b \rho}{\gamma^b U^1(x^b, y^b)}.$$

The repayment constraint in group a is $-\lambda^a \rho (y^a - d) + \pi^a \gamma^a \frac{\beta}{1-\beta} U^2(x^a, y^a) = 0$, so that

$$\frac{\partial \pi^a}{\partial d} = -\frac{1-\beta}{\beta} \frac{\lambda^a \rho}{\gamma^a U^1(x^a, y^a)}.$$

Therefore, increasing deposits from group a to b reduces the overall monitoring cost $\pi^a k^a + \pi^b k^b$ since

$$\frac{\partial \pi^a}{\partial d} k^a + \frac{\partial \pi^b}{\partial d} k^b = \frac{1-\beta}{\beta} \left[\frac{\lambda^b k^b \rho}{\gamma^b U^1(x^b, y^b)} - \frac{\lambda^a k^a \rho}{\gamma^a U^1(x^a, y^a)} \right] < 0,$$

where the inequality follows from $U^1(x^a, y^a) \leq U^1(x^b, y^b)$, $\gamma^a \leq \gamma^b$ and $k^b \leq k^a$.

Hence, from $d = 0$, only $d > 0$ can reduce total monitoring cost.

To prove the second part of the proposition, let $(\check{x}^a, \check{y}^a)$ solve $\max_{x,y} W^a(x, y)$, subject to the participation constraint for 2^a only. If

$$\bar{\pi} \equiv \frac{1-\beta}{\beta} \frac{\lambda^b (y^b + \check{y}^a)}{\gamma^b U^1(x^b, y^b)} \leq 1$$

then it is optimal to set $\pi^b = \bar{\pi}$, $d = \check{y}^a$, and $\pi^a = 0$. Then $\bar{\gamma}$ is defined as

$$\bar{\gamma} \equiv \frac{1-\beta}{\beta} \frac{\lambda^b (y^b + \check{y}^a)}{U^1(x^b, y^b)}.$$

This completes the proof. ■

Proof of Proposition 5: Consider an example with $U^1 = x - y$, $U^2 = u(\rho y) - x$, $\gamma^a = \gamma^b$, $\lambda^a = \lambda^b = 1$, $\omega_1 = \omega_2 = \omega$, and $\rho^a = \rho > 1 = \rho^b$. Given $\rho > 1$, we can have $\bar{\delta}^a < \bar{\delta}^b$, so that $(x^{*,a}, y^{*,a})$ is feasible in group a but $(x^{*,b}, y^{*,b})$ is not feasible in group b . Here we focus on the case where deposits in group b are PE. The condition $\delta^a > \bar{\delta}^a$ implies that $y^{*,a}$ is not IF in group a , so that deposits potentially have a

role. Consider the situation in group b . In the first case (a), agents in group b do not have a commitment problem because $\delta^b \leq \bar{\delta}^b$, although they do have inferior storage technology. Therefore, making deposits in group b requires agents in group a produce more to make up for a lower return to sustain a given level of consumption. The condition $\delta^a \rho > (\rho - 1) u'(\rho y^a)$ insures that δ^a is high enough so that $d > 0$ is PE. Case (b) is similar, except agents in group b have a binding repayment constraint when $\delta^b > \bar{\delta}^b$. Therefore they need to be compensated for taking deposits to prevent default. A transfer from group a does just that, but it comes on top of the additional production required from group a to cover for the loss in return. Hence, in this case, $d > 0$ is PE if $\delta^a \rho > \delta^b + (\rho - 1) u'(\rho y^a)$, which is stricter than case (a). Finally, if the commitment problem in group a is very severe, $u'(\rho y^a)$ will be large. In this case, if the investment technology in group a improves, their commitment problem must be worse for $d > 0$ to be PE.

The planner's problem with no interaction between groups is given by (25). The first best is y^{*i} solving $\rho^i u'(\rho^i y^{*,i}) = 1$. Denote by \underline{y}^i , the level of y^i that satisfies the repayment constraint at equality given δ^i . Also, Define $\bar{\delta}^i$ by $[u(\rho^i y^{*,i}) - y^{*,i}] / (\rho^i y^{*,i}) = \delta^i$ as the level above which the repayment constraint binds in group i at $y^{*,i}$. The next two claims establish the result.

Claim 1: Deposits in group b are PE if

$$\delta^a > \bar{\delta}^a, \delta^b \leq \bar{\delta}^b \text{ and } \delta^a \rho > (\rho - 1) u'(\rho y^a).$$

To verify this claim, note that given x_2^a and d , agents 1^a has to produce y^a such that $x_2^a = (y^a - d) \rho + d$. The repayment constraint is

$$u[(y^a - d) \rho + d] - \frac{y^a}{\rho} \geq \delta^a \rho (y^a - d).$$

To show deposits in group b are PE, we show increasing d relaxes the repayment constraint in group a . Hence it must be that at the allocation y^a

$$(1 - \rho) u'[(y^a - d) \rho + d] + \delta^a \rho > 0$$

$$\delta^a \rho > (\rho - 1) u'[(y^a - d) \rho + d]$$

So $d > 0$ is PE at \underline{y}^a iff $\delta^a \rho > (\rho - 1) u'(\rho \underline{y}^a)$, establishing the result.

Claim 2: Deposits in group b are PE if

$$\delta^a > \bar{\delta}^a, \delta^b > \bar{\delta}^b \text{ and } \delta^a \rho \geq \delta^b + (\rho - 1) u'(\rho \underline{y}_2^a).$$

To verify this, note that when $\delta^b > \bar{\delta}^b$, the solution to (25) in group b is \underline{y}^b . Deposits are incentive compatible only if agents 1^a make a transfer τ to agents 1^b . The repayment constraint in group b with τ and d , evaluated at \underline{y}^b , is $u(\underline{y}^b) - \underline{y}^b + \tau \geq \delta^b (\underline{y}^b + d)$. By definition, $u(\underline{y}^b) - \underline{y}^b = \delta^b \underline{y}^b$ and the minimum transfer τ that satisfies the constraint is $\tau = \delta^b d$. The repayment constraint in group a is

$$u[(y^a - d)\rho + d] - \frac{y^a}{\rho} - \tau \geq \delta^a \rho (y^a - d).$$

Substituting $\tau = \delta^b d$, we get

$$u[(y^a - d)\rho + d] - \frac{y^a}{\rho} - \delta^a \rho y^a + (\delta^a \rho - \delta^b) d \geq 0,$$

so the repayment constraint is relaxed whenever $\delta^a \rho - \delta^b \geq (\rho - 1) u'[(y^a - d)\rho + d]$. Evaluating at \underline{y}_2^a , $\delta^a \rho - \delta^b \geq (\rho - 1) u'(\rho \underline{y}_2^a)$. This establishes the claim, and concludes the proof. ■

Proof of Propositions 6 and 7: Problem (28) is a two-sided lack of commitment dynamic contracting problem in the literature. The method to solve this class of model uses some results of Thomas and Worrall (1994),²² which deliver the following: $\exists \tilde{V}^1$ such that $V^2(\tilde{V}^1) = 0$ and $V^2(V^1) \geq 0$ iff $V^1 \leq \tilde{V}^1$; and $V^2(V^1)$ is decreasing, strictly concave and continuously differentiable. Hence, we can rewrite (28) as

$$\begin{aligned} V^2(V^1) &= \max_{x, y, V_{+1}^1} U^2(y, x) + \beta V^2(V_{+1}^1) \\ \text{s.t. (29), (30) and } \tilde{V}^1 &\geq V_{+1}^1, \end{aligned} \tag{32}$$

where we replace (31) by $\tilde{V}^1 \geq V_{+1}^1$.²³ We will eventually determine \tilde{V}^1 , but first we solve (32) taking it as given. Letting μ_1 , μ_2 and μ_3 be multipliers on (29), (30) and

²²Also see Ljungqvist and Sargent (2010) for a textbook treatment.

²³We also need $V_{+1}^1 \geq 0$, but that is redundant given the repayment constraint $\beta \pi V_{+1}^1 \geq y$.

$\tilde{V}^1 \geq V_{+1}^1$, necessary and sufficient first-order conditions for a solution are

$$0 = U_x^2(y, x) + \mu_1 U_x^1(x, y) \quad (33)$$

$$0 = U_y^2(y, x) + \mu_1 U_y^1(x, y) - \mu_2 \quad (34)$$

$$0 = \beta V^{2'}(V_{+1}^1) + \beta \mu_1 + \beta \pi \mu_2 - \mu_3. \quad (35)$$

The envelope condition is $V^{2'}(V^1) = -\mu_1$, which combines with (35) to yield

$$0 = \beta V^{2'}(V_{+1}^1) - \beta V^{2'}(V^1) + \beta \pi \mu_2 - \mu_3. \quad (36)$$

Given \tilde{V}^1 , we consider separately two cases: (1) $\tilde{V}^1 > \underline{V}^1$ and (2) $\tilde{V}^1 \leq \underline{V}^1$.

Case (1): $\tilde{V}^1 > \underline{V}^1$. We describe the solution to (32) as a function of $V^1 \geq 0$. There are three possibilities: (1a) $V^1 \in [0, \underline{V}^1]$; (1b) $V^1 \in [\underline{V}^1, \tilde{V}^1]$; and (1c) $V^1 > \tilde{V}^1$. We first need the following:

Claim 3: In case (1), the solution to (32) is stationary if and only if $V^1 \in [\underline{V}^1, \tilde{V}^1]$.

To verify claim 3, first assume the solution is stationary, which means $V_{+1}^1 = V^1$. By (36), $V_{+1}^1 = V^1$ implies either: $\mu_2 = \mu_3 = 0$ or $\mu_2, \mu_3 > 0$. Suppose $\mu_2, \mu_3 > 0$. Since $\mu_2 > 0$ and the solution is stationary, (x, y) is on the stationary repayment constraint, and it is to the left of \mathcal{P} by virtue of (33)-(34). Type 1's payoff associated with this allocation is $V^1 = V_{+1}^1 < \underline{V}^1$. However, $\mu_3 > 0$ implies $V^1 = V_{+1}^1 = \tilde{V}^1$. This is a contradiction, since we are in case (1), where $\tilde{V}^1 > \underline{V}^1$.

So we cannot have $\mu_2, \mu_3 > 0$, and we must have $\mu_2 = \mu_3 = 0$. Then (33)-(34) imply $(x, y) \in \mathcal{P}$, and given stationarity, we can combine (29) and (30) to yield the stationary repayment constraint. Type 1's payoff thus satisfies $V^1 \geq \underline{V}^1$, since by definition \underline{V}^1 is the payoff associated with the point where the stationary repayment constraint intersects \mathcal{P} . Also, given stationarity, $U^1(x, y) / (1 - \beta) = V^1 = V_{+1}^1 \leq \tilde{V}^1$. This establishes that stationarity implies $V^1 \in [\underline{V}^1, \tilde{V}^1]$.

To show the converse, start by assuming $V^1 \in [\underline{V}^1, \tilde{V}^1]$, and recall that (29), (30), $V_{+1}^1 \leq \tilde{V}^1$, (33), (34) and (36) are necessary and sufficient for (32). Set $\mu_2 = \mu_3 = 0$ and $\mu_1 = -U_x^2/U_x^1$. Set $V_{+1}^1 = V^1$. Let (y, x) be the point on \mathcal{P} where $U^1(y, x) = (1 - \beta)V^1$. Then all of the relevant conditions are satisfied. Hence, $V^1 \in [\underline{V}^1, \tilde{V}^1]$

implies stationarity. Combined with the previous paragraph, we have established Claim 3.

Given this, we return to the analysis of (32) in case (1). Consider first (1a). We have the following:

Claim 1a: In case (1a) $V_{+1}^1 > V^1$.

This case is defined by $V^1 \in [0, \underline{V}^1]$. Suppose by way of contradiction that $V_{+1}^1 \leq V^1$. First suppose $V_{+1}^1 < V^1$. This implies $\mu_3 > 0$, by the strict concavity of V^2 and (36). But $\mu_3 > 0$ implies $V_{+1}^1 = \tilde{V}^1$, and we have $\tilde{V}^1 > \underline{V}^1$ since we are in case (1). Thus $V_{+1}^1 = \tilde{V}^1 > \underline{V}^1$, which contradicts $V_{+1}^1 < V^1$, since $V^1 < \underline{V}^1$ in case (1a). Now suppose $V_{+1}^1 = V^1$. This contradicts Claim 3, which tells us that the solution is stationary if and only if $V^1 \in [\underline{V}^1, \tilde{V}^1]$, and in case (1a) $V^1 \in [0, \underline{V}^1]$. So cannot have $V_{+1}^1 \leq V^1$, establishing Claim 1a.

Now consider case (1b), $V^1 \in [\underline{V}^1, \tilde{V}^1]$. By Claim 3 we know the solution is stationary and (x, y) is given by the intersection of \mathcal{P} and $U^1(x, y) = (1 - \beta)V^1$. Finally, consider case (1c), where $V^1 > \tilde{V}^1$. This implies $\mu_3 > 0$ by (36). Therefore $V_{+1}^1 = \tilde{V}^1$, and after one period we are back to case (1b), although as we show below, this cannot happen once we endogenize \tilde{V}^1 . In any event, we are done with case 1, and proceed to:

Case 2: $\tilde{V}^1 \leq \underline{V}^1$. The following three possibilities are now relevant: (2a) $V^1 < \tilde{V}^1$; (2b) $V^1 = \tilde{V}^1$; and (2c) $V^1 > \tilde{V}^1$. We need the following :

Claim 4: In case (2), the solution to (32) is stationary if and only if $V^1 = \tilde{V}^1$.

To verify claim 4, note one can show that if the allocation is stationary then $V^1 = \tilde{V}^1$ following the first part of the proof of Claim 3. To show the converse, assume $V^1 = \tilde{V}^1$. Again the first-order conditions are necessary and sufficient. It is easy to check that these are satisfied by $V_{+1}^1 = V^1 = \tilde{V}^1$, an allocation (x, y) solving $U^1(x, y) = (1 - \beta)\tilde{V}^1$ and $y = \beta\pi\tilde{V}^1$, and multipliers μ_1 , μ_2 and μ_3 given by the first-order conditions. This establishes claim 4.

We now return to (32) in this case. In (2a), where $V^1 < \tilde{V}^1$, we claim $V_{+1}^1 > V^1$. The argument is similar to case (1a). In case (2b), where $V^1 = \tilde{V}^1$, Claim 4 establishes

$V_{+1}^1 = \tilde{V}^1$, and the allocation is stationary. In case (2c), where $V^1 > \tilde{V}^1$, we claim $V_{+1}^1 = \tilde{V}^1$. The argument is similar to case (1c), but we also show below that this cannot happen once we endogenize \tilde{V}^1 . In any event, this completes the analysis of the solution to (32) for an arbitrary $\tilde{V}^1 \geq 0$.

We now need to find the value of \tilde{V}^1 that makes (32) equivalent to the original problem (28) – i.e., we seek a solution to $V^2(\tilde{V}^1) = 0$. We consider separately the two cases shown in Figure 7: a loose stationary repayment constraint, where $\underline{V}^1 < \bar{V}^1$; and a tight one, where $\underline{V}^1 > \bar{V}^1$.

Loose: $\bar{x} < \hat{x}$. In this case we claim $\tilde{V}^1 = \bar{V}^1$. To verify this, from the above results we know that $V^1 = \tilde{V}^1$ implies the optimal allocation is stationary. Given the allocation is stationary, $V^2(V^1) = U^2(y, x) / (1 - \beta)$, and we only need to find the stationary solution to $U^2(y, x) = 0$. This is obviously the (y, x) that occurs at the (nonzero) intersection of the indifference curves $U^2(y, x) = 0$ and $U^1(y, x) = \bar{V}^1$ in Figure 7. Hence, with a loose repayment constraint $\tilde{V}^1 = \bar{V}^1$. It is obvious from Figure 7 that $\underline{V}^1 < \bar{V}^1$ is equivalent to $\bar{x} < \hat{x}$. Hence, we have established proposition 6.

Tight: $\bar{x} > \hat{x}$. We now claim $\tilde{V}^1 = \hat{V}^1$. The argument is similar. If we set $V^1 = \tilde{V}^1$ then the optimal allocation is stationary. Thus, we need to find the best point (x, y) for 1 consistent with $U^2(y, x) = 0$ and the stationary repayment constraint. From the right panel of Figure 7, one can see that this is given by the intersection of participation constraint and repayment constraint, which is $V^1 = \hat{V}^1$. Again, it is clear from Figure 7 that $\underline{V}^1 > \bar{V}^1$ is equivalent to $\bar{x} > \hat{x}$. Hence, we have established proposition 7. ■

References

- Alvarez, F. and U. Jermann (2000) "Efficiency, Equilibrium, and Asset Pricing with Risk of Default," *Econometrica* 68, 775-798.
- Andolfatto, D. and E. Nosal (2009) "Money, Intermediation and Banking" *JME* 56, 289-294.
- Atkeson, A. and R.E. Lucas, Jr. (1993) "On Efficient Distribution with Private Information," *RES* 59, 427-453.
- Araujo, L. and R. Minetti (2011) "On the Essentiality of Banks," *IER*, in press.
- Biais, B., T. Mariotti, G. Plantin and J-C. Rochet (2007) "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications," *RES* 74, 345-390.
- Boyd, J. and E. Prescott (1986) "Financial Intermediary Coalitions," *JET* 38, 211-232.
- Calomiris, C. and C. Kahn (1991) "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *AER* 81, 497-513.
- Cavalcanti, R. and N. Wallace (1999a) "A Model of Private Bank Note Issue," *RED2*, 104-136.
- Cavalcanti, R. and N. Wallace (1999b) "Inside and Outside Money as Alternative Media of Exchange," *JMCB* 31, 443-457.
- Corbae, D., T. Temzelides and R. Wright (2003) "Directed Matching and Monetary Exchange," *Econometrica* 71, 731-756.
- Davies, G. (2002) *A History of Money From Ancient Times to the Present Day*, 3rd. ed. University of Wales Press.
- DeMarzo, P. and M. Fishman (2007) "Agency and Optimal Investment Dynamics," *RFS* 20, 151-188
- Diamond, D. and P. Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity," *JPE* 91, 401-419.
- Diamond, D. (1984) "Delegated Monitoring," *RES* 51, 393-414.

- Diamond, D. and R. Rajan (2001) "Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking," *JPE* 109, 287-327.
- Ennis, H. and T. Keister (2009) "Bank Runs and Institutions: The Perils of Intervention," *AER* 99, 1588-1607.
- Federico, G. (1997) *An Economic History of the Silk Industry, 1830-1930*. Cambridge University Press.
- Freixas, X. and J.C. Rochet (2008) *Microeconomics of Banking*. MIT Press.
- Ferguson, N. (2008) *The Ascent of Money*. London: Penguin.
- Galbraith, K. (1967) *The New Industrial State*. Princeton University Press.
- Gorton, G. and A. Winton (2002) "Financial Intermediation," in *Handbook of the Economics of Finance*, ed. by Constantinides et al. North Holland.
- Gorton, G. (2009) "Slapped in the Face by the Invisible Hand: Banking and the Panic of 2007," Mimeo.
- Gu, C., F. Mattesini, C. Monnet and R. Wright (2011) "Efficient Dynamic Credit Contracts," Mimeo.
- He, P., L. Huang and R. Wright (2005) "Money and Banking in Search Equilibrium" *IER* 46, 637-670.
- He, P., L. Huang and R. Wright (2008) "Money, Banking and Monetary Policy" *JME* 55, 1013-24.
- Huangfu, X. and H. Sun (2008) "Private Money and Bank Runs," *CJE*, forthcoming.
- Julien, B., J. Kennes and I. King (2008) "Bidding For Money," *JET* 142, 196-217.
- Joslin, D.M. (1954) "London Private Bankers, 1720-1785" *Econ History Rev* 7, 167-86.
- Kehoe, T.J., and D.K. Levine (1993) "Debt-Constrained Asset Markets," *RES* 60, 865-88.
- Kiyotaki, N. and J. Moore (2005) "Financial Deepening," *JEEA* 3, 701-712.
- Kiyotaki, N. and J. Moore (2008) "Liquidity, Business Cycles, and Monetary Policy," Mimeo.

- Kiyotaki, N. and R. Wright (1993) "A Search-Theoretic Approach to Monetary Economics," *AER* 83, 63-77.
- Koepl, T., C. Monnet and T. Temzelides (2008) "A Dynamic Model of Settlement," *JET* 142, 233-246.
- Kohn, M. (1999) "Early Deposit Banking," Dartmouth Economics Working Paper 99-03.
- Lagos, R. and R. Wright (2005) "A Unified Framework for Monetary Theory and Policy Analysis," *JPE* 113, 463-484.
- Leland, H.E and D.H. Pyle (1977) "Informational Asymmetries, Financial Structure and Financial Intermediation," *JFin* 32, 371-387.
- Ljungqvist, L. and T.J. Sargent (2010) *Recursive Macroeconomic Theory*. MIT Press.
- Mattesini, F., C. Monnet and R. Wright (2009) "Banking: A Mechanism Design Approach," Federal Reserve Bank of Philadelphia Working Paper No. 09-26.
- Mills, D. (2008) "Imperfect Monitoring and the Discounting of Inside Money," *IER* 49.
- Mueller, R.C. (1997) *The Venetian Money Market, Banks, Panics, and the Public Debt, 1200-1500*. Johns Hopkins University Press.
- Myers, S.C. and R. Rajan (1998) "The Paradox of Liquidity," *QJE* 113 (3), 733-771.
- Nosal, E. and G. Rocheteau (2011) *Money, Payments, and Liquidity*. MIT Press.
- Orsingher, R., (1967) *Banks of the World*. Walker and Company.
- Quinn, S. (1997) "Goldsmith-Banking: Mutual Acceptance and Interbanker Clearing in Restoration London," *Explor Econ History* 34, 411-42.
- Quinn, S. (2002) "Money, Finance and Capital Markets," in *The Cambridge Economic History of Britain since 1700, vol. 1, 1700-1860*, 3rd ed., ed. by R. Floud and D.N. McCloskey. Cambridge University Press.
- Sanches D. and S. Williamson (2010) "Money and Credit with Limited Commitment and Theft," *JET* 145.
- Selgin, G. (2006) "Banking," for *Encyclopedia Britannica*.

- Selgin, G. (2010) "Those Dishonest Goldsmiths,"Mimeo.
- Spufford, P. (1988) *Money and Its Uses in Medieval Europe*. Cambridge University Press.
- Thomas, J. and T. Worrall (1988) "Self-Enforcing Wage Contracts," *RES* 55, 541–554.
- Townsend, R. (1987) "Arrow-Debreu Programs as Microfoundations of Macroeconomics," *Advances in Economic Theory. Econometric Society Monographs Series* 12, ed. by T. Bewley. Cambridge University Press.
- Townsend, R. (1988) "Models as Economies," *EJ* 98, 1-24.
- Wallace, N. (1996) "A Dictum for Monetary Theory," in *Foundations of Research in Economics: How Do Economists Do Economics?* ed. by S. Medema and W. Samuels. Edward Elgar.
- Wallace, N. (2005) "From Private Banking to Central Banking: Ingredients of a Welfare Analysis," *IER* 46, 619-636.
- Wallace, N. (2010) "The Mechanism Design Approach to Monetary Theory," in *Handbook of Monetary Economics, 2nd Edition*, ed. by B. Friedman and M. Woodford.
- Weatherford, J. (1997) *The History of Money*. Crown Publishers.
- Williamson, S. (1986) "Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing," *JME* 18, 159-179.
- Williamson, S. (1987) "Financial Intermediation, Business Failures, and Real Business Cycles," *JPE* 95, 1196-1216.
- Williamson, S. and R. Wright (2010) "New Monetarist Economics: Models," in Benjamin Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*.