A Theory of Money and Banking

David Andolfatto
Simon Fraser University

Ed Nosal
Federal Reserve Bank of Cleveland

October 16, 2002

Abstract

We construct a very simple environment where an institution, which closely resembles a bank, arises endogenously. In our model a bank is an agent that: (1) creates money (a debt instrument that circulates as a means of payment); (2) lends it out (swapping it for less liquid forms of debt); (3) monitors those agents in control of the capital backing the illiquid debt; and (4) collects on money loans as they come due. Bank money in our model is a debt instrument that embeds within it important stipulations that are found in actual private money instruments. We show that when trade is decentralized, a Pareto efficient allocation can only be achieved if a bank is present. Our model goes some way in addressing the questions of why private money takes the form that it does, as well as why private money is typically supplied by banks.

1 Introduction

This article presents a simple environment that endogenously gives rise to “banks” that, among other things, create and lend out “money.” We define “money” to be any object that circulates widely as a means of payment. In our model, this object takes the form of a fully-secured and redeemable bearer bond. This monetary instrument is issued by an agent that can credibly commit to monitoring a pool of real investments; that is, this capital forms the requisite backing for a circulating private debt instrument. While direct trade in non-monetary securities is feasible in our model, we find that money can economize on monitoring costs, which enhances the efficiency of the exchange process.

In our model a “bank” is an agent that simultaneously issues money and monitors investments. In reality, banks also accept deposits of money, which are then redirected to borrowers. In our model banks do not accept deposits; we do not view this function as a defining characteristic of a bank.1 In particular, financial markets also accept “deposits” of money in exchange for marketable

1Needless to say, there are those who would disagree with this point of view. To our knowledge, Bullard and Smith (2001) have the only model featuring intermediaries that simultaneously take deposits, make loans, and issue circulating liabilities.
liabilities (equity and debt instruments). We think that banks differ from financial markets in two ways. First, bank liabilities are designed to be “high-velocity” payment instruments (money). Second, banks specialize in screening and monitoring their investments. Banks in our model perform both of these functions.

While our framework allows private “non-bank” liabilities to serve as the economy’s medium of exchange (as mentioned earlier, exchange is even possible without any money at all), we demonstrate that the cost-minimizing structure has a “bank” creating “liquid” funds, which are then lent to borrowers (for example, entrepreneurs) with suitable collateral (contingent claims against future output). These liquid funds constitute “real bills of exchange;” that is, they are backed by the issuing bank with enforceable claims against real assets (the collateral supplied by borrowers).

In reality, the vast bulk of “the” money supply consists of private debt instruments with contractual features similar to those embedded in the debt instruments that circulate in our model. In addition, the bulk of this money is created by “banks,” i.e., institutions that spend considerable resources monitoring their investment portfolio. Thus, our model goes some way in addressing the questions of why private money takes the contractual form it does, as well as why private money is typically supplied by “banks” (as opposed to other types of private agencies). In our model, money and banking are inextricably linked.

The institutions of “money and banking” in our model are not essential in the sense that there exist other organizational forms that are able to implement the same (efficient) allocation. For example, a centralized trading system, in which all trade is mediated by one agent, can implement the efficient allocation. However, when trade is decentralized, i.e., all trade is mediated by markets, then the institutions of money and banking are essential. That is, in a decentralized environment if the institutions of money and banking are not operative, then a Pareto efficient allocation of resources is not achievable. One can interpret our main result as being a Second Welfare Theorem for economies with asymmetric information. And one of our main messages is that markets by themselves are unable to implement the efficient allocations; markets need to be supplemented with certain institutions in order to achieve an efficient allocation.

Of course, we are not the first to explicitly model money and banking together. Some important recent contributions include Kiyotaki and Moore (2000), Bullard and Smith (2001), and Cavalcauti (2001). We view our work as complementary to these papers. In Kiyotaki and Moore (2000), banks are agents endowed with some sort of commitment technology that allows their liabilities to circulate. In Cavalcauti (2001), banks have verifiable histories but non-banks have not; as in Kiyotaki and Moore (2000), this special feature of banks allows their liabilities to circulate. In Bullard and Smith (2001), the pattern in which agents meet endows bank liabilities with relatively low transactions costs, making these instruments the preferred medium of exchange. Our setup is similar to that of Williamson (1986), which emphasizes the role of monitoring in the
business of banking; banks are endowed with no special characteristics relative to other agents in the economy.

2 The Physical Environment

Our framework utilizes an intertemporal version of Wicksell's triangle along the lines of Kiyotaki and Moore (2000). In particular, consider an economy with four periods, indexed by \( t = 0, 1, 2, 3 \). The economy is populated by a large (but finite) number (3N) of individuals who have linear preferences defined over time-dated (and possibly stochastic) consumption profiles \((c_1, c_2, c_3)\). We assume that there are three types of individuals, labelled A, B, C, with \( N \) individuals of each type. Individuals are specialized in the production of a nonstorables time-dated good \( y_t, t = 1, 2, 3 \): Type A individuals produce \( y_3 \); type B individuals produce \( y_1 \); and type C individuals produce \( y_2 \). We find it useful to think of these different types of people as being endowed with either short, medium, or long-term capital projects.

For simplicity only, we assume that individuals have specialized preferences: Person A desires \( c_1 \); person B desires \( c_2 \); and person C desires \( c_3 \). We also assume that each person values his own production good just a little bit. That is, for the fixed consumption bundle \((c_1, c_2, c_3)\), person A values it according to \( c_1 + \epsilon c_3 \); person B values it according to \( c_2 + \epsilon c_3 \); and person C values it according to \( c_3 + \epsilon c_2 \), where \( \epsilon \to 0 \) is a preference parameter. The table below describes, for each type of person, the good that he (really) desires, \( c_i \), and the good that he produces \( y_j \).

<table>
<thead>
<tr>
<th>Good</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( c_1 )</td>
<td>( y_1 )</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( c_2 )</td>
<td>( y_2 )</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( y_3 )</td>
<td></td>
<td>( c_3 )</td>
</tr>
</tbody>
</table>

Note that there is a complete lack of double-coincidence of wants: There are no gains from trade for any bilateral pairing of individuals. (There is still a complete lack of double-coincidence of wants even when one takes into account that each person attaches a small weight, \( \epsilon \), to the good that he produces).\(^2\)

Unlike many models that give rise to a demand for money, there is no spatial or temporal separation of agents that limits communication; all agents can potentially trade with each other at each of the four dates. We will think of period 0 as an \textit{ex ante} “contracting period” where individuals can meet and negotiate contracts before any uncertainty is resolved, but where no consumption or production actually takes place. In subsequent periods, output is produced and ultimately consumed. Unlike Kiyotaki and Moore’s (2000) motivation for

\(^2\)It should be pointed out, however, that the lack of double coincidence is not necessary to deliver our main results.
circulating private money, we assume that all agents are able to commit fully to negotiated agreements. In particular, we assume that there exists a contract enforcement mechanism (e.g., a court of law) that operates at zero resource cost to enforce verifiable aspects of privately negotiated contracts.

Suppose that there is risk associated with production. In particular, suppose that in each period only a fraction $1 - \lambda$ of the economy’s projects will succeed in producing an output level $y > 0$, with the remaining fraction $\lambda$ of projects failing to produce any output at all. Although there is no uncertainty over economy wide production—which is equal to $(1 - \lambda)Ny$ in each period—an agent does not know ex ante whether his project will be “good” or “bad.” As a result, each individual views his date $t$ project as being risky, where the payoff to the project, $y_t$, is given by

$$y_t = \begin{cases} y & \text{with probability } 1 - \lambda; \\ 0 & \text{with probability } \lambda. \end{cases}$$

As in Diamond (1984), Gale and Hellwig (1984) and Townsend (1979), the key friction in our model is the existence of private information, which is defined over the outcome of an agent’s project. Specifically, an agent can costlessly observe whether his own project is successful or not, but other individuals cannot. Other individuals can, however, observe the outcome if they incur a cost. One can interpret this cost as representing the effort exerted in auditing the project return, where the utility cost incurred for undertaking an audit is equal to $\mu$. As in Townsend (1979) and Gale and Hellwig (1984), an audit occurs (if at all) after project returns are realized.\(^3\) We assume, as in Diamond (1984) and Williamson (1986), that the information acquired in the audit remains private information (i.e., it cannot be sold to other parties).\(^4\)

The timing of events in the model is as follows. At date 0, agents can (if they wish) write and trade securities contracts. At each subsequent date $t \geq 1,$

1. agents who are scheduled to produce at date $t$ learn whether or not their projects are successful.
2. agents have the opportunity to meet and undertake exchanges in a spot market and/or deliver output according to the terms of previously negotiated securities contracts.
3. agents consume the output that they have in their possession.

The timing of events within subperiod 2 depends on what is negotiated contractually.

\(^3\)In Diamond (1984) and Andolfatto and Nosal (2001), the act of verification has to be undertaken concurrently with production. With this latter structure, an equilibrium has to feature monitoring effort exerted on all projects. In the current set-up, it is possible to have monitoring occur for only a subset of projects.

\(^4\)This information can, however, be communicated to the contract enforcement mechanism.
In a world with perfect information, a planner who weights all individuals equally can easily deliver an expected utility of \((1 - \lambda)\text{y}\) to each agent; call this the “optimal” or “first-best” allocation. In what follows, we ask whether an optimal allocation can be implemented as the equilibrium of some suitably designed “game.” The analysis considers two general types of economic mechanisms, which we refer to as “centralized” and “decentralized” trading arrangements. In a centralized trading arrangement, there is a date 0 market for competing contracts with the winner of this competition responsible for intermediating all subsequent trade in the economy. In a decentralized trading arrangement, trade takes place in a system of competitive markets (either securities markets or spot markets), in which case many different contracts are used to govern the pattern of trade.

### 3 Centralized Trading Arrangements

In a centralized trading arrangement, all trade is intermediated by a single agent (belonging to the set of \(A, B, C\) agents and not some fictitious planner); call this agent the “Contractor.” The Contractor draws up a set of contracts (or perhaps one “big” contract) that specifies an allocation that is a function of whatever is verifiable. In a world with symmetric information, one can implement the first-best allocation through centralized trade. In the context of our environment, the Contractor can offer three different kinds of contracts, one for each type of agent in the economy, each of which takes the following general form:

\[
\text{At date } d_1, \text{(name of agent) is entitled to } (1 - \lambda)y \text{ units of output, which is to be made available by (name of Contractor). The (name of Contractor) is entitled to the output produced by (name of agent) at date } d_2. \]

The contract for an \(A\) type agent has \(d_1 = 1\) and \(d_2 = 3\); for a \(B\) type agent has \(d_1 = 2\) and \(d_2 = 1\); and for a \(C\) type agent has \(d_1 = 3\) and \(d_2 = 2\). Since contracts are enforceable and output is observable and verifiable, the Contractor will receive \((1 - \lambda)y\) units of output in each period. Hence, the Contractor will have sufficient resources to meet all of his contractual obligations. Note that because of free-entry in the date 0 market for contracts (and because contracts can be drawn up and enforced at zero resource cost), the winning contract ensures that the Contractor also receives (expected) utility equal to \((1 - \lambda)y\).

In a world with asymmetric information, a successful producing agent may have an incentive to claim that he is unsuccessful in order to (illegitimately) consume the output of his project. It may appear that some form of auditing activity will be required in order to ensure that agents live up to their promises. It turns out that the first-best allocation can still be implemented with a suitably modified contract that takes the following general form:
At date \( d_1 \), (name of agent) is entitled to receive \((1 - \lambda)y\) units of output, which is to be made available by (name of Contractor). The (name of Contractor) is entitled to the output produced by (name of agent) at date \( d_2 \). If the output at date \( d_2 \) is not made available, then (name of Contractor) has the option to perform an audit on (name the agent). Any output that is discovered during the audit is to be seized and transferred to (name of Contractor).

Just as in the models of Diamond (1984) and Williamson (1986), where a financial intermediary is able to guarantee a certain payoff to its investors because it has claims to a certain aggregate payoff, a Contractor can guarantee a certain payoff to agents because the output that the Contractor has a claim on is certain. As well, by way of an audit the Contractor has to ability to obtain (at some personal cost) the required output in order to honor any outstanding liabilities.

If the Contractor offers a set of the above contracts to all agents in the economy and all the agents accept these contracts, then it is possible to implement the first-best allocation as a subgame perfect equilibrium. To see this, suppose that the Contractor offers the appropriate (type contingent) contract to each agent in the economy and that the equilibrium strategy for each agent is to accept the contract and to release \( y \) units of output to the Contractor if his project is successful. Is it profitable for an agent to defect from proposed equilibrium play? In particular, does the agent have any incentive to not release \( y \) units of production when his project is successful? If an agent defects from proposed play in this manner, then the Contractor will have strictly less than \((1 - \lambda)Ny\) units of output to distribute in that period. Since the Contractor is legally obliged to distribute \((1 - \lambda)Ny\) units of output in each period, the Contractor must audit those agents who did not submit any output in order to make up the shortfall in output. Through the auditing process the Contractor will recover the "missing" output. Since the defecting agent will be unable to consume any of the output that he does not release to the Contractor, he will have no incentive to defect from his equilibrium strategy.

We have shown that conditional on the Contractor offering contracts to all agents in the economy, there exists an equilibrium strategy for each agent that specifies: (i) accept the contract; and (ii) release output to the Contractor when the project is successful. Is it an equilibrium strategy for the Contractor to offer these contracts? Or put another way, does any agent in the economy want to be the Contractor? Since the Contractor receives the same level of utility as any other agent in the economy, \((1 - \lambda)y\), any agent is indifferent between being and

---

\(^5\)In fact, the contractor is legally obliged to distribute \((1 - \lambda)Ny\) units of output in only 2 of the three periods. In the period that the contractor consumes, he is legally obliged to distribute \((1 - \lambda)(N - 1)y\) units of output since the contractor himself is "entitled" to \((1 - \lambda)y\) units of output. However, if in this period a successful producer does not release his output, then the contractor would still be required to audit. That is, if the contractor does not perform an audit and forgos his own consumption, he would still be "short" \(\lambda y\) units of output on his contractual obligations.
not being the Contractor. Therefore, the Contractor has an incentive to offer the contracts.

A key insight, we believe, from the centralized environment is that if the economic agent that may end up auditing other agents can base his auditing decision—directly or indirectly—on an aggregate signal, then it may be possible to implement the first-best allocation. And, in order to obtain this aggregate signal, the “auditing agent” will have to be an “active participant” at the time when an initial set of contacts were written. In particular, the auditing agent—who audits in order to protect his asset—must have: (a) an offsetting liability in that period and; (b) negotiated the liability at the same time as the asset. The robustness of this insight should become apparent when we examine trade in a decentralized environment.

4 Decentralized Trade: Date 0 Contracting

4.1 Overview

This and the next two sections can be viewed as extensions of the Second Welfare Theorem. We have shown that it is possible to implement the first-best allocation in a centralized environment. We now examine if it is possible to achieve this allocation through decentralized trade. If the second welfare theorem holds, then we will be particularly interested in the decentralized trading arrangement that implements the optimal outcome.

The obvious place to start would be to ask if Arrow-Debreu style securities markets and prices can implement the first-best allocation. This is the question that we address in this section. It turns out that the answer is this question is “no.” We then ask if a sequence of “spot” markets can implement the first-best allocation. The answer here is “yes,” but depends critically on a number of factors. First, spot market transactions must involve a swap of goods for “money,” where money is a liability of some agent in the economy that circulates. Second, there has to be a relationship between the agent who issues to security that circulates and the agent(s) who might end up auditing other agents. In particular the agent who issues the security that circulates must also be the agent who might end up auditing other agents. The details that underlie these results are worked out in the subsequent sections.

For example, in the centralized trading environment at date 1, the Contractor’s “asset” is the output that he will receive from B type agents and his “liability” is the output that he promises to give to A type agents. The contracts that represent these assets and liabilities were all negotiated at date 0.
4.2 A Simple Securities Market

For this trading arrangement, a market opens only once, at date 0. The exchange of contracts is imagined to operate in a standard way. For example, an auctioneer calls out a system of exchange rates for a set of commodities, which in this case consist of contracts representing claims to output to be delivered under certain contingencies, with agents then submitting their buy/sell orders. When “supply equals demand” in all markets, contracts are signed and future trades are carried out according to contract specification. The only twist here is that contracts will have to respect the presence of private information and may require the holder of the contract to undertake a costly audit. That is, contracts will have to be incentive compatible. Consequently, a typical contract will have the following form:

This contract represents a claim against the output generated by [the issuer’s name] at date \( t \). When this contract is presented to the issuer at date \( t \), the bearer will be told by the issuer whether or not the project is successful. If successful, the bearer is entitled to redeem this contract for \( x \) units of output at date \( t \). If unsuccessful, the bearer is legally obliged to undertake an audit in order to verify the issuer’s claim that he is unsuccessful. Should the audit reveal that the issuer lied, the bearer may seize (using the force of law if necessary) \( x \) units of the issuer’s date \( t \) output.

Let us refer to such contract as an incentive compatible contract (IC contract). An IC contract is incentive compatible in the sense that the issuer has no incentive to deviate from making truthful statements, regardless of the state of the world.

In the context of the current environment, where preferences are linear and endowments symmetric, there is no need for any one agent to hold contracts of more than one issuing agent for diversification purposes. So, each agent issues an IC contract that sets \( x = y \) and \( t \) equal to the date that the project is expected to produce output. Note that a necessary condition for trade to occur is \( (1 - \lambda)y > \lambda \mu \). That is, \( (1 - \lambda)y - \lambda \mu \) is the expected utility that accrues to any agent from purchasing a security described above; with probability \( (1 - \lambda) \) the security holder receives \( y \) and with probability \( \lambda \) he is obliged to expend the monitoring effort \( \mu \). Given the symmetry that has been built into this environment, securities will exchange at par.

Figure 1 neatly describes the various trades undertaken by the agents in the economy. The trades at date 0 are as follows: \( A \) “sells a claim” against date 3 output, \( y_3 \), and “purchases a claim” on date 1 output, \( y_1 \); \( B \) sells a claim to date 1 output, \( y_1 \), and purchases a claim on date 2 output, \( y_2 \); and \( C \) sells a claim against date 2 output, \( y_2 \), and purchases a claim on date 3 output, \( y_3 \). As time unfolds, issuers learn whether they are successful and make announcements concerning the success or failure of their projects to the holders of their claims. (One can imagine that in each period there are sets
of pairwise meetings between issuers and holders of the contracts.) An agent holding a contract of an issuing agent who reports success, simply exchanges the contract for output and consumes the output. An agent holding a contract of an issuer who reports failure is contractually bound to verify the report at a utility cost of $\mu$. Under this trading arrangement, although agents end up consuming the good that they really desire, the first-best allocation is not implemented because (some) agents must undertake a costly audit. Here, each agent attains an expected (ex ante) utility payoff equal to $(1 - \lambda)y - \lambda\mu$ and the economy-wide audit costs associated with implementing this allocation of resources are $3\lambda N \mu$. Finally, note that as no security circulates, there is no object in this environment that can be readily identified as money.

In the “simple securities market” trading arrangement, producers and consumers are “paired off” and whether or not an audit occurs is determined by an idiosyncratic outcome, i.e., whether the producer in the pairing is successful or not. Note also that the potential agent who might perform an audit does not have an offsetting liability in that period. All this is in contrast to the centralized trading environment, where the decision to undertake an audit is based upon an aggregate signal and the Contractor does have an offsetting liability that was negotiated at the same time as the asset that might require an audit.

5 A Stepping Stone: Monetary Exchange and Delegated Monitoring

In the previous section, contracts are written and traded at date 0, and at dates 1, 2 and 3 agents have pairwise meetings and simply execute the contract as it is written. In the second decentralized trading arrangement that we consider, agents do not write any contracts at date 0. Instead, we now imagine that agents organize their trading activity in a sequence of spot markets that open subsequent to the realization of project risk in each period. We refer to this trading environment as a “stepping stone” because it has elements of both decentralization and centralization. In particular, spot markets mediate trade at dates 1 and 2 and a delegated monitor effectively mediates trade at date 3.

In period 1, type $B$ producers realize the returns on their projects. Type $A$ agents really value this output, but all they have to offer in exchange is their own IC contract, where $x = y$, which represents a stochastic claim against date 3 output. The question here is whether a type $B$ agent will have an incentive to exchange his output (which he values a little bit) for a contract that represent a claim to output he does not value at all. The answer to this question will be “yes” as long as the type $B$ agents anticipate being able to use type $A$’s contract as a payment instrument for output in the subsequent period (which $B$ values a lot). If this is the case, then the successful type $B$ producers have an incentive to sell their output for “money,” i.e., the IC contract issued by a type $A$ agent. We refer to the IC contract issued by a type $A$ agent as money because this
contract will circulate between agents in subsequent periods. The price of a unit of output in terms of an IC contract with \( x = y \) is equal to \( (1 - \lambda)y^{-1} \), (or, in other words, one IC contract trades for \((1 - \lambda)y \) units of date 1 output).

Notice something very important here. The first period spot market entails a *quid pro quo* exchange of money for output between type A and B agents. As such, only the *successful* type B agents will participate in this market since only these agents are in a position to release output for sale. But because these goods are willingly released by the producers, no audit costs need to be incurred in this exchange. This is in stark contrast to the audits that would have had to occur if this output was to be transferred by way of IC contracts issued at date 0. This is the way in which monetary exchange enhances the efficiency of the exchange process.

At date 2, individuals with output to sell are the successful type C producers. These are the people who highly value output at date 3; i.e., the output that is supplied by type A agents. At date 2, however, it is the (previously) successful type B agents who own all the claims to date 3 output. Consequently, it is indeed rational for type B agents to expect that type A contracts (which represent claims to date 3 output) will be valued in date 2. Similar to date 1, successful type C producers will willing release their output to type B agents in exchange for the IC contract issued by a type A agent. Once again, no verification costs are incurred in this period and the price of a unit of output, in terms of an IC contract is equal to \((1 - \lambda)y^{-1}\).

In the final period, the type C agents holding the type A contracts will want to present them for redemption. At this stage, the terms of these contracts are executed in the manner set out by each IC contract; i.e., auditing costs will have to be incurred at this stage if a type A's project is unsuccessful. In general, it will not be efficient for type C agents to *individually* redeem their securities; i.e., they will want to appoint a delegated monitor. Why might we need a delegated monitor? Remember that at date 1, there are \(N\) type A agents issuing securities but only \((1 - \lambda)N\) type B agents in a position to purchase them. Successful type B producers will hold \((1 - \lambda)^{-1} > 1\) securities. Consequently, as long as \((1 - \lambda)^{-1}\) is not equal to an integer, successful type B (and subsequently, type C) producers will be holding at least one fractional claim against some type A agent. In this situation, two potential problems arise if each type C agent intends to undertake the audit activity by himself. If \(\theta\) denotes the fractional value of a claim against a type A's output, then either \(\theta(1 - \lambda)y < \lambda\mu\) or \(\theta(1 - \lambda)y \geq \lambda\mu\) will be true. In the former case, it will make no sense for the contract holder to undertake an audit of the fractional claim, which implies that a type A agent that issued fractional claims will end up consuming some of his date 3 output. This clearly represents a misallocation of resources because the agent that values the good the most does not consume it. In the latter case, each unsuccessful type A producer that issued fractional claims will be audited by more than one type C agent, which again represents a misallocation of resources. There is "too much" auditing going on. Both of these problems
will disappear if one agent emerges as a delegated monitor.

We assume that there is free-entry into the delegated monitoring business. One of the C type agents will become the delegated monitor as it would be too costly to reimburse any other type of agent for his auditing costs. The delegated monitor will purchase all of the type A IC contracts from type C agents at date 3 in exchange for a contract or promissory note that guarantees a fixed payment of date 3 output. Assume, for simplicity, that the delegated monitor monitors and potentially audits all IC contracts simultaneously. This implies that the monitor will have to audit $\lambda N$ agents. Hence, competition among potential monitors imply that the fixed claim that the monitor exchanges for each IC contract is equal to $y - \mu \lambda / (1 - \lambda)$.\(^7,8\) If type C agents are holding fractional claims against type A producers, which in general they will be, then the delegated monitor's promissory note offers a higher expected utility payoff to (successful) type C agents, compared to each type C agent doing his own auditing.

Figure 2 summarizes the transactions undertaken by agents. There is no activity at date 0. At date 1, type B producers learn whether they are successful and each type A individual exchanges a IC contract to these type B agents for $(1 - \lambda)y$ units of date 1 output. At date 2, the successful type B agents can purchase $y$ units of date 2 output with their accumulated money holdings (i.e., IC contracts issued by type A agents) from successful type C producers. At date 3, successful type C agents exchange their money for a promissory note, that pays $y - \lambda/(1 - \lambda)\mu$ units of date 3 output, issued by the delegated monitor. The delegated monitor then collects output from the successful type A producers, in exchange for their IC contracts, and audits the unsuccessful type A producers. The delegated monitor collects $(1 - \lambda)Ny$ units of date 3 output, keeps $\lambda N\mu$ units as compensation for monitoring expenses, and then equally distributes the rest of the output to the type C agents who hold his promissory notes.

Under the present trading arrangement, the expected utility for both type A and B agents is $(1 - \lambda)y$, which is greater than the expected utility they would have attained under the trading arrangement where all agents write and exchange contracts at date 0. As for type C agents (including the delegated monitor), their expected utility is identical to what they would have attained under the date 0 contracting trading arrangement, i.e., each successful type C agent receives an expected payoff of $y - \mu \lambda / (1 - \lambda)$ and the probability of being

---

\(^7\)Note that total output, net of verification costs, in period 3 is $(1 - \lambda)Ny - \lambda N\mu$. Since $(1 - \lambda)N$ type C agents own all of the type A agents securities, each successful type C agent will receive a claim for $y - \mu \lambda / (1 - \lambda)$ units of date 3 output from the delegated monitor, (assuming zero profits). It is important to emphasize that in order to guarantee this fixed payment, the delegated monitor must purchase all of securities issued by type A agents, both fractional and unitary claims.

\(^8\)If the delegated monitor does not monitor simultaneously, but rather sequentially, then the expected number of audits that he will perform will be less than $\lambda N$. In this case, competition among potential delegated monitors will imply that the fixed claim that the delegated monitor issues will be greater than $y - \mu \lambda / (1 - \lambda)$. Qualitatively speaking, our results will be unaffected if we assume that the delegated monitor monitors sequentially.
successful is \((1 - \lambda)\). Consequently, when the trading arrangement that has agents writing and exchanging contracts at date 0 securities markets is replaced with the arrangement that has monetary exchange and a delegated monitor, a strict Pareto improvement is realized.

The introduction of a “monetary instrument” allows agents to economize on the costs associated with redemption activity. In particular, audit costs are completely averted in the first two periods because producers are willing to exchange their output for the monetary instrument. However, audit costs can not be avoided in the final period due to the usual incentive compatibility reason. Although the introduction of a delegated monitor reduces audit costs, compared to individual agents undertaking the audit, the delegated monitor bases his audit decision on idiosyncratic signals, which implies that, in equilibrium, costly audits will be undertaken. Note that the delegated monitor has assets (the IC contracts) and offsetting liabilities. But because these liabilities were not written at the same time as the contracts that represent the assets, the delegated monitor’s audit decision is not based on an aggregate signal. Instead, the audit decision is determined by an idiosyncratic outcome, (as was in the case of simple (date 0) security market trading environment).

6 Decentralized Trade: Money and Banking

In the previous section, the agents that issued monetary instruments (type A agents) operated independently of the agent who supplied the monitoring services (a type C agent). Here we show that this division of tasks is not “natural” in the sense that the efficiency of the exchange process can be further enhanced by having one agent responsible for both money-issuance and monitoring. We will refer to such an agent as the Bank. In this environment, trade is completely decentralized in the sense that spot markets mediate trade at all three dates.

Imagine now that a date 0, one of the type C agents emerges as the Bank. The Bank will issue a security—money—that is to serve as the economy’s monetary instrument. At date 0, each type A agent approaches the Bank and receives a “money loan.” Let the total amount of units of money that the Bank lends out be equal to \(M\). In exchange for the \(M/N\) monetary units borrowed from the Bank, each type A agent hands over a contract that represents a claim against his future output. The contract that each type A agent hands over is the IC contract with \(x = y\) and \(t = 3\), and this contract serves as collateral for the money loan. We now turn to the contractual nature of the monetary instrument created by the Bank and the nature of the money loan contract negotiated with type A agents.

The monetary instrument must be designed with the intent of having it circulate. In the end, it will only circulate if type C agents expect that they can use this money to purchase the goods that they desire; output at date 3. This might be seen as a tricky proposition in that the agents who possess output at
date 3, the type A agents, have no incentive on their own to exchange output—which they value a little bit—for paper—which they do not value at all and did not themselves issue. In order to protect the type C agents from the potential of “bad” behavior on the part of the type A agents, the Bank must embed within its monetary instrument a redemption clause that allows the money holder to convert each monetary unit for \(1/p^*\) units of date 3 output, where \(p^*\) is a price-level satisfying:

\[
p^* = \frac{M}{(1 - \lambda)Ny}.
\]

In order to make this convertibility option credible, the Bank must have the capacity to obtain up to \((1 - \lambda)Ny\) units of output at date 3 if the need so arises. But the Bank does have the wherewithal to honor the redemption clause since it holds as collateral IC contracts issued by type A agents.

Now let’s consider the nature of the money loan contract. Each type A agent borrows \(M/N\) dollars from the Bank and promises to repay \(M(1 - \lambda)^{-1}\) if his project is successful at date 3 and nothing if his project is unsuccessful. Upon payment of \(M(1 - \lambda)^{-1}\), the Bank extinguishes A’s debt obligation and returns the collateral (IC contract) back to A. Notice that if all successful type A agents behave “honestly,” then the Bank can expect an aggregate repayment of \(M\) dollars. Hence, if the aggregate repayment by type A agents less than \(M\), then at least some type A agents are “cheating” in the sense that they have output to sell but choose not to sell it.

If the Bank receives less than \(M\) dollars in repayment, then it does not have an incentive on its own to determine which of the successful type A producers did not sell their output. The Bank then might not insist that the money loan be repaid. After all, the Bank will receive only “pieces of paper” when the loan is paid off, which do not in themselves provide the Bank with any utility. But in order to determine which successful type A agents did not sell their output, the Bank will incur some monitoring costs. Given that the Bank has does not appear to have an incentive to collect on its money loan and that successful type A agents have an incentive to “cheat” by not selling their output at date 3, how is it possible for the Bank’s money to circulate? The key here is the redemption clause imbedded in Bank’s money. If a type C agent is unable to purchase \(1/p^*\) units of date 3 output per unit of Bank currency in the market, then he will forego market exchange and will demand redemption at the Bank.

At this point, the Bank is legally obliged to exchange the Bank’s notes for date

---

9 At this point it may not be obvious why the price level takes on this particular value. We will show, however, that if the money and banking setup is a feasible (equilibrium) outcome, then the price level described below will be the equilibrium price level.

10 Here, we anticipate that the equilibrium (gross) interest rate charged on each money loan is equal to \((1 - \lambda)^{-1} > 1\) so that the net interest earned on the bank’s loan portfolio is equal to zero. This zero interest rate reflects both the fact that zero banking costs will be incurred in equilibrium together with the assumption of free-entry into the banking business.

11 By “behaving honestly” we mean that at date 3 we mean that all successful type A producers exchange their output for bank money in the spot market and then repay their money loan debt.
3 output. The Bank will, therefore, audit those type A agents who did not pay off their money loan. Recall that the Bank holds as collateral an IC contract issued by type A agents and, therefore, has the means by which to conduct an audit and obtain date 3 output. Type A agents, understanding this, realize that it is fruitless to cheat on their money loan. Hence, successful type A agents will sell their date 3 output for Bank money in the date 3 spot market and will repay their money loan.

The equilibrium path for the money and banking trading environment is the following. At date 0, one of the type C agents becomes the Bank; each type A agent obtains a money loan of \( M/N \) (described above) from the Bank and hands over to the Bank an IC contract (described above) as collateral. At date 1, the spot market price for output is \( p^* \); successful type B producers will exchange \( y \) units of output for \( (1 - \lambda)N^{-1} M \) units of Bank money and all type A agents purchase and consume \( (1 - \lambda)y \) units of output. At date 2, the spot market price for output is \( p^* \); successful type C producers exchange \( y \) units of output for \( (1 - \lambda)N^{-1} M \) units of Bank money and type B agents that were successful purchase and consume \( y \) units of output. At date 3, the spot market price for output is \( p^* \); successful type A producers sell their output in exchange for \( (1 - \lambda)N^{-1} M \) units of Bank money and previously successful type C exchange their money holding for \( y \) units of output. Successful type A producers payoff their money loan, paying \( (1 - \lambda)N^{-1} M \) units of money. In exchange, type A agents receive their IC contract. If a successful type A producer attempts to keep some of all of his output for his own consumption, then some successful type C agent(s) will be unable to obtain \( y \) units of date 3 output at unit price \( p^* \). These agents will then demand redemption from the Bank, which is legally obliged to honor any such request. The Bank will then audit those type A agents that did not pay back their money loans. The type A agents who (implicitly) claimed that they were unsuccessful will be discovered; their output will be confiscated and transferred to the type C agents. Anticipating these events, no successful type A agent will have an incentive to hide his output for his own consumption.

In equilibrium, the money and banking trading environment entails zero monitoring costs. As a consequence, the expected utility of each agent in this economy (including the Bank) is equal to \((1 - \lambda)y\); i.e., the first-best allocation is implemented. The efficiency gain—compared to the monetary exchange and delegated monitor trading environment—arises because the delegated monitor issues money, redeemable in date 3 output, and extends a cleverly designed money loan that compels those debtors who have the capacity to seek out (thereby costlessly revealing their success) the dollars necessary to pay off their money loan. The subsequent money for goods transactions avoid the unnecessary costs associated with redemption activities.

Finally, the Bank, like the Contractor in the centralized trading environment, bases his audit decision on an aggregate signal. The signal here is a request from a type C agent to redeem bank notes for output at date 3. This is an aggregate
signal because a type $C$ agent will demand that the Bank redeem its note only if he (agent $C$) is unable to purchase $(1 - \lambda)y$ units of output in the market at price $p^*$ per unit, i.e., there must be less than $(1 - \lambda)Ny$ units of output for sale in the market.

7 Discussion

7.1 The Essentiality of Money and Banking

In our model, money and banking are not essential in the sense that the first-best allocation can be implement without use of either money or banking. For example, a centralized trading scheme is capable of implementing the first-best allocation. It is even possible to implement the first-best allocation using market trading in some periods and centralized trading in other periods. If, however, one is interested in decentralized economies, i.e., where all trade is mediated by markets, then money and banking are essential. That is, it is not possible to implement the first-best allocation when trade is mediated by markets if the economy does not have institutions that resemble what we have called “money and banking.” To see this, note that a claim on date 3 output that can be exchanged for goods in spot markets will provide an incentive for successful type $B$ and type $C$ agents to reveal and part with their output. Hence, for the first two dates money alone is essential in order to avoid costs associated with auditing when trade is mediated by markets. But if the claim that serves as money is issued by the type $A$ agent, then audit costs will necessarily be incurred at date 3. Therefore, the claim that circulates as money must at the same time represent a credible claim to date 3 output and can not be issued by agent $A$. “Another agent” can issue such a claim that circulates only if he, in turn, possesses claims to date 3 output that are issued by agent $A$, i.e., this claim that is issued by the “another agent” will not be accepted by agent $C$ if it does not, directly or indirectly, represent a real claim to date 3 output. Hence, a banking-type institution is necessary to implement efficient allocations costlessly when trade is mediated by markets. What we show that a banking type institution is also sufficient to implement the first-best allocation.

7.2 Risk Aversion

We have assumed throughout that all agents are risk-neutral. Suppose instead that all agents are risk-averse, where their preferences over the consumption good that they really desired, $c$, and the good that they produce, $y$, are given by $u(c) + \epsilon u(y)$ if the agent does not perform an audit and $u(c) + \epsilon u(y) - \mu$ if he does perform an audit. For these preferences, the unique first-best allocation (where a social planner equally weights all agents in the economy) is now characterized
by each agent consuming \((1 - \lambda)y\) units of the good that he really desires.\(^{12}\) This is precisely the allocation that the centralized trading environment implements.

In the money and banking trading environment studied above, each type \(A\) agent consumes \((1 - \lambda)y\) units of (date 1) output. But, type \(B\) and \(C\) agents consumption levels are stochastic: By itself, the money and banking trading environment can not implement the first-best allocation when agents are risk averse. If, however, we allow an audit to reveal not only the amount of goods that an agent produced but also the amount of money that an agent is holding, then it will be possible to implement the first-best in a money and banking trading environment when all agent are risk-averse.

To see this, imagine that at date 0, one of the agents in the economy emerges as an “Insurer.” An Insurer collects premiums from agents and pays out indemnities. Competition to be the Insurer will imply zero profit for this activity. In equilibrium, the Insurer will offer the following policy,

\[
\text{The premium for this insurance policy is } \lambda M/[(1 - \lambda)Np^*] \text{ units of money. If at date } t, t = 1, 2, \text{ (name of policy holder) are not a successful producer, then at date } t + 1, \text{ you will receive } M/[(1 - \lambda)Np^*] \text{ units of money from (name of Insurer). The premium payment is due at the beginning of date } t + 1 \text{ and the (net) indemnity can be collected in date } t + 1, \text{ after premiums are paid. (Name of Insurer) always has the option to audit any policy holder. If (name of policy holder) is audited and is found to have more than } M/(Np^*) \text{ units of money, then (name of Insurer) will confiscate all of it and will equally distribute this amount among the date } t + 1 \text{ policy holders.}
\]

Note that the premium payment and the indemnity payoffs are chosen so that, independent of whether or not the policy holder was a successful producer, he will have exactly the right number of units of money to purchase \((1 - \lambda)y\) units of the output that he really desires.

Consider now the money and banking trading environment with an Insurer. At date 0, an insurer offers the above contract. All type \(B\) agents will purchase insurance contracts that pay off at date \(t = 2\) and all type \(C\) agents will purchase insurance contracts that pay off at date \(t = 3\). Type \(A\) agents do not purchase insurance contracts. In addition, at date 0, one of the type \(C\) agents emerges as a Bank. All type \(A\) agents approach the Bank and receive a money loan. The nature of the various (contractual) exchanges between the type \(A\) agents and the Bank are identical to that described in the previous section (where there was no Insurer). As well, the date 1 exchanges are identical to the exchanges in the previous section. Date 2 starts off with the each successful type \(B\) agent paying his premium, equal to \(\lambda M/[(1 - \lambda)Np^*]\), to the Insurer. If all successful

\(^{12}\)When agents are assumed to be risk-neutral, the first-best allocation of goods is indeterminate. Any allocation that provides each agent with an expected utility of \((1 - \lambda)y\) is a first-best allocation of goods.
type $B$ agents pay this premium, then all of the unsuccessful type $B$ agents can receive their net indemnity payment of $\frac{M}{(1 - \lambda)Np^*}$, without having the Insurer audit any agent. Note that successful type $B$ agents will always pay their premiums. If they did not, the Insurer would have to undertake an audit. The audit would reveal the successful agent and, given the nature of the insurance contract, the Insurer would seize all of the successful agent’s money. Hence, it is a dominant strategy for the successful agent to pay his premium. Now at date 2, all type $B$ agents have the same amount of money and each will purchase $(1 - \lambda)y$ units of output with their money holdings. Successful type $C$ agents sell all of their output in the market in exchange for money. In terms of insurance payments, date 3 is identical to date 2 except that it is the type $C$ agents who pay premiums and receive indemnities. After insurance contracts are settled, all type $C$ agents have the same money holdings and, assuming that all successful type $A$ producers sell their output in the market, all type $C$ agents will purchase $(1 - \lambda)y$ units of output with their money holds. Successful type $A$ producers will have an incentive to sell their output in the market for exactly the reasons given in the previous section.

If agents are risk averse, then a money and banking trading arrangement can implement the first-best. However, a new institution—an insurance company—must arise if this allocation is to be implemented. A profit motive to become an insurance company will ensure that it does, in fact, arise.
8 References


5. D. Gale and M. Hellwig, 1984,

