Banking: A Mechanism Design Approach*

Fabrizio Mattesini
University of Rome Tor-Vergata

Cyril Monnet
Federal Reserve Bank of Philadelphia

Randall Wright
University of Wisconsin and
Federal Reserve Bank of Minneapolis

April 29, 2011

Abstract

We study banking using mechanism design, without prior assumptions about what banks are or what they do. Given preferences, technologies and certain frictions, including limited commitment, we describe incentive-feasible allocations and interpret them as banking arrangements. Our bankers accept deposits, make investments or loans, and their liabilities (e.g., circulating banknotes) facilitate others' transactions. Banking is essential: without it, the set of feasible allocations is inferior. We show it can be efficient to sacrifice high investment returns in favor of more trustworthy bank deposits. We identify characteristics making for good bankers (e.g., patience and visibility), and compare these predictions with economic history.

*We thank many colleagues for comments and discussions, especially Chao Gu, Steve Williamson, Erwan Quintin, and Chris Phelan. 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 Philadelphia, the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

> The lending system of ancient Babylon was evidently quite sophisticated. 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... But the foundation on which all of this rested was the underlying credibility of a borrower's promise to repay. (It is no coincidence that in English the root of 'credit' is *credo*, the Latin for 'I believe'.) Ferguson (2008, p. 31)

1 Introduction

The goal of this project is to study banking without making prior assumptions about what banks are, who they are, or what they do. To this end we use mechanism design. This method, in general, begins by describing an economic environment, including preferences, technologies, and certain frictions – e.g., spatial or temporal separation, information problems, commitment issues, and so on. Then, one studies the set of allocations that are attainable, respecting both resource and incentive feasibility constraints, or the set of allocations that are optimal according to particular criteria. One then looks at these allocations and tries to interpret the outcomes in terms of institutions that can be observed in actual economies. We want to see if something that resembles banking emerges out of such an exercise. To reiterate, we do not take a bank as a primitive concept: our primitives are preferences, technologies and frictions. We want banking to arise endogenously.

Much has been written about the virtues of mechanism design generally. Our particular approach is close to that advocated by Townsend (1987, 1990). He describes the method as asking if institutions that we see in actual economies, such as observed credit or insurance arrangements, can be derived from simple but internally consistent economic models, where internal consistency means that one cannot simply *assume* that some markets are missing, contracts are incomplete, prices are sticky etc. Of course, something that looks like missing markets or incomplete contracts may *emerge*, but the idea is to specify the environment explicitly and derive this as an outcome.[1] Simple models, with minimal frictions, often do not generate arrange-

[1] As Townsend (1988) puts it: "The competitive markets hypothesis has been viewed primarily as a postulate to help make the mapping from environments to outcomes more precise ... In the end

ments that resemble actual economies; e.g., they typically predict that credit and insurance work better than the outcomes we observe. So, one asks, what additional complications can be introduced to bring the theory more in line with experience? We apply this method to banking.

Obviously some frictions are needed, since models like Arrow-Debreu have no role for banks. As has been discussed often, frictionless models have no role for any institution whose purpose is to facilitate the process of exchange. The simplest example is the institution of money, and a classic challenge in monetary economics is to ask what frictions make money *essential*, in the following technical sense: money is essential when the set of allocations satisfying incentive and other feasibility conditions is bigger or better with money than without it (see Wallace 2010). We study the essentiality of banks in this sense. Just like monetary economists ought not take the role of money as given, for this issue, we cannot take the role of banks as given. In our environment, the planner – or the mechanism – may choose to have some agents perform certain functions resembling salient elements of banking: they accept deposits, they make investments or loans, and their liabilities (claims on deposits) are used by others to facilitate exchange. This activity is essential, in the sense that if it were ruled out the set of feasible allocations would be inferior.[2]

The vast literature on banks and financial intermediation is surveyed by Gorton

though it should be emphasised that market structure should be endogenous to the class of general equilibrium models at hand. That is, 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." Relatedly, speaking more directly about banking, Williamson (1987b) says "what makes financial intermediation potentially worthy of study are its special functions (such as diversification, information processing, and asset transformation). We cannot expect to generate these special activities or derive many useful implications if our approach does not build on the economic features that cause financial intermediaries to arise in the first place."

[2] A different but related way to motivate the project is to say that we adhere to a generalized version of the Wallace dictum (Wallace 1996): "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 theory." We interpret this as follows. In Arrow-Debreu there is a set of agents consisting of firms and households, as well as a set of tradable objects consisting of consumption goods and productive inputs. The dictum to us says that one ought not simply expand the latter set to include a third object, money, as a primitive – unlike consumption goods or productive inputs, the existence and function of money should be derived endogenously, rather than assumed. Our extended version of the dictum is that one ought not expand the former set to include a third type of agent, banks – unlike consumers or producers, banks should emerge endogenously. In other words, a bank should not be a primitive in banking theory.

and Winton (2002) and Freixas and Rochet (2008). Much of this research is based on informational frictions, including adverse selection, moral hazard, and costly state verification, that hinder the channeling of funds from investors to entrepreneurs. One can distinguish broadly three main strands. One approach originating with Diamond and Dybvig (1983) interprets banks as coalitions of agents providing insurance against liquidity shocks. Another approach pioneered by Leland and Pyle (1977) and developed by Boyd and Prescott (1986) interprets banks as coalitions sharing information in ways that induce agents to truthfully reveal the quality of investments. A third approach based on Diamond (1984) interprets banks as delegated monitors taking advantage of returns to scale (see also Williamson 1986, 1987a). These papers provide many useful insights. We think we have something different to offer, especially when we study which agents should be bankers and when we highlight the role of bank liabilities in payments.[3]

Relative to information-based theories, we focus on commitment problems, although imperfect monitoring is also part of the story.[4] We are not the first to highlight commitment issues. Rajan (1998) criticized standard banking theory on the grounds that it typically assumes agents have a perfect ability to contract, and argues instead for models that rely on incomplete contracting or incomplete markets based on limited enforcement (see also Calomiris and Kahn 1991, Myers and Rajan 1998 or Diamond and Rajan 2001). We agree that limited enforcement or commitment should be central, but rather than taking market incompleteness as given, we delve into this further using the tools of mechanism design. There is related work on money and credit, including Sanches and Williamson (2010), Koeppl, Monnet and Temzelides

[3] There is too much work on Diamond-Dybvig models to list here; see Ennis and Keister (2009) for a recent contribution with references. Usually these models do not interpret the bank as a self-interested agent, nor do they derive which agents should be bankers. In papers that emphasize information and delegated monitoring, banks are agents, but their role is restricted to solving information problems, and they typically do not derive which agents will play this role. The fact that bank liabilities are useful in transactions is usually not discussed at all; for exceptions see Andolfatto and Nosal (2009), Huangfu and Sun (2008), He, Huang and Wright (2005, 2008), Cavalcanti and Wallace (1999a, 1999b), Wallace (2005) and Mills (2008). Arguably, these are all papers trying to get something that looks like banks into monetary theory, rather than papers on banking studying the role of their obligations in transactions.

[4] Some people seem to think mechanism design is only about dealing with private information; we see the method applying to other frictions as well, including limited commitment.

(2008) or Andolfatto (2008), and there is much work in general building on the limited commitment models in Kehoe and Levine (1993,2001) or Alvarez and Jermann (2000). All these papers study similar environments, although our application is very different.

Commitment issues are central because banking concerns intertemporal resource allocation, and we want to take seriously incentives to make good on one's obligations. Agents in our model make investments, which they can in principle use as collateral to ameliorate commitment problems. But this does not work well if investments are easily *liquidated* – e.g., if a debtor can simply consume the collateral (this is related to ideas in, e.g., Kiyotaki and Moore 2005, 2008). An implication is that delegated investment may be useful. If you deposit resources with a third party to invest on your behalf, and they have less of an incentive or ability to liquidate for strategic reasons, a seller may be more willing to extend you credit. Hence, claims on deposits can facilitate other transactions, which we think is a key component of banking. Therefore, although it is better if a bank has good investment opportunities, other things equal, it may be efficient to sacrifice rate of return by depositing with a bank that is less inclined to renege on obligations, because this helps with other transactions. Related to Hicks' (1935) well-known rate of return dominance question, our agents hold assets with lower returns because these facilitate exchange when they constitute the liabilities of more trustworthy parties.

Without doubt, even if venders will not accept one's personal IOU, they may accept the obligations of third parties, which throughout history took the form of notes, checks, credit/debit cards and other instruments issued by commercial banks. Of course, this begs the question, why are banks less inclined to renege on obligations? In the model, future rewards and punishments mitigate strategic behavior, so patience is relevant, but monitoring is imperfect (opportunistic behavior is detected only probabilistically). Agents with a higher likelihood of being monitored, or greater visibility, have more incentive to make good on obligations, and so they are better suited for the responsibility of accepting and investing deposits. However, we go beyond simply assuming that some can be monitored while others cannot, by allowing agents to have different probabilities of gaining from economic activity, or different stakes in the system. Even with equal visibility, those with higher stakes are less inclined to deviate from prescribed behavior because they have more to lose. This allows us to endogenize monitoring when we analyze which agents, and how many, should be bankers.[5]

In brief, we show that agents are better suited to perform the activities of banking (accepting and investing or lending deposits) to the extent that they have a good combination of the following characteristics making them more trustworthy (less inclined to renege on obligations):

- they are relatively patient;
- they are more visible, meaning 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 liquidating investments for strategic reasons.

Some of these findings, like patience being good for incentives, are fairly obvious. Others seem less so, like the idea that it can be better to delegate investments to parties with a greater stake in the system, even if they have poor investment opportunities, because their trustworthiness facilitates third-party transactions. And, as we discuss in some detail below, these results are all broadly consistent with the history of banking.

The rest of paper is organized as follows. Section 2 describes the basic environment, emphasizing the roles of temporal separation, commitment, collateral and monitoring. Section 3 characterizes incentive feasible and optimal allocations in a version of the model with a single group of agents that are heterogeneous with respect

[5] Imperfect monitoring has been studied by many people, but in theories of money and banking it is worth mentioning Kocherlakota (1998), Kocherlakota and Wallace (1998), and Cavalcanti and Wallace (1999a, 1999b). Our setup, where agents are monitored probabilistically, differs from those papers, and also from the literature in game theory where players observe only signals about each other's actions (see, e.g., Mailath and Samuelson 2006).

to type, so that at certain points in time one type wants to borrow while the other is willing to lend, but all agents of a given type (borrowers or lenders) are homogenous. Section 4 allows types to differ across groups, and in particular, allows borrowers to differ with respect to visibility, connection to the system, and so on, to derive our results on essentiality. Section 5 examines the role of bank liabilities in exchange. Section 6 presents several extensions and applications of the basic model. Section 7 discusses the implications of imposing or relaxing stationarity. Section 8 reviews some banking history. Section 9 concludes.

2 The Environment

Time is discrete and continues forever. Agents belong to one of $N \geq 1$ groups, and in each group they can be one of 2 types. Within a group, agents of a given type are homogeneous, while across groups types can be heterogeneous. The role of heterogeneous groups will be clear later; for now we focus on a representative group with a set of agents $\mathcal{A}$.[6] Each period, all agents of type j in the group can be active or inactive, and we partition $\mathcal{A}$ into three subsets: inactive agents $\mathcal{A}_0$; active type 1 agents $\mathcal{A}_1$; and active type 2 agents $\mathcal{A}_2$. These sets have measure γ_0, γ_1, and γ_2, respectively, and type j agents take as given that they belong each period to $\mathcal{A}_j$ or $\mathcal{A}_0$ with probabilities γ_j and $1 - \gamma_j$. To ease the presentation we set $\gamma_1 = \gamma_2 = \gamma$. Active agents can produce, consume, and derive utility each period, as described below. Inactive agents get utility normalized to 0, say, because they have no desire to consume or ability to produce, that period. Letting γ differ across groups captures the idea that they can have different degrees of connection to the economic system, since a bigger γ means agents have more frequent gains from trade.

In each period there are two goods, 1 and 2. Agents in $\mathcal{A}_1$ consume good 1 and produce good 2, while agents in $\mathcal{A}_2$ consume good 2 and produce good 1. Letting x_j and y_j denote consumption and production by type j, we assume utility $U^j(x_j, y_j)$ is increasing in x_j, decreasing in y_j and satisfies the usual differentiability and curvature

[6] No special restirctions on $\mathcal{A}$ are needed – it could be a continuum, countably infinite, or finite with as few as two agents.

conditions. Also, we normalize $U^j(0,0) = 0$, and we assume normal goods. A key friction is temporal separation: each period is divided in two, and good j must be consumed in subperiod j. This generates a role for credit, since type 1 must consume before type 2. To have a notion of collateral, we assume good 2 is produced in the first subperiod and invested in some way that delivers goods for consumption in the second subperiod, with a fixed gross return ρ. Investment here may be as simple as pure storage, or it could in principle involve neoclassical capital, or any other project; it is only for tractability that we use a linear investment technology with fixed return ρ.

To facilitate the presentation, there are no investment opportunities across periods, only across subperiods. Also, we do not allow type 2 agents to invest for themselves – or, more generally, at least not as efficiently as type 1 – since this would eliminate gains from intertemporal trade. Thus, only a producer of good 2, a type 1 agent, can invest it. In the formal model we usually assume that agents all discount across periods at the same rate β, but in the economic discussion we sometimes proceed as if patience differed across groups, since it is apparent what would happen if it did. This is mainly to reduce notation, but also to avoid some technical issues that can arise with heterogeneous discount rates. Our treatment of differences in patience is therefore relatively heuristic; we are more rigorous in modeling differences in visibility, connection to the system and the opportunity to liquidate.

Suppose we offer good 1 to type 1, in exchange for good 2 that in the first subperiod will be produced and invested, with the proceeds delivered to type 2 in the second subperiod. One can say that the type 1 agent is getting a loan to consume good 1, with a promise to deliver good 2 later, backed by their investment. Such a collateralized loan works well if type 1 agents get no payoff from what we call liquidating the investment, since when it comes time to deliver the goods, the production cost has been sunk. To make collateral work less well, we let type 1 derive payoff λ per unit liquidated out of investment proceeds, over and above the payoff $U^1(x_1, y_1)$. If $\lambda = 0$, as we said, collateral works well, but if $\lambda > 0$ there is an opportunity cost to delivering goods even if the production cost is sunk. This is motivated by the general idea that, as Ferguson (2008) puts it, "Collateral is, after all, only good if a creditor can get his hands on it."[7]

We assume $U^1(x_1, y_1) + \lambda\rho y_1 \leq U^1(x_1, 0)$ for all x_1, so that it is never efficient ex ante for type 1 agents to produce and invest for their own consumption, although they might consider consuming the proceeds opportunistically ex post. Type 1 agents derive the same liquidation payoff from good 2 even if it was produced by another type 1 agent, including one from a different group. But for type j agents in any group, only goods produced in the same group enter the function U^j (this is what defines a group). Imagine two groups living on two different islands, say a and b. While the output of type 1 or 2 from island a is not an argument of U^2 or U^1 for those on island b, type 1 on island b could receive and invest the produce of type 1 from island a, and pass the proceeds back to type 2 on island a. To generate the potential of opportunistic behavior in this activity, although the proceeds from investment do not enter U^1 for type 1 on island b, they can liquidate them for a per-unit payoff of λ. In this setup, by design, any trade or other interaction across islands is only interesting for its incentive effects, not the more standard mercantile effects discussed in (international) trade theory.

We focus mainly on symmetric and stationary allocations (Section 7 shows how stationarity can be restrictive). These are given by vectors $(x_1^i, y_1^i, x_2^i, y_2^i)$ for each group i, and when there is more than one group, descriptions of cross-group transfers, investment, and liquidation. We sometimes proceed as if the planner, or mechanism, collects all production and allocates it to consumers. This is merely for convenience – all a mechanism really does is make suggestions concerning production, investment and exchange. When there are no transfers across groups or liquidation, since $\gamma_1 = \gamma_2 = \gamma$, allocations are resource feasible if $x_1 = y_2$ and $x_2 = \rho y_1$, which means they can be summarized by (x_1, y_1). To reduce notation, we drop the subscript and write (x, y). Finally, there is no outside (government) money in the model. Although it may be interesting in the future to add it, for now, we focus on inside (bank) money.

[7]Formally, liquidation can be interpreted as type 1 consuming the proceeds of his investments, but this is meant to stand in for the more general idea that he could abscond with them, try to sell them, or in any other way default on obligations.

In the sequel, mainly to introduce some simple concepts and notation, we start by studying incentive feasible allocations within a single group, with borrowing and lending between types. We then allow multiple groups, across which agents of a given type (borrowers or lenders) can differ. In this context, we first show how transfers across groups can affect incentive feasible allocations, where by transfers we mean that type 1 in group a produces some y and gives it to type 1 in group b, who invests it and liquidates (consumes) the proceeds. Then we show that we can do even more with deposits than transfers, where by deposits we mean that type 1 in group a produces some y and gives it to type 1 in group b, who invests it and returns the proceeds back to group a. This to us suggests thinking about type 1 in group b as a banker. Building on this rudimentary idea, we present several applications: we allow investment returns to differ across groups and show our deposit scheme may be a good idea even if bankers do not have access to the best investment opportunities; we discuss implementation of efficient outcomes using inside inside money; we endogenize monitoring; and we study intermediated lending. Rather than starting with a full-blown general model, for pedagogical reasons, we begin with a relatively simple setup, and progressively add features as appropriate for the particular applications.

3 A Single Group

Beginning with a single group, $N = 1$, all the 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,

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

Another notion is partial commitment, where agents can commit at the start but only after knowing their type. Then IF allocations entail two participation constraints

$$U^1(x,y) \geq 0 \tag{2}$$
$$U^2(\rho y, x) \geq 0. \tag{3}$$

The case in which we are actually interested involves no commitment. This means that at the start of each period, and not just at the beginning of time, we have the participation conditions

$$U^1(x,y) + \beta V^1(x,y) \geq (1-\pi)\beta V^1(x,y) \tag{4}$$
$$U^2(\rho y, x) + \beta V^2(x,y) \geq (1-\pi)\beta V^2(x,y), \tag{5}$$

where $V^j(x,y)$ is the continuation value of type j. In (4)-(5) the LHS is type j's payoff from following the recommendation, while the RHS is the payoff from deviating, where without loss in generality we restrict attention to one-shot deviations.[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); and with probability $1-\pi$ it goes undetected and hence unpunished.[9] 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 the dynamic participation conditions hold iff (2)-(3) hold, or $U^j(x,y) \geq 0$. This means that any interesting dynamic considerations emerge from repayment considerations between subperiods, and this makes the framework uncommonly tractable.

Suppose that type 1 consumes, produces and invests in the first subperiod, in exchange for a promise to deliver good 2 in the second, but he can always renege and liquidate the proceeds ρy for a short-term gain $\lambda \rho y$. If he is caught, he is punished

[8] At the suggestion of a referee, we mention that although we do not explicitly define a formal game here we can still use methods from game theory, including the one-shot deviation principle – which, in fact, is for our purposes really just the unimprovability principle of dynamic programming (see, e.g., Kreps 1990, Appendix A2.3).

[9] At the suggestion of a referee, we clarify that while there may be many ways to rationalize our version of random monitoring, the most straightforward is to assume imperfect record keeping. Thus, information concerning deviations simply "gets lost" with probability $1-\pi$ across periods, in which case it cannot be used as a basis for punishments.

with autarky, but again he is only caught with probability π. This random monitoring technology, once we allow heterogeneity in π, captures the idea that some agents are more visible than others, and hence less inclined to renege. In any case, agent 1 delivers the goods in the second subperiod only if

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

where the RHS describes deviating by liquidating, detected with probability $1-\pi$. Inserting $V^1(x,y)$ and letting $\delta \equiv \lambda(1-\beta)/\pi\gamma\beta$, this simplifies to what we call the *repayment constraint*

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

In (6), ρy is the obligation – promised payment – in subperiod 2, and δ is an effective discount factor used when contemplating whether to make good. A low monitoring probability π, a low rate of time preference β, a low stake in economic activity γ, or a high liquidation value λ all make δ big, and hence increase temptation to default. We call an agent *more trustworthy* when he has smaller δ, since this makes him less inclined to renege. More trustworthy agents can get bigger loans, naturally, because they can credibly promise bigger repayments. Moreover, in the general model, with $N > 1$, it will be precisely the more trustworthy agents that make good bankers.

Let $\mathcal{F}_0$ denote the set of IF allocations with no commitment. Since (6) makes (2) redundant, $(x,y) \in \mathcal{F}_0$ satisfies the participation constraint (3) for type 2 and the repayment constraint (6) for type 1. For comparison, the IF set with partial commitment $\mathcal{F}_P$ satisfies (2)-(3), while the IF set with full commitment $\mathcal{F}_F$ only requires (1). Figure 1 shows $\mathcal{F}_0$ delimited by two curves defined by the relevant incentive conditions at equality,

$$\begin{aligned} C_2 &\equiv \left\{(x,y) : U^2(\rho y, x) = 0\right\} && (7) \\ C_r &\equiv \left\{(x,y) : U^1(x,y) = \delta\rho y\right\}. && (8) \end{aligned}$$

Clearly, $\mathcal{F}_0$ is convex, compact, and nonempty since $(0,0) \in \mathcal{F}_0$. We assume that it contains points other than $(0,0)$, so there are gains from trade, which holds under the usual Inada conditions. Let ξ be the unique point other than $(0,0)$ where C_2 and C_r

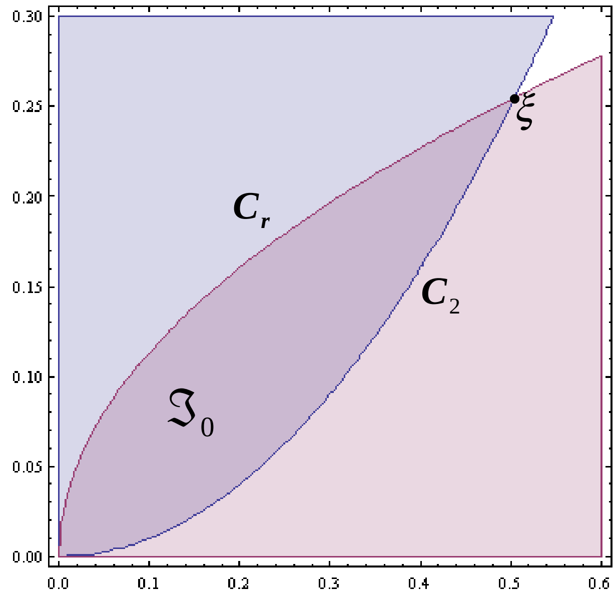

Figure 1: Incentive constraints.

intersect, as shown in Figure 1.[10] The following result is too obvious to give a proof, but still useful for what follows to state it formally:

Lemma 1 *If $\delta^b < \delta^a$, then ξ^b lies northeast of ξ^a in (x, y) space.*

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 $S(x, y)$, while a natural welfare criterion for ex post welfare is to maximize

$$\max_{(x,y)} \mathcal{W}(x, y) = \omega_1 U^1(x, y) + \omega_2 U^2(\rho y, x) \tag{9}$$

for some weights ω_1 and ω_2. As the weights vary we get the contract curve,

$$\mathcal{P} = \left\{ (x, y) \mid \rho \frac{\partial U^1(x, y)}{\partial x} \frac{\partial U^2(\rho y, x)}{\partial y} = \frac{\partial U^2(\rho y, x)}{\partial x} \frac{\partial U^1(x, y)}{\partial y} \right\}. \tag{10}$$

The core unconstrained by repayment, which may be relevant under partial commitment, is $\mathcal{K}_P = \mathcal{P} \cap \mathcal{F}_P$. The constrained core $\mathcal{K}$ is the solution to (9), as we vary the weights, subject to $(x, y) \in \mathcal{F}_0$. Given our assumptions, $\mathcal{K} \neq \emptyset$.

[10] This is drawn for an example with $U^1(x, y) = \sqrt{x} - y$, $U^2(\rho y, x) = \sqrt{\rho y} - x$, $\beta = 1/2$, $\gamma = 0.07$, $\pi = 1/2$, $\rho = 1$ and $\lambda = 1/2$.

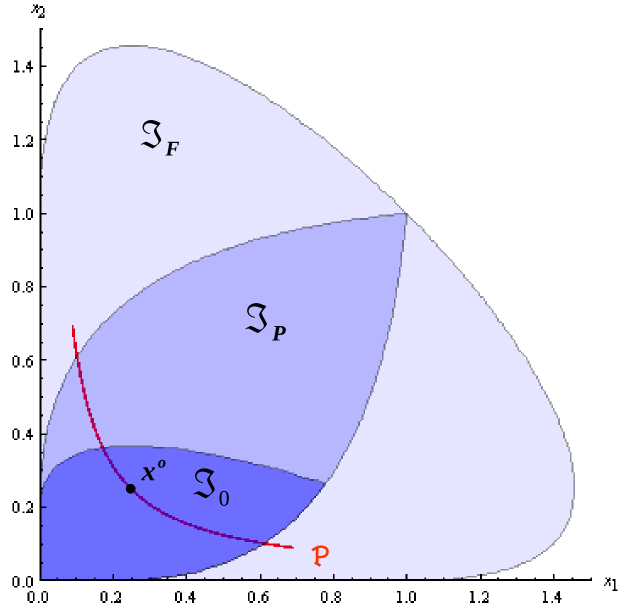

Figure 2: Pareto and incentive feasible sets.

The above results are rudimentary – basically undergraduate micro – as our objective is to keep the baseline environment totally transparent. The same can be said for the next two results (with proofs in the Appendix):

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

Lemma 3 *Let* $(\hat{x}, \hat{y})$ *maximize* $\mathcal{W}(x, y)$ *s.t.* $(x, y) \in \mathcal{F}_0$*. If the repayment constraint (6) is not binding then* $(\hat{x}, \hat{y}) \in \mathcal{P}$*.*

Figure 2 shows the IF set when we have no commitment $\mathcal{F}_0$, partial commitment $\mathcal{F}_P$ and full commitment $\mathcal{F}_F$.[11] Clearly, commitment matters: $\mathcal{F}_0 \subset \mathcal{F}_P \subset \mathcal{F}_F$. Also shown is $\mathcal{P}$, which in this example happens to intersect $\mathcal{F}_0$, but it is possible that $\mathcal{F}_0$ and $\mathcal{P}$ do not intersect, in which case the constrained core $\mathcal{K}$ and $\mathcal{P}$ do not intersect. Focusing on the case of no commitment, it is easy to see that $\mathcal{F}_0$ and $\mathcal{K}$ shrink when δ increases. Also, $\mathcal{F}_0$ and $\mathcal{K}$ shrink when ρ decreases (although this is less easy to

[11] This uses the same utility functions as the previous example, with $\beta = 3/4$, $\gamma = \pi = 1/2$ and $\rho = \lambda = 1$.

see in the Figure). We think this is a nice stylized model of credit with imperfect commitment, monitoring and collateral, yet so far it has nothing to do with banking. That comes next, when we introduce heterogeneity across groups.[12]

4 Multiple Groups

For now it suffices to consider $N = 2$ groups, labeled a and b (in Section 6.4 we use $N = 3$). Each group i has two types 1^i and 2^i, and two goods 1^i and 2^i, specialized as above; both types are active each period with probability γ^i; type 1^i has liquidation value λ^i; and we detect deviations with probability π^i. For now ρ, β and the cardinality of the set of agents $\mathcal{A}$ are the same in each group. Let $\delta^i = \lambda^i (1-\beta)/\pi^i\gamma^i\beta$ and consider the case $\delta^a > \delta^b$. This means type 1^a have more of a commitment problem than 1^b, in the sense that $\mathcal{F}_0^a \subset \mathcal{F}_0^b$, where $\mathcal{F}_0^i$ is the no-commitment IF set for group i. The IF set for the economy as a whole is given by allocations (x^i, y^i) for each group, plus a description of transfers across groups and liquidation, as discussed below, subject to relevant incentive constraints.[13]

Now suppose that in addition to producing and investing their own output, we have all active type 1^b agents produce an extra $t > 0$ units of good 2^b and transfer it to type 1^a, who invest it and liquidate the proceeds for their own benefit. Since there are γ^b/γ^a active type 1^b agents for each active type 1^a agent, the payoffs are

$$\hat{U}^1(x^a, y^a, t) \equiv U^1(x^a, y^a) + \lambda^a \rho t \gamma^b/\gamma^a \quad (11)$$

$$\hat{U}^1\left(x^b, y^b, t\right) \equiv U^1\left(x^b, y^b + t\right). \quad (12)$$

[12] Before proceeding, we mention some ways in which one can change the environment. First, rather than having permanently different types, one can assume that agents are randomly selected to be type 1 or type 2 each period, and all the results all go through but the analysis is messier (see the working paper Mattesini et. al. 2009). Second, we have a pure-exchange version that also works, similar to what one sees in much of the limited-commitment literature, but we find richer the model with production and investment. We also have a version with neoclassical investment, where the repayment constraint can be interpreted as a condition on profit, and γ as a productivity shock. Finally, although we are generally interested here in the entire IF set, one can alternatively study equilibria under some particular pricing mechanism, including different bargaining and competitive (price-taking) mechanisms, as in Gu and Wright (2010).

[13] To be clear, type 1^i can invest the output of group a or group b, and the return and liquidation value ρ and λ are the same (it is easy enough to let them differ). Also recall that goods produced by one group do not enter U^j for agents in the other group. Thus, any interesting interactions across groups will be exclusively due to incentive considerations, not the usual gains from trade.

We call t a pure transfer, and one can think of it as a lump sum tax on type 1^b, with the proceeds going to type 1^a (although it is not compulsory, as agents can chose not to pay the tax, at the risk of future autarky). Transfers in the other direction are given by $t < 0$, and it is never useful to have simultaneous transfers in both directions. Pure transfers have incentive effects that we need to analyze 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, but instead of liquidating it, they transfer the proceeds back to the first group. 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 pure transfers.

With transfers, the participation conditions for type 2^i in each group i are as before,

$$U^2\left(\rho y^i, x^i\right) \geq 0,\ i = a, b, \tag{13}$$

but the repayment constraints for type 1^i change to

$$\hat{U}^1\left(x^i, y^i, t\right) \geq \delta^i \rho y^i,\ i = a, b. \tag{14}$$

The IF set with transfer t satisfies (13) and (14).[14] 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 impact of reneging. Since type 1^b are better off and type 1^b worse off in the long run with $t > 0$, this relaxes the repayment constraints in group a and tightens them in group b. 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 pure transfers, consider the biggest transfer from group b to a satisfying (13) and (14). This standard maximization problem has a unique solution $\tilde{t}$ and implied allocation $(\tilde{x}^i, \tilde{y}^i)$ for each group i. Since the RHS of constraint (14) is increasing in δ^b, $\tilde{t}$ rises as δ^b falls (when agents are more

[14] In case it is not clear, (14) is the incentive condition for type 1^i to make a payment to type 2^i (i.e., to agents in their own group). For a type 1^a agent, who is meant to liquidate the returns from investing a transfer $t > 0$, this can be written

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

which simplifies to (14).

patient, more visible, or more connected to the system, we can extract more from them). By way of example, suppose $U^1(x,y) = x - y$, $U^2(\rho y, x) = u(\rho y) - x$, and, to make the case stark, set $\lambda^b = 0$. Then IF allocations in group b solve

$$
\begin{aligned}
u\left(\rho y^b\right) - x^b &\geq 0 && (15)\\
x^b - y^b - t &\geq 0. && (16)
\end{aligned}
$$

The maximum IF transfer and the implied allocation for group b satisfy $\tilde{y}^b = y^*$, $\tilde{x}^b = u(\rho y^*)$, and $\tilde{t} = u(\rho y^*) - y^*$, where y^* solves $\rho u'(\rho y) = 1$.

In this example, with transfer $\tilde{t}$, production by type 1^b agents $\tilde{y}^b$ is efficient, type 2^b agents give all of their surplus to 1^b by producing $\tilde{x}^b$, and we tax away the entire surplus of group b, because with $\lambda^b = 0$ we do not have to worry about repayment. Giving the proceeds of this tax to type 1^a agents allows us to relax the incentive constraint in group a, since 1^a agents now have more to lose. Giving them $\tilde{t}$ is the best we can do by way of relaxing their constraints, since any bigger tax-transfer would entail defection in group b. The point we make next is that, although pure transfers can sometimes improve the IF set, we can do even more with deposits, defined as follows: deposits $d > 0$ are units of good 2^a produced by type 1^a and transferred to type 1^b for investment, but rather than liquidating the proceeds, as they did with pure transfers, now type 1^b transfer it to group a for consumption by 2^a.

Deposits $d > 0$ entail delegated investment, with 1^b investing the output of 1^a (obviously, for deposits going the other way, set $d < 0$). This changes the repayment constraints as follows. Since type 1^a is now only obliged in the second subperiod to pay $\rho(y^a - d)$, their constraint is

$$\hat{U}^1(x^a, y^a, t) \geq \delta^a \rho (y^a - d); \tag{17}$$

and since type 1^b is now obliged to pay $\rho\left(y^b + d\gamma^a/\gamma^b\right)$, their constraint is

$$\hat{U}^1\left(x^b, y^b, t\right) \geq \delta^b \rho \left(y^b + \gamma^a d/\gamma^b\right). \tag{18}$$

These conditions allow transfers, in addition to deposits, since they use the payoffs defined in (11) and (12). We also face a resource constraint

$$0 \leq d \leq y^a. \tag{19}$$

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 (13) and (17)-(19). Notice that we relax the repayment constraint in group a while tightening it in group b with $d > 0$, as we did before with $t > 0$. But it is critical to understand that deposits and transfers are different in the way they impact incentives: t only affects continuation payoffs, while d affects directly the within-period benefits to reneging by changing the obligations of types 1^a and 1^b. Putting these observations together implies that delegated investment is essential in the following sense: if we start with $d = 0$, and then introduce deposits, the IF set may expand.[15]

Proposition 1 *$\mathcal{F}_0 \subset \mathcal{F}_d$ and for some parameters $\mathcal{F}_d \backslash \mathcal{F}_0 \neq \emptyset$.*

Proof: Since any allocation in $\mathcal{F}_0$ can be supported once deposits are allowed by setting with $d = 0$, it is trivial that $\mathcal{F}_0 \subset \mathcal{F}_d$. To show that more allocations may be feasible with deposits it suffices to give an example. To make the example easy, set $\lambda^b = 0$, so that holding deposits does not affect the incentive constraints for group b. We claim that 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 group b to a, as discussed above. Given $\left(x^b, y^b, t\right) = \left(\tilde{x}^b, \tilde{y}^b, \tilde{t}\right)$, all incentive constraints are satisfied in group b. In group a, the relevant conditions (13) and (17) are

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

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 and hence expands the IF set. ■

While the case discussed in the proof has $\lambda^b = 0$, which means 1^b agents never have incentive to renege, it should be clear that this is not necessary. Moreover, as long as it does not violate the repayment constraint for 1^b, which is certainly the case when $\lambda^b \approx 0$, we could always set $d = y^a$ and let 1^b invest *all* of the output of 1^a. In

[15] We are not claiming $\mathcal{F}_0 \subset \mathcal{F}_d$ for any $d = \bar{d} > 0$, since then the repayment constraint in group b may be violated. The claim is that deposits may be essential when we get to choose d.

this case the repayment constraint for 1^a reduces to $\hat{U}^1(x^a, y^a, \tilde{t}) \geq 0$, which is their participation condition. Thus, when λ^b is small, having agents in group a delegate all of their investment eliminates entirely their commitment problem. Another extreme case is $\pi^a \approx 0$, which means 1^a *never* repays a loan, and so credit and investment in group a cannot even get started unless 1^a deposits his output with 1^b.

By way of explicit example, suppose for both groups $U^1(x, y) = u(x) - y$, $U^2(y, x) = y - x$, $\rho = 1$, $\delta^i = \delta$, and $\lambda^i = \lambda$. Also, let $\omega_1^i = 1$, so type 2 agents get no surplus, and $x^i = y^i$ (as in a bargaining equilibrium where type 1 gets to make a take-it-or-leave-it offer). The IF sets in group a and b, when group b makes a transfer t and accepts deposits d, are defined by (x_a, x_b) satisfying

$$u(x^a) - x^a + \lambda t \geq \delta (x^a - d) \tag{20}$$

$$u\left(x^b\right) - x^b - t \geq \delta\left(x^b + d\right). \tag{21}$$

We can obtain the IF sets when group a makes a transfer and accepts deposits in a similar way. We claim that deposits expand the IF sets beyond what is achieved with transfers alone. To verify this, note that $t > 0$ relaxes constraint (20) by λt while tightening (21) by t. Now, consider deposits $d = \lambda t/\delta$. This relaxes (20) by the same amount λt, but only tightens (21) by $\lambda t < t$. Therefore, to obtain the same level of slack for group a, we require less tightening for group b with deposits rather than pure transfers.

Figure 3 shows the IF sets for group a (x-axis) and b (y-axis) in three cases. First, with $t = d = 0$, it is given by the red square. Second, using transfers from group b to a but no deposits, the IF set in group a expands by the dark blue area, and symmetrically, using transfers from group a to group b but no deposits, the IF set of group b expands by the darker red area. Finally, using deposits in group b, the IF set for group a expands more to also include the light blue area, and symmetrically, using deposits in group b, the IF set for b expands to also include the light red area. In terms of economics, suppose you are a type 1^a agent, and want to consume now in exchange for pledging to deliver later something to 2^a. When the time comes to make good, you are tempted to renege. If $\lambda^a \approx 0$, this temptation is not an issue, and your

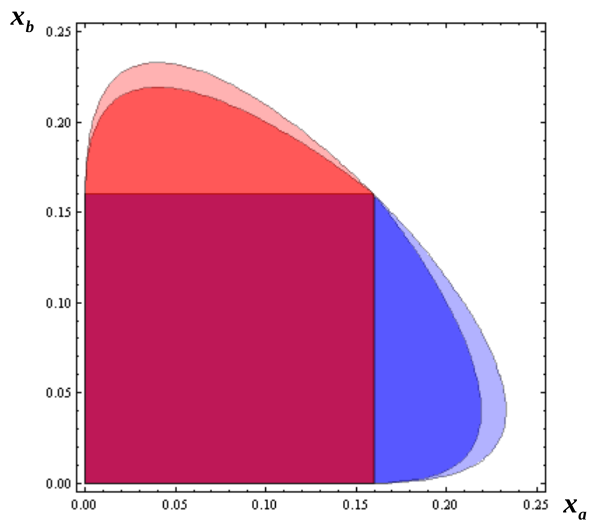

Figure 3: An example where deposits are essential.

investments are good collateral. But when $\lambda^a > 0$, collateral is imperfect, and your credit is limited. By depositing $d > 0$ with a third party, who invests on your behalf, the temptation to renege is relaxed. Now, we must consider the temptation of 1^b, in general, but as long as the third party is more trustworthy, in the sense that $\delta^b < \delta^a$, depositing $d > 0$ allows you to get more than a personal pledge.

We interpret type 1^b as a bank in the above narrative because: (1) it accepts deposits; (2) it makes investments on behalf of its depositors; and (3) its liabilities (claims on deposits) facilitate transactions. Deposits facilitate transactions because you can get more when your promises are backed by deposits – by your banker's good name, so to speak. The next section presents a particular way to implement this idea using liabilities that resemble circulating banknotes.

5 Inside Money

Having bank liabilities that are useful in payments is, to us, critical to a complete theory of banking, because various bank-issued instruments have served exactly this function over time, from notes to checks to debit cards. Provision of these instruments is one of the roles played by banks, as is commonly understood by the general public and stated in standard reference books. Selgin (2006), e.g., in his entry on "Banks" for the recent *Encyclopedia Britannica* puts it as follows: "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." To us, no theory of banking is fully satisfactory unless it speaks to this issue.

We do not mean to pick too much on any one approach, but consider a typical Diamond-Dybvig model. In that setup, agents with a desire to consume withdraw deposits and eat them. This presumably stands in for the idea that the agents want to or need to buy something. But why can't they buy it using their claims on deposits as a means of payment? In that model, agents don't actually want to buy something with their deposits, they want to eat them. This is fine for some purposes; for others it seems relevant to think about the role of deposits as a means of payment, or inside money, not merely stored consumption goods. Some discussion along these lines can be found in analyses of the model in Cavalcanti and Wallace (1999a, 1999b). What we describe next is a version of Cavalcanti-Wallace adapted to fit our environment. But we would argue that this as more than one more version of Cavalcanti-Wallace: existing work on that model may discuss the role of bank-issued liabilities as inside money, but the framework has nothing resembling deposits, delegated investments or lending, which also seem interesting components of banking.

We begin with a heuristic discussion, then give a more formal presentation. The question with which we begin is, how does the mechanism keep track of the agents' actions in the arrangement with deposits and delegated investment? One way that is appealing whenever record keeping is costly is the following: when 1^a wants to consume in the first subperiod, he produces and deposits output with 1^b *in exchange for a receipt.* Think of the receipt as a bearer note. He gives this note to type 2^a in exchange for consumption goods. Type 2^a accepts it, since the note is backed by the promise of 1^b, more readily than he would accept a personal pledge from 1^a, who is less trustworthy. Then 2^a agents carry the notes until they want to consume, which happens here to be next subperiod, although this could be easily generalized. At that point 2^a agents redeem the notes for consumption. Banker 1^b pays 2^a out of deposits – the principle plus return on investments – to clear (settle) the note. One can think these circulating liabilities as banknotes, or, with a little imagination, checks or debit cards.[16]

To flesh this out we need to be more specific about how agents meet and what is observable, since if all trade is centralized and all actions are observable it is hard to come up with a role for a medium of exchange. Hence, we now explicitly interpret groups a and b as inhabiting different locations, or islands. To ease the exposition, each group has equal measures of each type, and we set $\gamma^a = \gamma^b = 1$, $\rho^a = \rho^b = 1$, $U^1(x, y) = u(x) - y$ and $U^2(y, x) = u(y) - x$. Also let $\pi^a = 0 < \pi^b$. To be able to talk about circulating liabilities, each agent can costlessly produce indivisible, durable, and intrinsically worthless objects that can function like bearer notes. To avoid some technical details, we assume agents can store at most one note, although this does not affect any substantive results.[17] Aside from these details, everything is the same as in the benchmark model.

Using ideas in search theory, consider the following matching structure, illustrated in Figure 4. Within each group j, each type 1^j agent is randomly matched with one type 2^j agent for the entire period. We know from standard arguments (e.g., Kocherlakota 1998 or Wallace 2010) that some medium of exchange is necessary for trade on island a, given $\pi^a = 0$. But notes issued by type 1^a have no value in any first subperiod match between 1^a and 2^a, because 1^a has no incentive to redeem them when $\pi^a = 0$ (and, perhaps less obviously, these notes cannot have value as fiat objects because no one would produce to get one when he can print his own for free).

[16] We like this story about circulating liabilities, but another scheme that might also work is this: suppose 1^a gives his output directly to 2^a who then gives it to 1^b to invest. This is delegated investment, but it does not have inside money. One can however rule this out with the assumption that 2^a agents cannot transport first-subperiod goods, just like they cannot invest them. Then receipts, which anyone can transport, are essential.

[17] This assumption is for convenience only: It means we do not have to consider possible deviations where agents accumulate notes over time, and cash them in in bundles. It should be clear that agents would not want to do so, anyway, but it complicates the presentation to have to prove it.

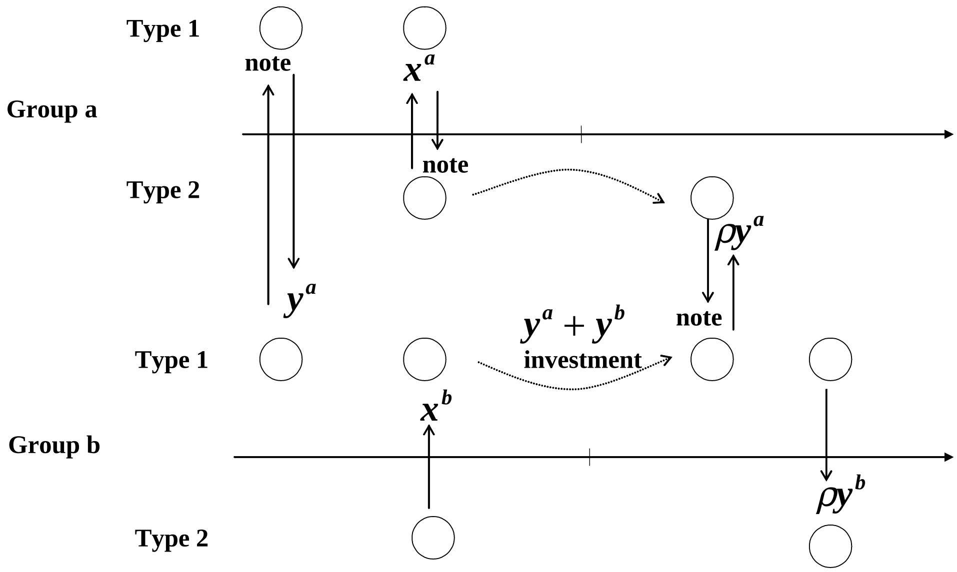

Figure 4: Circulating banknotes

Consider notes issued by agents 1^b, who will be the bankers in this discussion, as in the previous section. In addition to the above matching structure, before types 1^j and 2^j pair off for the period, type 1^a agents travel to island b where they meet some 1^b chosen at random. Similarly, in subperiod 2, type 2^a travel to island b, where they can try to meet anyone they like – i.e., search by 2^a is directed.

To anticipate the outcome, along the equilibrium path, every type 1^a visits a type 1^b agent, *his banker*, where he deposits y^a in exchange for a note. He then gives the note to a type 2^a agent in exchange for x^a. All this happens in the first subperiod. Agent 2^a carries the note to the second subperiod, when he visits 1^b to redeem the note for y^a. We also have to say what happens off the equilibrium path. Suppose that $n > 1$ type 2^a agents try to visit the same agent 1^b, say, because they all have notes issued by him, perhaps notes that they have been holding for some time. In this case, every type 2^a has the same probability $1/n$ of matching with the 1^b agent, but only one actually meets him, so some agents holding notes will not be able to redeem them.[18]

[18]Notice that we focus attention on interactions between 1^a and 2^a and how they use deposits with the third party 1^b to facilitate these interactions. There are of course also interactions between 1^b and 2^b but these look like what we saw in the previous sections. That is, they engage in credit-like

Consider direct mechanisms that suggest an allocation to the agents, but in each match they can accept or reject the suggestion. If they both accept, the pair implement the suggested actions; otherwise, nothing happens in the meeting. But if someone rejects a suggested trade, with probability π^i they are punished with future autarky, as above. There are four types of trades we need to consider: (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 they do not 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 subperiod, without the exchange of notes.

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

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

In words, if 1^a has $m = 0$ notes when he meets a banker, he produces/deposits y^a in exchange for a note; if he already has a note, he simply holds onto it. Then, when he meets 2^a, if he has a note he trades it for x^a and if he does not he leaves without consuming, and in either case starts next period with $m = 0$. Similarly, for 2^a

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

transactions supported by punishment to autarchy when someone reneges. There are various ways to think about decentralizing this, including the approach in Kocherlakota and Wallace (1998).

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 2^a meets 1^b.

Since defections in group a cannot be punished when $\pi^a = 0$, the relevant incentive conditions are

$$
\begin{aligned}
v_1^a(1) - y^a &\geq v_1^a(0) \geq 0 && (22)\\
\tilde{v}_2^a(1) - x^a &\geq \tilde{v}_2^a(0) \geq 0. && (23)
\end{aligned}
$$

Thus, (22) says 1^a agrees to produce/deposit y^a in exchange for a note, and (23) says 2^a agrees to produce for a note. We can reduce these to

$$
\begin{aligned}
u(x^a) &\geq y^a && (24)\\
u(y^a) &\geq x^a, && (25)
\end{aligned}
$$

which constitute a special case of the participation constraints in the previous sections that applies with $\pi^a = 0$ and quasi-linear utility.

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

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

The important decision for 1^b is repayment. If he reneges on his obligation to either 2^a or 2^b, he is detected with probability π^b, and punished accordingly. But 2^a only gets y^a if he gives 1^b one of his notes; otherwise, the mechanism says 1^b can liquidate his delegated investments for an instantaneous payoff $\lambda^b y^a$. It is this part of the implementation scheme that gives 2^a the incentive to produce in exchange for a note in the previous subperiod back on island a.

Continuing, the payoff of 2^b is simply

$$
v_2^b = u\left(y^b\right) - x^b + \beta v_2^b.
$$

The relevant incentive conditions for group b easily simplify to

$$u\left(y^{b}\right) \geq x^{b} \tag{26}$$

$$u\left(x^{b}\right) \geq y^{b} \tag{27}$$

$$\pi^{b}\beta\frac{u\left(x^{b}\right)-y^{b}}{1-\beta} \geq \lambda^{b}\left(y^{b}+y^{a}\right) \tag{28}$$

Notice (26)-(27) are exactly the participation constraints, and (28) is the repayment constraint, in previous sections, specialized to the case of quasi-linear utility. We saw above that for group a (24)-(25) are exactly the participation constraints in previous sections, and there is no repayment constraint (since $\pi^a = 0$ implies 1^a deposits all his output).

Summarizing the above discussion, we have:

Proposition 2 *Any* (x^a, y^a) *and* $\left(x^b, y^b\right)$ *satisfying (24)-(28) can be decentralized as an equilibrium using a direct mechanism with banknotes. Since these same constraints define the IF set, any IF allocation can be decentralized in this way.*

The economic content of this Proposition is the following. First, deposit-backed notes issued by bankers are used as payment instruments by group a. This is essential since $\pi^a = 0$ implies there can be no trade on island a without said notes. Bankers have the incentive to redeem notes, by (28), and 2^a always wants to give up a note for y^a (holding it to the next period is feasible but not desirable). Again, it is the fact that he needs a note to get y^a that makes 2^a willing to produce x^a in the first place. With this decentralization, all trades on island a are spot transactions of goods for notes; we have effectively monetized all their intertemporal exchange. Indeed, the use of notes here is similar to the use of money in Kiyotaki and Wright (1989, 1993).[19] The difference from pure monetary theory is that we have banks; the difference from most banking theory is our bank liabilities facilitate transactions; and the difference from other models where bank liabilities also serve this function, like Cavalcanti-Wallace, is that our banks do more than just issue notes, they also accepts deposits and make investments. And, as we show in the next section, can also make loans.

[19] One technical difference is that here, rather than purely random search, we have partially directed search, making the setup more like Corbae, Temzelides and Wright (2003) or Julien, Kennes and King (2008).

6 Extensions and Applications

Having shown that banking is essential, in the sense that it can expand the IF set, and that IF allocations can be decentralized using deposit-backed nabk notes, we now explore several related issues. We first go into more detail concerning who should hold deposits (previously we took for granted that it should be 1^b). We then study how to monitor when it is costly. We then expand on the rate-of-return dominance issue. And we discuss how to extend the model so that banks lend to investors, as opposed to making investments for themselves. For all of this we us the welfare criterion in (9), and assume the weights are equal across groups: $\omega_j^a = \omega_j^b$.

6.1 Who Should Hold Deposits?

Consider two groups with $\delta^a > \delta^b$. We claim that it may be desirable, in a Pareto sense, to have group a deposit resources with group b, but it is *never* desirable to have group b deposit with group a. Let $(\hat{x}^i, \hat{y}^i)$ be the best IF allocation for group i with no transfers or deposits, solving

$$\max_{x^i, y^i} \mathcal{W}^i \left(x^i, y^i\right) \text{ s.t. } \left(x^i, y^i\right) \in \mathcal{F}_0^i, \tag{29}$$

where welfare $\mathcal{W}$ is defined in (9), given some weights. At $(\hat{x}^i, \hat{y}^i)$, obviously, without deposits no IF allocation for group i makes 1^i better off without making 2^i worse off, and vice-versa. Then we ask, given $(\hat{x}^a, \hat{y}^a)$ and $\left(\hat{x}^b, \hat{y}^b\right)$, can deposits make agents in one group better off without hurting the other? Transfers cannot help, in this regard, since the group making the transfer is always worse off, so we ignore them for this discussion.

If deposits can help, we say they are Pareto essential, or PE.[20] Consider the allocation $(\tilde{x}^i, \tilde{y}^i)$ that, for some d, solves

$$\max_{x^i, y^i} \mathcal{W}^i \left(x^i, y^i\right) \text{ s.t. } \left(x^i, y^i\right) \in \mathcal{F}_d^i, \tag{30}$$

where we note that the constraint sets are different in (29) and (30). Deposits are PE if there is d such that $\mathcal{W}^i(\tilde{x}^i, \tilde{y}^i) \geq \mathcal{W}^i(\hat{x}^i, \hat{y}^i)$ for both i with one strict inequality.

[20] Recall that essential means the IF set becomes bigger or better. By PE, we mean better, according to the Pareto criterion.

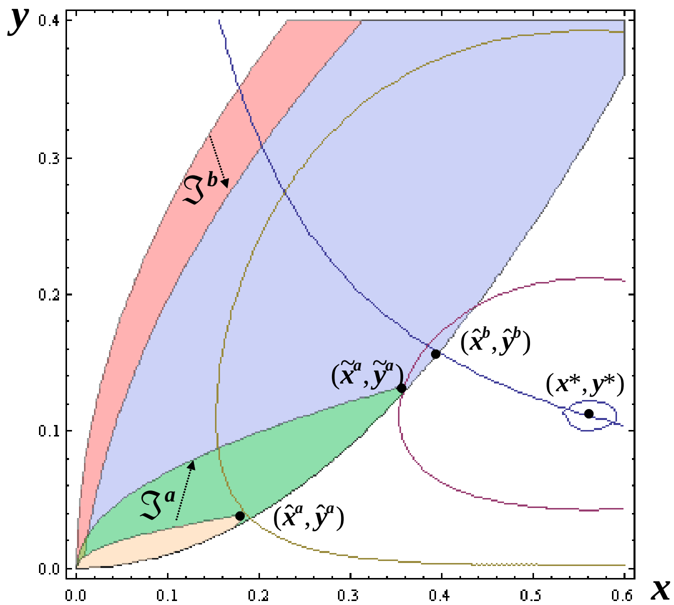

Figure 5: Illustration of Proposition 3.

A necessary condition for PE in group i is that the repayment constraint does not bind at $(\hat{x}^i, \hat{y}^i)$, since otherwise, deposits will make the repayment constraints tighter, shrink the IF set, and lower $\mathcal{W}^i$.

Proposition 3 *Deposits are PE iff the repayment constraint binds for one group and not the other.*

Proof: Let $(\hat{x}^i, \hat{y}^i)$ be the best IF allocation for group i with no deposits. Notice deposits only affect the repayment constraint. There are two possible cases for each group: either the repayment constraint is binding or it is not at $(\hat{x}^i, \hat{y}^i)$. If the repayment constraint binds for both groups, then $d \neq 0$ tightens it in one group, making it worse off, so deposits cannot be PE. If it does not bind in either group, deposits again cannot be PE. But when the repayment constraint binds for one group and not the other, a small $d \neq 0$ can relax it in the group where it binds without affecting the other group. In this case deposits are PE. ■

In economies where the repayment constraint binds in group a but not b, bankers are selected from group b, not a. This is the case if, e.g., $\rho^a = \rho^b$ and $\delta^a > \delta^b$. In this economy $\mathcal{F}_0^a \subset \mathcal{F}_0^b$, and since the welfare weights are the same, if the repayment constraint does not bind in group a at $(\hat{x}^a, \hat{y}^a)$ then it cannot bind in group b. Other things equal, bankers should come from the group with less of a commitment problem. This is illustrated in Figure 5 for a case in which (x^*, y^*) is not feasible in either group. When $d = 0$, $(\hat{x}^b, \hat{y}^b) \in \mathcal{P}$ solves (29) for group b, but the commitment problem is so severe in group a that $(\hat{x}^a, \hat{y}^a) \notin \mathcal{P}$. Introducing $d > 0$ shifts the repayment constraint for group b in and the one for group a out. This has no effect on group b since $(\hat{x}^b, \hat{y}^b)$ is still feasible, but makes group a better off since $(\tilde{x}^a, \tilde{y}^a)$ becomes feasible. Hence, we can make group a better off without hurting b, with $d > 0$, but we cannot make group b better off without hurting a. Similar logic applies if, instead of δ, we allow ρ or ω to vary across groups. In conclusion, bankers should come from the group with less of a commitment problem, with better investment opportunities, or with a higher welfare weight on type 1.

6.2 How Should We Monitor?

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

$$\max_{(x,y,\pi)} \mathcal{W}^i (x, y) - \pi k^i \text{ s.t. } x \in \mathcal{F}_0^i \text{ and } 0 \leq \pi \leq 1. \tag{31}$$

The repayment constraint must be binding, $U^1 (x^i, y^i) = \delta^i \rho y^i$, or we could reduce monitoring costs. Also, notice that (x^*, y^*) is typically 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.

In this application we are interested in minimizing total monitoring costs, rather than asking if deposits are PE; below we discuss using transfers to compensate agents for changes in cost. Also, for now, there is only one active agent in each group at each date, which means there is a single candidate banker in each group; below we discuss a more general case. Obviously, if agents in one group deposit output with the other group, we can reduce the cost of monitoring the former only at the expense of 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 $d < 0$ cannot be. Also, we show that when 1^b has a big enough stake in the economy γ^b, he should hold all the deposits, so that we can give up monitoring type 1^a entirely.

Proposition 4 *Fix* (x^a, y^a) *and* $\left(x^b, y^b\right)$*. 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 show that $d > 0$ may be desirable even if 1^a must compensate 1^b for increased monitoring costs. The working paper provides details but, briefly, suppose we distinguish between the probability of monitoring participation, which is fixed, and of monitoring repayment, which we endogenize. In this case, characterize the efficient repayment-monitoring probability and cost. With quasi-linear utility, we show 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 also consider the efficient number of bankers more generally. Fewer bankers reduce total monitoring cost, but imply more deposits per banker, meaning that we need to monitor them more vigorously. In fact, even if there is only one group, if one considers asymmetric allocations, it can be desirable to designate some subset as bankers and concentrate all monitoring on them. In the working paper we analyze the optimal number of bankers and show, e.g., that there should be fewer of them when they have more at stake in the economy.

6.3 Rate Of Return Dominance

We now show that the best bankers need not have the best investment opportunities. This has implications for perhaps *the* classic issues in monetary economics, rate of return dominance. Each unit invested in group i now returns ρ^i, different across groups. We claim that for some parameters deposits in group b are PE, despite higher investment returns in group a. Intuitively, this helps us understand why individuals keep wealth in demand deposits, despite the existence of alternatives with higher yields: these deposits make good payment instruments, i.e., they are more liquid.

Proposition 5 *For all $\delta^b < \delta^a$ there exists $\rho^b < \rho^a$ such that $d > 0$ is PE.*

Proof: When $\rho^a = \rho^b$, we already established that $d > 0$ is PE if $\delta^b < \delta^a$. By continuity, for small ε this is also true when $\rho^b = \rho - \varepsilon$. ■

Of course there is an interaction between trustworthiness and return. When group a deposits with group b, they give up $\rho^a - \rho^b$, and if this difference is large $d > 0$ may not be PE, since it can tighten repayment constraints. In the Appendix, we demonstrate the following in 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 $\rho^a = \rho > 1 = \rho^b$.

Proposition 6 *Deposits in group b are PE if $\delta^a > \overline{\delta}^a$, and either: (a) $\delta^b \le \overline{\delta}^b$ and $\delta^a \rho > (\rho - 1) u'\left(\rho \underline{y}^a\right)$; or (b) $\delta^b > \overline{\delta}^b$ and $\delta^a \rho > \delta^b + (\rho - 1) u'\left(\rho \underline{y}^a\right)$, with the thresholds $\overline{\delta}^a$, $\overline{\delta}^b$, and $\underline{y}^a$ defined in the proof.*

6.4 Intermediated Lending

So far, we have assumed banks directly undertake investments – i.e., we have consolidated the activities of deposit taking and investing into one type of agent. In reality, although banks do invest some deposits directly, other deposits are lent to borrowers who make their own investments. The reason this is particularly relevant here is that, if we introduce borrowers explicitly, one may wonder how they can credibly commit to repay a bank but not commit to repay depositors. What is the use of banks as intermediaries if depositors could lend directly to investors? One can appeal to standard theories (referenced in the Introduction), and in this Section we present our own version of the idea.

Assume that beside a and b, there is a third group, c. For the sake of illustration, all parameters are the same across groups, except $\pi^b > \pi^c > \pi^a = 0$. Also assume $\rho^a = \rho^b = 1$, and for agents 1^i and 2^i in each group i, $U^1(x, y) = u(x) - y$ and $U^2(y, x) = u(y) - x$. To incorporate lending we assume that agents in group c have a special technology $f(I)$ that requires at least $\bar{I}$ units of good y^a. Precisely, for $I \ge \bar{I}$, we have $f(I) = \alpha I$ of the same good with $\alpha > 1$, and for $I < \bar{I}$ we have $f(I) = 0$. It is therefore potentially beneficial for agents in group a to let those in group c have some y^a to invest. All is well if the minimum investment $\bar{I}$ is small. However, 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 get many 1^a agents to pool their output of y^a and lend it to someone in group c. In this case, direct lending is fine. But consider an additional friction, that agents in group c can meet at most n other agents from any group each period.[21]

In this case, direct lending may fail since 1^a would have to produce enough y^a to meet the minimum investment level, which may not be worthwhile. Also, since monitoring is so poor in group a, $\pi^a = 0$, it is impossible for them to pool their resources and have one type 1^a agent lend it all to a type 1^c agent, since the former would abscond with the proceeds. Now intermediated lending can help. In this case, a trustworthy type 1^b collects resources from many 1^a agents and makes a loan to a type 1^c investor. By delegating lending to a bank, the minimum investment level can be met more efficiently.

To formalize this, first note that $\pi^a = 0$ implies group a agents cannot consume at all unless they use deposits. In principle, 1^a could deposit resources with 1^b who could then invest them, with return ρ, but we want to consider the case where it is preferable for 1^b to lend these deposits to 1^c because $f(\cdot)$ constitutes a better investment opportunity. We first define the IF set when group a lends $d = y^a$ directly to investors in group c. The relevant incentive constraints for group a are

$$
\begin{aligned}
u(x^a) - y^a &\geq 0 \qquad (32)\\
u(\alpha y^a) - x^a &\geq 0 \qquad (33)
\end{aligned}
$$

(there is no repayment constraint since 1^a invests nothing). For group b, the relevant constraints are

$$
\begin{aligned}
u\left(y^b\right) - x^b &\geq 0\\
u\left(x^b\right) - y^b &\geq \delta^b y^b,
\end{aligned}
$$

[21] We do not regard this as particularly deep; it is a simple way to capture the notion that intermediation may be efficient using elementary search theory, as has been used in related applications (e.g., the model of middlemen in Rubinstein and Wolinsky 1987).

since participation for 1^b is implied by repayment. And for group c the relevant constraints are

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

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 agent 1^c can meet at most n agents, the minimum resources that 1^a must commit is $\bar{I}/n$ (assuming symmetry). If $u\left(\alpha\bar{I}/n\right)<\bar{I}/n$, this is too large for agents in group a to use direct lending, and the only IF allocation with direct lending implies autarky for group a. Now consider intermediated lending. The relevant constraints are still (32) and (33) for group a, while in group b they become

$$
\begin{aligned}
u\left(y^b\right)-x^b &\geq 0 \\
u\left(x^b\right)-y^b &\geq \delta^b\left(y^b+\hat{n}y^a-d+\alpha d\right).
\end{aligned}
$$

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 investors from group c. In the second subperiod, he gets αd back from these loans. 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

$$
\begin{aligned}
u\left(y^c\right)-x^c &\geq 0 \\
u\left(x^c\right)-y^c &\geq \delta^c\left(\alpha\tilde{n}d+y^c\right)
\end{aligned}
$$

where $\tilde{n}d\geq\bar{I}$, and $\tilde{n}\leq n$ is the number of bankers lending d to an investor from group c.

If $\tilde{n}>1$, the minimum investment is now $\bar{I}/\tilde{n}n<\bar{I}/n$. For $\tilde{n}$ large, $u\left(\alpha\bar{I}/\tilde{n}n\right)>\bar{I}/\tilde{n}n$, and the IF set for group a now contains points other than autarky. The smaller is δ^b, the larger we can set $\tilde{n}$. It is easy to see that if the repayment constraint in group b is not binding with $d=0$, then intermediated lending can be PE: it makes group a better off without hurting other agents. This relies on some restrictions in the meeting technology, and in particular it is not easy for a group c investor to meet a large number of group a lenders, as well as the fixed investment cost $\bar{I}$, but this seems reasonable. In reality firms may need funds beyond what a single lender is willing or able to provide. Another reason (not modeled here) is that a single lender may not want the risk exposure implied by large single investment. The bottom line is that we can, at the cost of some simplicity, extend the framework to explain how banks are useful in a situation where they intermediate between depositors and investors, based in part on their trustworthiness, and in part on other frictions.

7 Nonstationary Allocations

To this point we have assumed the mechanism is restricted to stationary allocations, which one may want to relax. The reason is this: one might imagine that using nonstationary allocations can relax repayment constraints, say by backloading rewards, and if this works well deposits may no longer be useful. A moment of reflection indicates that this is not the case, in general. Suppose $\pi^a = 0$ and $(x^*, y^*) \in \mathcal{F}_0^b$. Then the only IF allocation in group a without deposits is autarky, while the best IF allocation in group b is $\left(x_t^b, y_t^b\right) = (x^*, y^*)$ for all t. As long as the repayment constraint is not binding in group b, deposits are PE. With deposits, the optimal allocation in group a may be non-stationary, but this is beside the point – the key result is that deposits are PE. The case of $\pi^a > 0$ is a little more involved, and the best IF allocation in group a may be nonstationary, but deposits can still be PE. We now investigate this in detail using a finite-horizon version of our model.[22]

There are 2 periods, each having two subperiods. Period utility at date t is $U^1(x_t, y_t) = u(x_t) - y_t$ for type 1 and, for now, any specification $U^2(x_t, y_t)$ for type 2. Let V_t^j be the type j lifetime payoff in period t. Then at $t = 2$ we have $V_2^1 = u(x_2) - y_2$ and $V_2^2 = U^2(y_2, x_2)$, while at $t = 1$,

$$\begin{aligned} V_1^1 &= u(x_1) - y_1 + \gamma\beta\left[u(x_2) - y_2\right] \\ V_1^2 &= U^2(y_1, x_1) + \gamma\beta U^2(y_2, x_2). \end{aligned}$$

To reduce notation, set $\rho = \lambda = 1$. More substantively, with a finite horizon, pun-

[22] This is based on comments and suggestions by the referees and editor.

ishments that involve taking away future credit do not work at $t = 2$, and hence, do not work at $t = 1$. So we add an exogenous punishment: the payoff to type j agents of getting caught in a deviation is $-P_j$, where $P_1 = P > 0 = P_2$. In addition to the obvious participation constraints, we have the repayment constraints

$$\begin{aligned} \pi\gamma\beta\left[u(x_2) - y_2\right] &\geq y_1 - \pi\gamma\beta P \\ 0 &\geq y_2 - \pi P. \end{aligned}$$

Consider maximizing welfare $\mathcal{W}$ defined in (9) with $\omega_1 = 1$ and $\omega_2 = 0$. Since participation constraints for type 1 are still redundant given repayment constraints, we can maximize $\mathcal{W}$ subject to participation for type 2 and repayment for type 1:

$$U^2(y_1, x_1) + \gamma\beta U^2(y_2, x_2) \geq 0 \tag{34}$$

$$U^2(y_2, x_2) \geq 0 \tag{35}$$

$$\pi\gamma\beta\left[u(x_2) - y_2\right] + \pi\gamma\beta P \geq y_1 \tag{36}$$

$$\pi P \geq y_2. \tag{37}$$

Clearly, (34) binds. If (35)-(37) do bind, it is easy to show stationarity: $(x_1, y_1) = (x_2, y_2)$. To show how nonstationarity can arise, suppose $U^2(x, y) = y_2 - x_2$. This allows us to use the binding constraint (34) to reduce the problem to

$$\begin{aligned} \max \mathcal{W} &= u(x_1) - x_1 + \gamma\beta\left[u(x_2) - x_2\right] \\ y_2 &\geq x_2 \\ x_1 &\leq \pi\gamma\beta\left[u(x_2) + P\right] + y_2(1-\pi)\gamma\beta - \gamma\beta x_2 \\ \pi P &\geq y_2. \end{aligned}$$

Since y_2 does not affect $\mathcal{W}$, and increasing y_2 slackens the first two constraints, to solve this, we start by making y_2 as big as possible without violating the third constraint: $y_2 = \pi P$. Then the problem becomes

$$\max \mathcal{W} = u(x_1) - x_1 + \gamma\beta\left[u(x_2) - x_2\right] \tag{38}$$

$$x_1 \leq \pi\gamma\beta\left[u(x_2) - x_2\right] + (2-\pi)\pi\gamma\beta P \tag{39}$$

$$x_2 \leq \pi P. \tag{40}$$

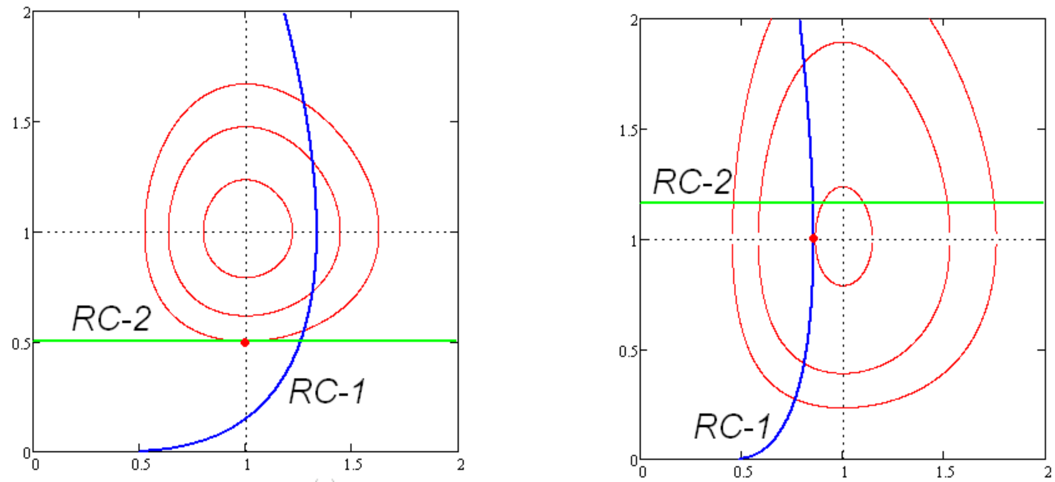

Figure 6: Left: RC-1 slack and RC-2 binds. Right: RC-1 binds and RC-2 slack.

We illustrate the results in Figures 6-7, where first repayment constraint (39) is satisfied to the left of the blue curve labeled RC-1, and the second (40) is satisfied below the green line labeled RC-2.[23] It is possible that both constraints are slack, whence the solution is $(x_1, x_2) = (x^*, x^*)$, as in the left panel of Figure 7, drawn for high values of πP and $\gamma\beta$. It is also possible to have (39) bind but not (40), as in the right panel of Figure 6, drawn for a relatively low πP and high $\gamma\beta$, making it easier to enforce repayment in the first subperiod and harder in the latter. In this case $x_1 = x^*$ and $x_2 < x^*$. It is also possible to have (40) bind but not (39), as in the left panel of Figure 6, drawn for higher πP and lower $\gamma\beta$. In this case $x_1 < x^*$ and $x_2 = x^*$. The final case in the right panel of Figure 7 has both constraints binding, but this is nongeneric.

So, it is possible to have x_1 either above or below x_2, and generally it is restrictive to impose $x_1 = x_2$. We imposed this in the baseline model because it greatly simplifies the analysis. One may regard this as a first step towards understanding this kind of theory. But there is no presumption that our substantive results rely on this simplifying restriction. It is important to emphasize that, in the finite case, with

[23] Here we use the following easily-verified properties of the problem: the indifference curves in (x_1, x_2) space are centered around (x^*, x^*), where $u'(x^*) = 1$; these curves are horizontal when $x_1 = x^*$ and vertical when $x_2 = x^*$; the curve defined by constraint (39) also has a slope that is vertical when $x_2 = 1$; and constraint (40) is a horizontal line at $x_2 = \pi P$.

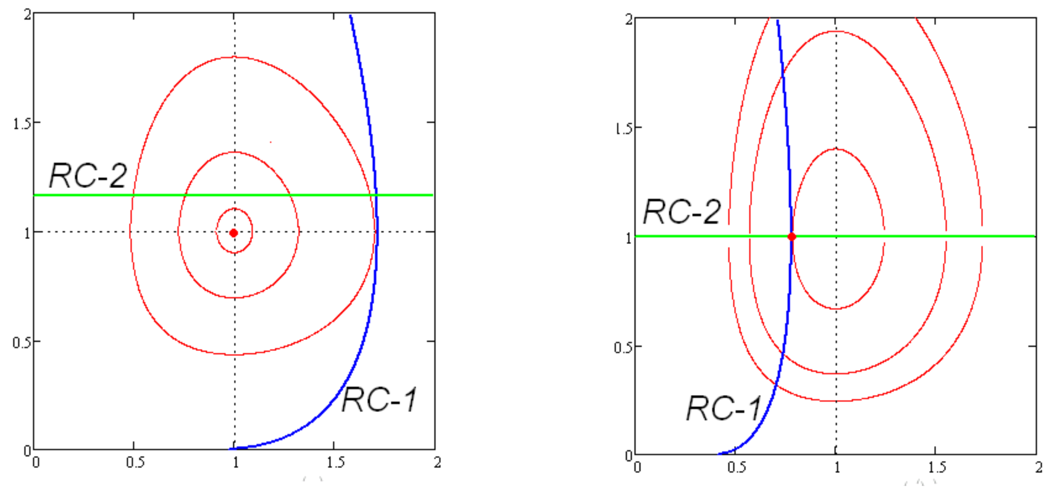

Figure 7: Left: RC-1 and RC-2 slack. Right: RC-1 and RC-2 bind (nongeneric).

nonstationary allocations, a role for banking is still present. Consider two groups, a and b, where the repayment constraints are binding for a but not b. Deposits are still essential: we improve the IF set by having 1^a deposit produce with 1^b, as the former uses claims on the latter to facilitate trade with 2^a. Type 2^a are more willing to trust the promises of 1^a that 1^b because, in this case, 1^a have a greater incentive to honor their obligations given the values of β, γ, π or P (and, in general, also λ and ρ, although we set $\lambda = \rho = 1$ in this section). So in the big picture, without stationarity, our theory of banking based on commitment goes through. While there are reasons to prefer the infinite horizon – in particular, one might like punishments to be endogenous – we can certainly illustrate the main ideas in the finite version.[24]

[24] In general, the advantage of allowing nonstationarity is that we can increase the incentive to behave well by pushing rewards into the future – backloading. This is because, when we give type 1 a big x today, after he consumes it the rerward is *sunk* and has no further incentive effects. Giving him less today and more in the future encourages good behavior today and tomorrow. Mitigating against this is the fact that agents like smooth consumption. This is generally a hard problem. See Burdett-Coles (2003,2010), e.g., for complications that arise in the context of Burdett-Mortensen (1998). We hope our model, even with the stationarity restriction, is still interesting, the way the large literature is interesting on stationary versions of Burdett-Mortensen (where firms offer constant wages rather than wage-tenure contracts). Moreover, based in part on Burdett-Coles, it is reasonable to conjecture the following. While backloading implies we want to push rewards to the future, we cannot do this indefinitely. So, rewards should rise over time, but they are bounded by the physical environment. Hence we think that the allocation may converge over time to exactly the stationary outcome studied here. Investigating this is left for future research.

8 A Digression on History

We have established that, because of incentive issues, it can be beneficial for a party who wants something from a second party to deposit resources with a third party – an intermediary – who invests on his behalf until the resources are needed for trade (in the formal model, the resources are deposited and used for trade in the same subperiod, but it is obviously possible, and potentially desirable, to generalize the setup so that type 1 produces and deposits at one point in time, then trades using the deposits at some later, possibly random, point in time, just as it is possible to generalize the setup so that type 2 not necessarily cash in his claims to the deposits in the very next subperiod). The reason this can be efficient is that the third party may be more trustworthy, in terms of honoring obligations, than the first party. He can be more trustworthy because he is more patient, he is more visible, he has more at stake in the economic system, or because his gains from liquidating the investment and absconding with the returns are lower. This arrangement can be efficient, even if the third party does not have access to the best investments. We say this resembles banking, and here we go into more detail.

We begin by recalling that, although the deposit receipts or banknotes discussed above constitute inside money, our theory by assumption involves no outside money. While it may be interesting to include outside money in future work, from the historical perspective, institutions that accepted deposits in goods came long before the invention of coinage in 7th century Lydia. In ancient Mesopotamia and Egypt, mainly for security and to economize on transportation costs, goods were often deposited in palaces and temples and, in later periods, also private houses. As Davies (2002) describes the situation:

> 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.

We think it is interesting that deposit receipts so long ago led to transfers to the order of third parties, so they could facilitate transactions and payments, just like in the model. Moreover, in his detailed description of medieval Venetian bankers, Mueller (1997) describes the practice of accepting two types of deposits: regular deposits, which were specific goods that bankers had to deliver on demand; and irregular deposits, involving specie or coins that only had to be repaid with the same value but not the same specie or coins. The depositor making an irregular deposit tacitly agreed to the investment by the banker of the deposits. Like in modern times, when you put money in the bank you do not expect to withdraw the same money later, only something of some specific value. This is true in the model, too: the liability of the bank is not the deposit, per se, but claims on the returns to investment.[25]

Many regard the English goldsmiths as the first modern bankers. Originally, their depositors were again mainly interested in safe keeping, which is a simple type of investment. But early in the 17th century their deposit receipts began circulating in place of money for payment purposes – the first incarnation of British banknotes – and shortly thereafter they allowed deposits to be transferred by "drawn note" or cheque. Nice discussions of the English goldsmith bankers can be found in *Encyclopedia Britannica* (we looked at the 1941 and 1954 editions). For more specialized treatments, see Joslin (1954), Quinn (1997) and Selgin (2010). Although many call the goldsmiths the first modern bankers, others mention the Templars (see Weatherford 1997 or Sanello 2003). During the crusades, because of their skill as warriors, these knights became specialists in protecting and moving money and other valu-

[25] Because they are making investments, our bankers are more than pure storage facilities. We mention this because Chris Phelan in a comment on an earlier version of the paper said that, according to our theory, coat-check girls at restaurants are bankers. Not quite. When you leave your hat at the coat-check you expect the same hat back; when you deposit resources at your bank you expect something different. Still, most people agree the origin of deposit banking did have something to do with safe keeping, something like a coat-check.

ables. At some point, rather than e.g. shipping gold from point A to point B for one party and shipping different gold from point B to point A for another, they saved on security and transportation costs simply by reassigning the parties' claims to gold in different locations. It is less clear, however, if their liabilities circulated as a medium of exchange the way goldsmiths' receipts did.

It is also interesting to note that other institutions that engaged in the type of banking we have in this paper – accepting deposits of goods that facilitated other transactions – were still common after the emergence of modern banks. In colonial Virginia, tobacco was commonly used in transactions because of the scarcity of precious metals (Galbraith 1975). The practice of depositing tobacco in public warehouses and then exchanging authorized certificates, attesting to its quality and quantity, was extremely common and survived for over 200 years. Similarly, in the 19th century, to facilitate transactions and credit arrangements between cocoon producers and silk weavers, warehouses were established that stored dried cocoons or silk and issued warrants that could be used to pledge for credit. As discussed by Federico (1997), the first of these warehouses was funded by a group of entrepreneurs in Lyons in 1859. The Credit Lyonnais established its own warehouse in 1877 and was soon imitated by a series of Italian banks.

The main thing to take away from these examples is that an early development in the evolution of banking was that deposits came to be used to facilitate exchange. As in the model, throughout history, a second party is more likely to give you something if you can use in payment the liability of a credible third party, rather than your own promise. As we said, notes, cheques, debit cards and related instruments issued by commercial banks have this feature. Returning to Venice, Mueller (1997) explains how deposit banking came to serve "a function comparable to that of checking accounts today; that is, it was 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."[26]

[26] According to some, these early deposits did not actually circulate, in the sense that transferring

This system only works well if bankers are relatively trustworthy. Our theory says that the more patient or visible an agent is, or the more he has at stake, the more credible he becomes. The Rialto banks in medieval Venice 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 compagnia – 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 1997, p. 97). It is also interesting to point out that, although direct evidence is scant, Venetian bankers seem to have been subject to occasional monitoring as in the model: "In order to maintain 'public faith,' the Senate in 1467 reminded bankers of their obligation to show their account books to depositors upon request, for the sake of comparing records." (Mueller 1997, p. 45). While it may have been prohibitively costly for depositors to continuously audit the books, one can imagine monitoring every so often. And if caught cheating, the punishment was indeed lifetime banishment from any banking activity in Venice, although apparently this happened rarely in history (as in the theory).[27]

We also mention that many bankers historically started as merchants, who almost by definition have a greater connection to the market than a typical individual. As Kohn (1999) describes it, 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

funds from one account to another "generally required the presence at the bank of both payer and payee" (Kohn 1999). This is the argument for regarding the goldsmiths the first modern bankers. See also Quinn (2002). But even if they did not circulate, in this sense, these deposits cleary still facilitated payments. And Spufford (1988) documents that the Florentines were already using cheques in 1368.

[27] We think it is obvious that visibility and monitoring have always been crucial for good banking, but if one wants more evidence, 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."

small investors. They were not the wealthiest group; wealth then was concentrated in the hands of landowners, who controlled agriculture, forests, and mineral rights. But the merchants arguably had the most to lose from reneging on obligations. Thus, "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 evidence on the first bankers being individuals who had a great connection to the market is given by Pressnell (1956) in its study of the origins of country banking in England during the Industrial Revolution. Almost all of the early country banks grew up as a by-product of some other main activity, usually some kind of manufacturing.

Also, 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. Their organizational form was generally either that of the fraterna or that of the partnership, the latter often concluded between a citizen and a noble. (Mueller 1997, p. 82)

As in our model, there seem to have been interesting issues concerning the efficient number of bankers, and revolving around greater credibility or commitment and larger amounts of deposits per bank. While we do not claim to be experts on the history of banking, in general, we hope that all of these examples illustrate how our theory is based not only on assumptions that seem to make common sense, but that are also consistent with the more-or-less objective record.[28]

[28] Of course, such a cursory examination of the extensive history of banking has to be selective, and one can always try to pick and choose to make one's case. Indeed, as a referee pointed out, it is perhaps not clear from the history if, e.g., causation runs from our criteria for good bankers to

Finally, moving to more recent history, what does our theory say about banking panics, and the recent financial crisis in particular? Gorton (2009) argues that the recent crisis was a wholesale panic, whereby some financial firms ran on others by not renewing sale and repurchase agreements. This resembles a retail panic in which customers withdraw demand deposits. By analogy, depositors in the current crisis were firms that lent money in the repo market. The location of subprime risks among their counterparties was unknown, depositors were confused about which counterparties were really at risk, and consequently ran all banks. While our framework is too simple to capture all the intricacies, we can use it to highlight some issues. Suppose the probability of being active each period γ is subject to shocks, where uncertainty surrounding these shocks could induce agents to not renew their deposits (or deposit less) to re-establish incentives. Such shocks may depend on the nature of the firm's business – say, γ could be affected by the housing market if that business involves originating mortgage loans; or it could be affected by political events. Generally, whenever visibility goes down or monitoring becomes more difficult, theory predicts that credit is hindered. But this is efficient: when π falls, not only can credit dry up, it *should* dry up. We are not arguing that recent events were unproblematic. We are suggesting that it may be interesting to look at them through the lens of mechanism design.

9 Conclusion

This has been an attempt to study banking with minimal assumptions about what banks are or what they do. We specified preferences, technologies, and certain frictions, including commitment issues and imperfect monitoring. We then examined feasible or efficient outcomes, and interpreted them in terms of banking arrangements. It can be desirable for certain agents, chosen endogenously, to perform functions commonly associated with banks – they accept deposits, they invest or make loans to other investors, and their liabilities facilitate exchange. This can be efficient even

activity in banking, or the other way around – did the Medici family get into banking because they were heavily invested in the market, or vice versa? This is, to us, an interesting open question.

when bankers do not have access to the best investment opportunities, if they are more trustworthy, which here means they are more patient, more visible, or more connected to the economic system. Other things equal, it is better if bankers have good investment opportunities, but it can be efficient to sacrifice rate of return for trustworthiness. This resembles salient aspects of banking in modern and historical contexts. And this arrangement is essential: if we were to rule it out, the set of feasible allocations would be inferior.

We found mechanism design useful for thinking about these issues. One can alternatively study equilibria in this environment under particular pricing mechanisms, as in Gu and Wright (2010). With either approach, the model captures in a tractable way interesting aspects of intertemporal exchange, like imperfect commitment and collateral. It is tractable mainly because much of the interesting activity takes place across subperiods within a period, making the analysis similar to a two-period model, yet is genuinely dynamic and makes use of the infinite horizon to endogenize punishments and hence credit constraints. The model can be generalized in many directions. It may be desirable to add uncertainty, perhaps private information, potentially giving rise to additional functions for banks discussed elsewhere like diversification and information processing. We abstracted from these to focus on other issues. Of course our approach captures some, but not all, nuances of banking. For instance, it is silent on the fact that bank liabilities typically have shorter duration than their assets. While the current specification does not deliver everything a complete theory of banking should, one can certainly try to extend it in any number of directions to deliver more. And we think it does deliver some interesting results not in the existing literature.

Appendix

Proof of Lemma 2: From (10), y is a function of x with

$$\frac{dy}{dx}=-\frac{\rho U_1^1\left(U_{22}^2U_1^2/U_2^2-U_{12}^2\right)-U_2^2\left(U_{11}^1U_2^1/U_1^1-U_{21}^1\right)}{\rho\left[U_1^2\left(U_{22}^1U_1^1/U_2^1-U_{12}^1\right)-U_2^1\left(U_{11}^2U_2^2/U_1^2-U_{12}^2\right)\right]}.$$

When all goods are normal, the four terms in parentheses are positive. Remembering that $U_1^1>0$, $U_2^1<0$, $U_1^2<0$ and $U_2^2>0$, we are done. ■

Proof of Lemma 3: Define the Lagrangian

$$\mathcal{L}=\omega_1U^1(x,y)+\omega_2U^2(\rho y,x)+\eta U^2(\rho y,x)+\varphi\left[U^1\left(x,y\right)-\delta\rho y\right] \tag{41}$$

where η and φ are multipliers. The FOCs are

$$\begin{aligned}\omega_1U_1^1+\omega_2U_1^2+\eta U_1^2+\varphi U_1^1 &= 0\\ \omega_1U_2^1+\omega_2\rho U_2^2+\eta\rho U_2^2+\varphi\left(U_2^1-\delta\rho\right) &= 0\end{aligned}$$

plus the constraints. Rearranging implies

$$\frac{\left(\omega_1+\varphi\right)U_1^1}{\left(\omega_1+\varphi\right)U_2^1-\varphi\delta\rho}=\frac{\left(\omega_2+\eta\right)U_1^2}{\rho\left(\omega_2+\eta\right)U_2^2}.$$

If (6) is not binding then $\varphi=0$, and $\rho U_1^1/U_2^1=U_1^2/U_2^2$, which means $(\hat{x},\hat{y})\in\mathcal{P}$. ■

Proof of Proposition 4 :Since $\gamma^b>\gamma^a$, it must be that $U^1\left(x^b,y^b\right)\geq U^1\left(x^a,y^a\right)$. With deposits d, and since there is one candidate banker in each group, the repayment constraint in group b becomes $-\lambda^b\rho\left(y^b+d\right)+p^b\gamma^b\frac{\beta}{1-\beta}U^1\left(x^b,y^b\right)=0$. Therefore, we obtain

$$\frac{\partial p^b}{\partial d}=\frac{1-\beta}{\beta}\frac{\lambda^b\rho}{\gamma^bU^1\left(x^b,y^b\right)}.$$

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

$$\frac{\partial p^a}{\partial d}=-\frac{1-\beta}{\beta}\frac{\lambda^a\rho}{\gamma^aU^1\left(x^a,y^a\right)}.$$

Therefore, increasing deposits from group a to b reduces the overall monitoring cost $p^ak^a+p^bk^b$ since

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

where the inequality follows from $U^1\left(x^a,y^a\right)\leq U^1\left(x^b,y^b\right)$, $\gamma^a\leq\gamma^b$ 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 $(\breve{x}^a, \breve{y}^a)$ solve $\max_{x,y} \mathcal{W}^a(x, y)$, subject to the participation constraint for 2^a only. If

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

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

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

This completes the proof. ■

Proof of Proposition 7: Given $\rho > 1$, we can have $\overline{\delta}^a < \overline{\delta}^b$, so that $(x^{*,a}, y^{*,a})$ is feasible in group a but $\left(x^{*,b}, y^{*,b}\right)$ is not feasible in group b. Here we focus on the case where deposits in group b are PE. The condition $\delta^a > \overline{\delta}^a$ 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 $\delta^b \leq \overline{\delta}^b$, 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 $\delta^a \rho > (\rho - 1) u' \left(\rho \underline{y}^a\right)$ 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 $\delta^b > \overline{\delta}^b$. 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 $\delta^a \rho > \delta^b + (\rho - 1) u' \left(\rho \underline{y}^a\right)$, which is stricter than case (a). Finally, if the commitment problem in group a is very severe, $u' \left(\rho \underline{y}^a\right)$ 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 (**??**). 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 (**??**) at equality. Define $\overline{\delta}^i$ by $\left[u(\rho^i y^{*,i}) - y^{*,i}\right] / \left(\rho^i y^{*,i}\right) = \delta^i$ as the level below which (**??**) binds in group i. The next two claims establish the result.

Claim 1 *Deposits in group b are PE if*

$$\delta^a > \overline{\delta}^a, \; \delta^b \leq \overline{\delta}^b \text{ and } \delta^a \rho > (\rho - 1) u' \left(\rho \underline{y}^a\right).$$

Proof: Given x_2^a and d, type 1^a has to produce y^a such that $x_2^a = (y^a - d)\,\rho + d$. The repayment constraint is

$$u\left[(y^a - d)\,\rho + d\right] - \frac{y^a}{\rho} \geq \delta^a \rho\,(y^a - d)\,. \tag{42}$$

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

$$\begin{aligned}(1-\rho)\,u'\left[(y^a - d)\,\rho + d\right] + \delta^a\rho &> 0\\ \delta^a\rho &> (\rho - 1)\,u'\left[(y^a - d)\,\rho + d\right]\end{aligned}$$

So $d > 0$ is PE at $\underline{y}^a$ iff

$$\delta^a\rho > (\rho - 1)\,u'\left(\rho\underline{y}^a\right)$$

This establishes the claim. ■

Claim 2 *Deposits in group b are PE if*

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

Proof: When $\delta^b > \overline{\delta}^b$, the solution to (**??**) 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 $u(\underline{y}^b) - \underline{y}^b + \tau \geq \delta^b\left(\underline{y}^b + d\right)$. 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

$$u\left[(y^a - d)\,\rho + d\right] - \frac{y^a}{\rho} - \tau \geq \delta^a\rho\,(y^a - d)\,. \tag{43}$$

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

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

so the repayment constraint is relaxed whenever

$$\delta^a\rho - \delta^b \geq (\rho - 1)\,u'\left[(y^a - d)\,\rho + d\right].$$

Evaluating at $\underline{y}_2^a$,

$$\delta^a\rho - \delta^b \geq (\rho - 1)\,u'\left(\rho\underline{y}_2^a\right).$$

This establishes the claim and concludes the proof of Proposition 7. ■

References

[1] Alvarez, F. and U. Jermann (2000) "Efficiency, Equilibrium, and Asset Pricing with Risk of Default," *Econometrica* 68, 775-798.

[2] Andolfatto, D. (2008) "The Simple Analytics of Money and Credit in a Quasi-linear Environment," Mimeo.

[3] Andolfatto, D. and E. Nosal (2009) "Money, Intermediation and Banking" *Journal of Monetary Economics* 56, 289-294.

[4] Boyd, J. and E. Prescott (1986) "Financial Intermediary Coalitions," *Journal of Economic Theory* 38, 211-232.

[5] Burdett, K. and Coles M. (2003). "Equilibrium Wage-Tenure Contracts," *Econometrica*, 71(5), 1377-1404.

[6] Burdett, K. and Coles, M. (2010). "Wage/tenure contracts with heterogeneous firms," *Journal of Economic Theory*, 145(4), 1408-1435.

[7] Calomiris, C. and C. Kahn (1991) "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *American Economic Review* 81, 497-513.

[8] Cavalcanti, R. and N. Wallace (1999a) "A Model of Private Bank Note Issue," *Review of Economic Dynamics* 2(1), 104-136.

[9] Cavalcanti, R. and N. Wallace (1999b) "Inside and Outside Money as Alternative Media of Exchange," *Journal of Money Credit and Banking* 31(3), 443-457.

[10] Corbae, D., T. Temzelides and R. Wright (2003) "Directed Matching and Monetary Exchange," *Econometrica* 71, 731-756.

[11] Davies, G. (2002) *A History of Money From Ancient Times to the Present Day*, 3rd. ed. Cardiff: University of Wales Press.

[12] Diamond, D. and P. Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* 91, 401-419.

[13] Diamond, D. (1984) "Delegated Monitoring," *Review of Economics Studies* L1, 393-414.

[14] Diamond, D. and R. Rajan (2001) "Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking," *Journal of Political Economy* 109, 287-327.

[15] Ennis, H. and T. Keister (2009) "Bank Runs and Institutions: The Perils of Intervention," *American Economic Review* 99, 1588-1607.

[16] Federico, G. (1997) *An Economic History of the Silk Industry, 1830-1930.*Cambridge: Cambridge University Press.

[17] Freixas, X. and J.C. Rochet (2008) *Microeconomics of Banking.* Cambridge: MIT Press.

[18] Ferguson, N. (2008) *The Ascent of Money.* London: Penguin.

[19] Galbraith, K. (1967) *The New Industrial State.* Princeton: Princeton University Press.

[20] Gorton, G. and A. Winton (2002) "Financial Intermediation," in *Handbook of the Economics of Finance*, ed. by Constantinides et al. Amsterdam: North Holland.

[21] Gorton, G. (2009) "Slapped in the Face by the Invisible Hand: Banking and the Panic of 2007," Mimeo.

[22] Gu, C. and R. Wright (2010) "Endogenous Credit Cycles," Mimeo.

[23] He, P., L. Huang and R. Wright (2005) "Money and Banking in Search Equilibrium" *International Economic Review* 46, 637-670.

[24] He, P., L. Huang and R. Wright (2008) "Money, Banking and Monetary Policy" *Journal of Monetary Economics* 55, 1013-24.

[25] Hicks, J.R. (1935) "A Suggestion for Simplifying the Theory of Money," *Economica* 2, 1-19.

[26] Huangfu, X. and H. Sun (2008) "Private Money and Bank Runs," *Canadian Journal of Economics,* forthcoming.

[27] Julien, B., J. Kennes and I. King (2008) "Bidding For Money," *Journal of Economic Theory* 142, 196-217.

[28] Joslin, D.M. (1954) "London Private Bankers, 1720-1785" *Economic History Review* 7, 167-86.

[29] Kehoe, T.J., and D.K. Levine (1993) "Debt-Constrained Asset Markets,"*Review of Economic Studies* 60, 865-88

[30] Kehoe, T.J., and D.K. Levine (2001) "Liquidity Constrained vs. Debt Constrained Markets,"*Econometrica* 69, 575-598.

[31] Kiyotaki, N. and J. Moore (2005) "Financial Deepening," *Journal of the European Economic Association* 3, 701-712.

[32] Kiyotaki, N. and J. Moore (2008) "Liquidity, Business Cycles, and Monetary Policy," Mimeo.

[33] Kiyotaki, N. and R. Wright (1989) "On Money as a Medium of Exchange," *Journal of Political Economy* 97, 927-954.

[34] Kiyotaki, N. and R. Wright (1993) "A Search-Theoretic Approach to Monetary Economics," *American Economic Review* 83, 63-77.

[35] Kocherlakota, N. (1998) "Money is Memory," *Journal of Economic Theory* 81, 232-251.

[36] Kocherlakota, N. and N. Wallace (1998) "Incomplete Record-Keeping and Optimal Payment Arrangements," *Journal of Economic Theory* 81, 272-289.

[37] Koeppl, T., C. Monnet and T. Temzelides (2008) "A Dynamic Model of Settlement", *Journal of Economic Theory* 142, 233-246.

[38] Kohn, M. (1999) "Early Deposit Banking," Working Paper 99-03, Department of Economics, Dartmouth College.

[39] Kreps, D.M. (1990) *A Course in Microeconomic Theory*, Princeton: Princeton University Press.

[40] Lagos, R. and R. Wright (2005) "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy* 113, 463-484.

[41] Leland, H.E and D.H. Pyle (1977) "Informational Asymmetries, Financial Structure and Financial Intermediation," *Journal of Finance* 32, 371-387.

[42] Mailath, G.J. and Samuelson L. (2006) *Repeated Games and Reputations: Long-Run Relationships*. Oxford: Oxford University Press.

[43] Mattesini, F., C. Monnet and R. Wright (2009) "Banking: A Mechanism Design Approach" Working Paper, Federal Reserve Bank of Philadelphia.

[44] Mills, D.C. (2008) "Imperfect Monitoring and the Discounting of Inside Money," *International Economic Review* 49.

[45] Mueller, R.C. (1997) *The Venetian Money Market, Banks, Panics, and the Public Debt, 1200-1500*. Baltimore: Johns Hopkins University Press.

[46] Myers, S.C. and R. Rajan (1998) "The Paradox of Liquidity," *Quarterly Journal of Economics*, 113 (3), 733-771.

[47] Nosal, E.and G. Rocheteau (2009) *Money, Payments, and Liquidity*. Cambridge: MIT Press.

[48] Orsingher, R., (1967) *Banks of the World*. New York: Walker and Company.

[49] Quinn, S. (1997) "Goldsmith-Banking: Mutual Acceptance and Interbanker Clearing in Restoration London," *Explorations in Economic History* 34, 411-42.

[50] 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: Cambridge University Press.

[51] Rubinstein, A. and A. Wolinsky (1987) "Middlemen," *Quarterly Journal of Economics* 102, 581-593.

[52] Rajan, R. (1998) "The Past and Future of Commercial Banking Viewed through an Incomplete Contract Lens," *Journal of Money, Credit, and Banking* 30 (3), 524-550.

[53] Sanches D. and S. Williamson (2010) "Money and Credit with Limited Commitment and Theft," *Journal of Economic Theory* 145.

[54] Sanello, F. (2003) *The Knights Templars: God's Warriors, the Devil's Bankers*. Lanham: Taylor Trade Publishing.

[55] Selgin, G. (2006) "Banking", written for *Encyclopedia Britannica*.

[56] Selgin, G. (2010) "Those Dishonest Goldsmiths," Mimeo.

[57] Spufford, P. (1988) *Money and Its Uses in Medieval Europe*. Cambridge: Cambridge University Press.

[58] Townsend, R. (1987) "Arrow-Debreu Programs as Microfoundations of Macroeconomics," *Advances in Economic Theory, Fifth World Congress. Econometric Society Monographs Series* 12, ed. by T. Bewley. Cambridge University Press.

[59] Townsend, R. (1988) "Models as Economies," *Economic Journal* 98, 1-24.

[60] Townsend, R. (1990) *Financial Structure and Economic Organization*. Basil Blackwell.

[61] 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.

[62] Wallace, N. (2005) "From Private Banking to Central Banking: Ingredients of a Welfare Analysis," *International Economic Review* 46, 619-636.

[63] Wallace, N. (2010) "The Mechanism Design Approach to Monetary Theory," in *Handbook of Monetary Economics, 2nd Edition*, ed. by B. Friedman and M. Woodford.

[64] Weatherford, J. (1997) *The History of Money*. New York: Crown Publishers.

[65] Williamson, S. (1986) "Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing," *Journal of Monetary Economics* 18, 159-179.

[66] Williamson, S. (1987a) "Financial Intermediation, Business Failures, and Real Business Cycles," *Journal of Political Economy* 95, 1196-1216.

[67] Williamson, S. (1987b) "Recent Developments in Modeling Financial Intermediation," *Federal Reserve Bank of Minneapolis Quarterly Review* 1132.