Why Pay?
An introduction to payments economics*

September, 2006

Charles M. Kahn
Department of Finance
University of Illinois
c-kahn@uiuc.edu

William Roberds
Research Department
Federal Reserve Bank of Atlanta
william.roberds@atl.frb.org

Abstract:

JEL Fields:
Keywords: payments, money

* The views expressed in this paper are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. An earlier version of this essay was presented by the first author at a conference, "The Economics of Payments II" held at the Federal Reserve Bank of New York, March 29-30, 2006.

1. Introduction

Payment systems are the plumbing of the economy—a collection of conduits that is essential, pervasive, and boring until there's a malfunction. Economists studying payment systems have long labored under the shadow of this unsavory metaphor. Recently, however, payments economics has begun to achieve some respect, because of the significant changes in payments technology and infrastructure, because of the important policy concerns associated with the industry, and possibly because of the sheer magnitude of payment activity.

This essay is an (admittedly idiosyncratic) introduction to some of the issues of payments economics and the empirical work on payment systems, but above all to the models used to understand the role of payments in an economy and the forms which payments take. By the end we hope the reader will conclude that plumbing can be an interesting object of study.

A payment occurs when one economic agent transfers an asset to another agent for the purpose of discharging a debt. In developed economies, the action of payment is so mundane and apparently simple that this definition hardly seems necessary. But economists have struggled to understand the nature of this everyday activity: constructing a convincing model of payments is a challenging task. Payments and payment systems, so ubiquitous and obviously essential to real-world economies, are conspicuously absent from the world of Arrow-Debreu.

But what precisely are the frictions that give rise to payment arrangements? Do these frictions, and the payment systems designed to overcome them, matter for economic allocations more generally? How do payment systems interact with the machinery of financial intermediation? And how will these systems evolve with the ongoing improvements in information technology?

1.1. Payments—big business and getting bigger

Payments are big business, even by the standards of macroeconomists. For example, 85 billion payments of $883 trillion were recorded in the United States in 2004, not counting payments made in currency.[1] This figure corresponds to about $75 in payments for each dollar of GDP, a ratio that has been steadily rising over time. Payment statistics in all developed economies display similar levels and trends (summary statistics can be found in Committee on Payment and Settlement Systems 2006).

Much of the recent pickup in payments activity comes from the *large-value* or *wholesale* payment systems used to settle obligations between banks. Since the "funds side" of financial market trades[2] must ultimately settle through such systems, volume over these systems has mushroomed with the upsurge in financial market trading.[3] But volume is also expanding quite rapidly over certain types of *small-value* or *retail* payment systems, those used by households and non-bank firms. In the U.S., for example, usage of debit cards and direct transfers[4] is expanding at double-digit rates (measured either in terms of the volume or value of transactions), even as the usage of checks and cash is declining (Humphrey 2004; Gerdes et al. 2005; Garcia-Swartz et al. 2006a; Klee 2006a). Transactions that were once only conducted solely with cash are increasingly made using cards that electronically link buyers to their payment histories.[5]

The past decade has also witnessed tremendous innovation in payments. Besides the spectacular rise of card payments, innovative web-based payment systems such as PayPal have enabled individuals to "wire" funds across great distance, instantaneously and at low cost (Kuttner and McAndrews 2001). At the wholesale level, the multicurrency CLS system (beginning operations in 2002) has allowed banks to coordinate the settlement of foreign-exchange transactions across national borders, binding together all major national payment systems in a way that was unthinkable a decade ago. And there is no reason to believe that the rate of innovation will slow down anytime soon. The "pay-

[1] This figure also does not include the U.S. dollar payments made through the international, multicurrency CLS system, for which no precise figures are available. Adding these in would drive the total close to $1 quadrillion.

[2] Residual amounts left after any netting of securities trades.

[3] In fact there is a close connection between the study of payments system and the study of clearing and settlement systems for financial markets. Many of the policy issues are similar and many of the techniques used to design them are parallel. In Europe as a result of monetary unification, there has been concern with the competitiveness and industrial organization of financial settlement systems. For an introduction to these issues, see European Central Bank (2006) and Milne, (forthcoming).

[4] I.e., those made through the Automated Clearinghouse (ACH), such as direct deposit of payrolls.

[5] For a valuable examination of the history and growth of the use of credit cards see Evans and Schmalensee (1999); an informal theory of the process is found in Kahn (2006).

ment card" of the future will likely be just a "smart card," cell phone, or other portable computing device, able to instantaneously transfer value to anyone so equipped.[6]

1.2. Policy issues

While the technology-fueled expansion of payments activity has clearly generated tremendous economic benefits, it has also generated noteworthy and in many cases, unresolved questions for policymakers.

Important policy issues for wholesale payment systems include *access*, *liquidity*, and *systemic risk.* Important policy issues for retail payment systems include *competition* and *fraud*. *Cross-border coordination* is an important problem at both levels.

1.2.1 Wholesale systems

Economic models of banks have usually focused on the roles of banks as financial intermediaries, parties able to exploit the synergies between the provision of extremely liquid deposits and equally illiquid loans (e.g., Diamond and Dybvig 1983; Kashyap, Rajan and Stein 2002). But banks have always served another, equally essential role, as providers of payment services—as parties able to transfer liquid claims quickly and cheaply and with a minimum of legal uncertainty.[7]

Financial innovations such as securitization, syndication, and credit insurance have eroded banks' advantages as financial intermediaries, subjecting them to competition from both financial markets and non-bank financial institutions. But, for the moment at least, banks continue to enjoy significant advantages over non-banks as payment service providers. These advantages include, most critically, full access to wholesale payment systems. Such access is usually restricted to (regulated) banks, and sometimes only

---

[6] Although in the case of smart cards, it probably makes sense to be a little cautious. Observers have for many years been predicting their arrival as a major payment method. While they have begun to have an impact in other countries, they still are relatively unused in the U.S.

[7] The history of banking (see Kohn 1999 for a survey) indicates that banks arose first as providers of payment services—moneychanging and book-entry payments—and then diversified into the business of financial intermediation. Central bank participation in payment arrangements dates to at least the 15th century *Banco di San Giorgio* in Genoa (Fratianni and Spinelli 2006). By the 18th century, book-entry payments at the Bank of Amsterdam—the leading central bank of its day— were the dominant wholesale payment system in Europe (Quinn and Roberds 2006). For a reexamination of banking history from the point of view of payment systems, see Speight et al. (2006).

to banks whose home charter is in the same jurisdiction as the system itself.[8] Indeed it has been argued that such access has become *the* definitive feature of a bank (Lacker 2006). Recent policy discussions in the U.S. regarding the acquisition of banks by non-financial firms have focused on this issue. Wal-Mart and other would-be acquirers of banks have stated that their principal motivation is not to compete with banks as financial intermediaries, but instead to gain unfettered access to interbank payment arrangements.[9]

Liquidity is a concern because of the very high value of payments routed through wholesale systems, due to settlement of financial market trades. No intraday markets exist to allocate this liquidity. Instead, such liquidity has traditionally been allocated implicitly through agreements between banks to settle payments on a net basis. During the 1990's, however, net settlement of large-value interbank payments came under criticism as being too subject to *systemic risk*, roughly defined as the risk that a failure by one bank to settle might result in multiple settlement failures.[10] Chains of failures have rarely been observed in practice, but there have been enough "near misses" to produce considerable regulatory angst.[11] Consequently net settlement has in many cases been abandoned in favor of *real-time gross settlement* (RTGS).

Under RTGS, each payment by a bank to another bank consists of an irrevocable transfer of central bank funds. Until the payor has access to such funds, no payment can be made. RTGS systems[12] thus insulate an individual payee from systemic risk, but their operation requires many times more central bank liquidity than net settlement systems. How such liquidity should best be provided remains an open question. The Federal Reserve System, for example, routinely allows banks using Fedwire to incur uncollateralized intraday overdrafts, but other central banks (Euro area, Japan, U.K.) only grant such

---

[8] A condition for full participation in such systems is an account with the relevant central bank, even in cases where the system is privately operated. Central banks have traditionally limited access to such accounts, justifying this limitation both on moral hazard grounds (Bank of England 2002), and on a reluctance to compete with commercial banks (Green and Todd 2001).

[9] See for example the written testimony of Jane J. Thompson, Wal-Mart Financial Services, at the Wal-Mart Bank Federal Deposit Insurance Application Public Hearing, April 10th, 2006, Washington, DC. Available online at http://www.fdic.gov/regulations/laws/walmart/publichearings/walmartjane.html.

[10] See the "Lamfalussy report" (Committee on Payment and Settlement Systems 1990).

[11] Among the most famous occurrences resulted from the 1974 failure of Bankhaus Herstatt and the 1991 failure of BCCI; see the "Allsopp report" (Committee on Payment and Settlement Systems 1996) for the details of these and other incidents.

[12] Including Fedwire in the U.S., TARGET in the Euro area, BOJ-NET in Japan, CHAPS in the U.K., and many others.

credit against collateral. The Fed charges fees against overdrafts (over a ceiling), but many other central banks do not. Other large-value payment systems—CHIPS in the U.S., LTVS in Canada, and the multicurrency CLS—have opted for arrangements that employ modified versions of net settlement, often in conjunction with queuing arrangements that attempt to optimize the extent to which queued payments may be netted (Intraday Liquidity Management Task Force 2000; McAndrews and Trundle 2001, Willison 2005).

1.2.2 Retail systems

For retail systems, the principal policy controversies have centered on issues of competitive efficiency. Payment systems are a relatively expensive component of financial infrastructure, as much as 3 percent of GDP in the U.S. (Humphrey, Pulley, and Vesala 2000). In the U.S. case, one reason this figure is so high is the continued widespread use of checks, but the cost of electronic alternatives can also be substantial, particularly for card-based payments. Merchant dissatisfaction with the fees charged for card payments has resulted in some spectacular antitrust litigation in the U.S. (the "Wal-Mart case" against Visa and MasterCard, settled in 2003 for $3 billion; additional lawsuits are pending) and outright regulation of card fees in countries such as Australia (Lowe 2005). As is the case with some other "network" industries, there is little consensus on what constitutes an efficient fee structure for card-based payments (Rochet and Tirole 2006a).

Another policy controversy associated with retail payments has been the emergence of new forms of payments fraud such as "identity theft." Electronic payment systems can offer tremendous efficiency gains, by allowing for rapid and easy transmission of information across system participants. But the flipside of this efficiency is that these same systems can allow for rapid propagation of fraud. A single loss of confidential data, such as the compromise of 40 million credit card accounts at CardSystem Solutions Inc. in June 2005, can lead to significant fraud losses.[13] Payment service providers and policymakers have struggled to find a balance between the inherent advantages of information-sharing, and the potential costs that can arise due to fraud and loss of privacy.

[13] See Isabelle Lindenmayer, "Security Watch," American Banker, Friday, September 2, 2005.

Payment systems have traditionally observed national boundaries, but over the past decade cross-border arrangements have become increasingly common. Cross-border credit card and ATM card transactions are now routine, and cross-border direct (ACH) transfers are starting to see increasing usage. At the wholesale level, the Euro-area-wide TARGET system (beginning operations in 1999) and the multicurrency CLS system are now two of the world's largest payment systems, measured by the value of funds transferred. The advent of these cross-border arrangements has no doubt facilitated international commerce, but at the same time their operation has required a much greater degree of coordination between central banks and regulatory authorities than was previously the case.

2. Theory of payments

Payments arise in environments characterized by two potential impediments to exchange: a *time mismatch* of trading demands, and *limited enforcement* of pledges about future behavior.

The notion of a time mismatch is less stringent that Menger's celebrated "lack of double coincidence of wants." What is key is not an absence of counterparty-by-counterparty matchups in consumption good demands, but instead an inadequate supply of liquid (desirable) assets to allow for exchange to proceed as a sequence of spot trades.

Limited enforcement can arise in many ways. Enforcement of obligations might be hindered by geographical displacement of potential transactors, or by an inadequate legal system. Informational frictions can also play a role. For example, account-based payment systems must incorporate at least two critical information technologies, one for keeping track of an individual's actions over time, and another to verify identities. The scope and application of such systems is clearly limited by the efficiency of these underlying technologies.

Broadly defined, a *payment system* is any arrangement that enables exchange by overcoming the paired frictions of time mismatch and limited enforcement; payments economics is the study of such arrangements. As such, this field has drawn on many fields of economics, but as noted by Green (2004), its most important antecedent is monetary theory. The study of payments is in part a study of the techniques institutions use to increase transactions velocity, i.e., to make payments more efficiently with the same stock of money (see, for example, Temzelides and Williamson 2001).

Despite its overlap with monetary economics, the study of payments is focused on very different issues from the study of money. Monetary theory focuses on questions about how money obtains its value, the relationship between inside and outside money, and how the supply of money should be optimally regulated. Payment economics is in a sense more fundamental, relying on information economics and mechanism design to characterize alternatives to money, and to explain how certain assets may take on money-like traits. Over the long run, changes in underlying features of economic environments and consequent structure of payments will continue to redefine what may be considered "money."

2.1. Two types of payment systems

A fundamental distinction should be made between payment systems that are *store-of-value systems* and those that are *account-based systems.* Store-of-value systems, such as commodity money, fiat money, and stored value cards, are founded on the transfer of some payments object (be it coins, notes, or electronic stored value) between payor and payee, and they depend critically on a payee's ability to verify the payments object. Account-based systems, such as charge accounts, checks, and credit cards, require the keeping of accounts in the name of the payor and payee. The success of account-based system hinges, most fundamentally, on the ability of its participants to verify the identities of account holders, to ascertain the link between transactors and histories. This dichotomy necessarily represents something of an idealization, and there are some types of systems (e.g., debit cards) that arguably do not fall so neatly into either of these camps. Nonetheless the overwhelming proportion of payments arrangements can be classified into one of these two forms. Why this should be so remains an interesting challenge (first posed by Green 2004) for the theory of payments.

For each of these types of system we first consider a "Platonic ideal"—a baseline model of how such a system might function in its most abstract form. For account-based systems this ideal will be pure, costless credit. For store-of-value systems it is spot trades using perfectly liquid assets. Whenever considering more complex systems, we will treat these as attempts to approximate one these two ideals (for a similar approach, see Nosal and Rocheteau 2006).

2.2. Baseline 1: pure credit

Theoretical models of exchange there often examine "gift-giving" equilibria, but of course economic agents never give gifts selflessly. Instead successful gift-giving arrangements depend on agents giving gifts in anticipation of possible future rewards, possibly combined with fear of punishment if gifts are not handed over. Gift-giving arrangements are thus more likely to be successful given a sufficiently low discount rate, high probabilities of and stringent penalties for defection, and high anticipated future benefits. A classic example of gift giving arises is in Diamond's (1990) model of pairwise meetings with credit, where autarky is the punishment. Jin and Temzelides (2004) extend this idea in their study of the interactions of money and credit. In their model, the frequency of meetings between agents determines the possibility of credit relationships. Agents simultaneously maintain credit relationships with individuals with whom they meet frequently, and cash relationships with individuals whom they rarely meet.

In pure credit models, retaliation against defectors can take a variety of forms. One obvious possibility is a fine or nonpecuniary punishment ("jail"). Others include an individual refusal to trade, expulsion from a group, or even "the nuclear option"—economy-wide reversion to autarky.

In these highly stylized models of exchange, gift giving can be surprisingly robust. Araujo (2004) shows that gift-giving equilibria can be supported even if the identity of a defector is *not* known and acts of defection are *not* publicly revealed, as long as agents are sufficiently patient. Through a "contagion equilibrium," even a single refusal to trade can eventually lead to an economy-wide breakdown in reciprocity; thus agents are induced to adhere to the "social norm" of trade. Although Araujo's result provides a useful theoretical starting point, it is clear that much less patience will be required when defectors can be identified and punishments more sharply focused.

2.3. Credit with "hostages"

When credit is inadequate by itself, other institutions can arise to help support promises. These institutions allow, in effect, for the taking of hostages: agents agree on a (possibly costly) action making retaliation for defection easier or more potent.

Corbae and Ritter (2004) provide one example of such an arrangement where, in addition to the possibility of interacting in random meetings, agents can form partnerships that extend voluntarily over time. The message in the paper is that even if an individual is currently matched with a less-than-ideal partner, the continuation of the partnership can enforce promises by enhancing the cost of defection.

In the context of payments, a more typical form of hostage is collateral. There are numerous examples of collateral serving this role. In Shi (1996) and Li (2001) collateral is modeled as the taking of an agent's "tools" until a debt is repaid. The absence of the necessary implements makes consumption impossible. Koeppl and Monnet (2005) construct a model of central counterparties and margin posting, whereby payments are facilitated by collateral held in escrow by a third party. Kahn and Roberds (2001a) place a similar interpretation on the CLS system for international large value payments. They argue that CLS acts as an "escrow agent" able to impose greater penalties for nonperformance than could be imposed through purely bilateral arrangements.

Such arrangements depend on the ready availability of a collateral good for which there is a low cost of transfer into the creditor's control (low transportation costs, low likelihood of damage while hostage, no legal uncertainty about ownership). The cost of a loss of the collateral to the debtor should be high relative to its value to the creditor (Williamson 1985). Collateral also tends to reduce costs of keeping track of a debtor, since the debtor has an incentive to return to retake possession of the collateral.

2.4. Baseline 2: pure spot exchange

From hostage taking it is a small jump to spot trades in assets. Hostage taking occurs when a creditor does not value an asset posted as collateral, but anticipates that he original owner will pay to get it back. Trading in assets occurs when the "hostage" collateral good can be readily sold to third parties, so that accepting such a good in payment makes sense, even if the payee does not value the asset itself. In this sense, commodity money is simply an asset with a negligible cost of transfer (Ostroy and Starr 1974, Kiyotaki and Wright 1989). At some point there is no longer the need to maintain accounts and the transferred asset becomes a “store of value.”

Models with store-of-value payments often place agents into “wilderness” settings. An agent wanders about (“searches”) through an uncharted rain forest, occasionally encountering a lone stranger. These encounters are never repeated, and agents are beyond the reach of any court. The only feasible interactions between agents are spot trades: each agent simultaneously holds out his own good while grabbing for the other. Absent a double coincidence of wants, goods can only trade for a store of value.

Working through the models of these environments is valuable for emphasizing the parsimony of information needed to effect payment under a store-of-value system. A trader needs to verify the genuineness of the proffered store of value, but need know nothing about his counterparty.

The canonical model of this type is the well-known search model of Kiyotaki and Wright (1989). But the essential idea of nonrecurring encounters can be found in earlier models such as Samuelson’s (1958) overlapping-generations model and Townsend’s (1980) turnpike model. “Island” models such as Freeman (1996) motivate the use of money through both overlapping generations and geographic dispersion. “Day/night” models such as Lagos and Wright (2005) have sought to increase the tractability of search by allowing for alternating periods of anonymous trade and trade in Walrasian markets.

The ingenious details of these models can sometimes obscure their key point, which is that trade in assets is useful if malefactors cannot be identified, caught, and punished. In other words, a market could be completely centralized, but if it were also completely anonymous, trade in stores of value will still be useful.

2.5. Fiat money as a store of value

Any good with the necessary physical characteristics can serve as a store of value. But if such assets are in short supply, one possibility is to create them in the form of fiat money. The are numerous theories about how intrinsically worthless fiat money may come to represent a store of value, but these can generally be split into two groups. The first group—the “soufflé theories”—postulates that money has value today because it is believed to have value tomorrow (see references in the previous section). The other group—the "Mahagonny theories"[14]—asserts that money has value because it is needed to pay taxes (as in e.g., Starr 1974; Aiyagari and Wallace 1997). From the point of view of payments foundations, it does not really which group of theories is the more accurate. Once the money has future value, for whatever reason, it can function as a store of value.

When trade based on credit fails or is too costly due to limited enforcement, then trade in assets, including fiat money, may be an attractive alternative, imposing fewer informational and enforcement costs. But payment in money, like all store-of-value arrangements, is inherently a little less flexible than recordkeeping, because it is subject to a physical restriction in various stochastic realizations (as anyone who has ever searched for an ATM can attest). There is a strand of monetary theory which emphasizes this distinction (Kocherlakota and Wallace 1998).

Despite these apparently obvious and practical distinctions, at the theoretical level it can be difficult to distinguish between the functioning of money and credit. Numerous papers (e.g., Townsend 1987; Taub 1994; Aiyagari and Williamson 2000; Corbae, Temzelides, and Wright 2003; Berentsen et al. forthcoming) have shown, in some circumstances at least, that trade with money can deliver identical outcomes to trade with credit. Kocherlakota (1998) argues that this is because money functions as a cheap, easily transportable proxy for expensive public record keeping ("memory"), i.e., an agent's money holdings can effectively serve as a sort of portable account balance. Because this proxy is sometimes imperfect, however, Kocherlakota's argument implies that as the costs of recordkeeping fall, the use of payment systems based on store of value should likewise fall. Temzelides and Yu (2004) develop this idea somewhat further in a model in which money is used when it is not worthwhile to track credits and debts in more complex ways: small transactions use money. Over time, the margin where account-based payment systems become effective has grown ever smaller.[15]

---

[14] After the1930 Brecht musical *Rise and Fall of the City of Mahagonny*, which takes place in a city where the only crime is to be caught without money.

[15] The cost to a merchant of accepting card payments remains stubbornly high, in the U.S. about $.60 for a small ($11.52) grocery store transaction according to Garcia-Swartz et al. (2006a), while the same transaction in cash would cost the merchant only $.30. They argue that cards become more competitive for small transactions, when one takes into account consumers' costs of acquiring and carrying cash, as well as the benefits of paying by card.

Further complicating the comparison between account-based and store-of-value models is the fact that in practice, the latter may not be free from serious informational and enforcement frictions. Counterfeiting, virtually eradicated in the developed world, is much more prevalent in developing countries (Judson and Porter 2003). The limited information associated with cash leads to its use in payment for illegal activities (Camera, 2001). Another potential problem with store-of-value systems, as yet unmodeled in the literature, is how to enforce spot trades, i.e., how to keep anonymous, isolated spot trades from degenerating into swindles or simple robbery.[16]

We would argue that these and other practical difficulties have served to constrain the use of store-of-value payment systems, including the use of fiat money. But society is not limited to a choice of fiat money or autarky: instead, society can (and does) create alternative types of payment infrastructures. Specifically it can develop better courts, better surveillance techniques and communication systems, better ways of identifying people, and better recordkeeping. All of these increase the costs of non-adherence in account-based systems.

3. Hybrid payment systems

Most actual payment systems incorporate features from the two basic models outlined above. An idea common to all of these hybrid arrangements[17] is the use of privately issued, transferable debt (inside money) in one form or another. We can think of this debt as privately-, rather than publicly-issued pieces of paper (although nowadays they are more likely to take some electronic form). These pieces of paper derive their value from the issuer's credible promise to redeem them. Through the use of transferable debt, one person's credit (on account) becomes another's means of payment (store of value). Both sorts of verification will be necessary for a successful transferable debt: a payee needs to verify that a note is what it claims to be, and to verify that the issuer is good for it.

Papers such as Cavalcanti and Wallace (1999), Williamson (1999), Kiyotaki and Moore (2000), and Andolfatto and Nosal (2003) investigate a basic model of inside

[16] However, see He et al. (2005).

[17] Our use of the term "hybrid" as a combination of store-of-value and account-based systems, is in contrast to the policy literature, where the term "hybrid" is used to describe wholesale payment features that combine features of gross and net settlement (Willison 2005).

money. In this framework, only certain agents (who function as bankers) are able to transact for credit, because they possess some special advantage—they are easy to locate, famous, or just unusually upstanding individuals. Others cannot borrow on their own recognizance, but must use whatever store-of-value payment mechanisms are available: real assets or outside money. If these are in limited supply, then adding bankers' transferable debt to the set of alternatives extends the economy's payment capacity. In Williamson (1999) and Kiyotaki and Moore (2000) bankers function as financial intermediaries as well, holding illiquid claims while simultaneously issuing liquid debt; in short, bankers are engaged in the task of "liquidity transformation."

In these models, separation between own-debt payors and inside-money payors is strict. In other words, no one in these models ever uses a debt to pay a debt, in marked contrast to the arrangements we observe in practice. Kahn and Roberds (forthcoming) give an expanded model of transferable debt in which agents at various times move from one category to another. The move is occasioned by the mix of assets currently held by the agent (think of the conversion of raw materials to finished product over a production cycle). In this world it becomes useful to pay outstanding debts with other forms of debt, thereby enabling the removal of agents from a "credit chain" (Kiyotaki and Moore 1997) in a timely fashion.

We would argue that this alternative approach to inside money is potentially important for understanding the role of liquidity in firms. In the corporate finance literature, there has been recently been a renewed interest in the precautionary demand for liquidity, and the determinants of a firm's cash holdings (Opler et al.; 1998, Kim et. al.; 1998, Almeida et al. 2004). Focus has centered on the decisions to raise liquidity by foregoing investment projects or engaging in external finance. But the timing of payment is one more method by which a firm adjusts its financial portfolio—the act of payment is a reordering of the seniority of its liabilities. Thus, timing of payments is another means of liquidity management. In order to build a credible model of the timing of payments by financial institutions, and the costs and benefits of delay, an account of payment as settlement of debt is necessary.

Bullard and Smith (2003) present another model of the usefulness of inside money. In their account, inside money economizes on transportation costs. Bullion is more expensive to move than paper is. Suppose a chain of debts needs resolution, and suppose that this chain wanders off temporarily into Frontierland, before returning to an ultimate creditor who is in proximity to the ultimate debtor. Better to send a lightweight store of value following the path, and let the gold move the short distance to make final settlement of the chain. In other words, geographical separation can also make transferable debt valuable. Transportation costs could well be significant for explaining movements of paper and specie in, say the 19th century U.S., although the explanation requires the assumption that a store of value form of payment was for some reason mandatory on intermediate links in the credit chain.

Although often overlooked, the simplest inside-money scheme is in fact "net settlement," a payment system where, by prior agreement, obligations owed *by* a particular agent are automatically cancelled or "set off" using obligations due *from* other participants; remaining obligations must be discharged in some other fashion. By compelling the use of an institution's "due-froms" as inside money, netting in effect allows for mechanical and virtually costless exercise of claims against a counterparty. Kahn, McAndrews, and Roberds (2003) present a model in which netting can effectively reduce the amount of collateral necessary to sustain trade.

Netting is an ancient method of payment that is still widely used today, e.g., in large-value payment systems such as CHIPS and CLS. But netting is also subject to some noteworthy limitations. To be effective, netting arrangements must allow their participants automatically to exercise prior claims against other participants, a power that in some legal settings can conflict with bankruptcy law (Bliss 2003). Netting is also unlikely to work unless all parties involved are reasonably creditworthy. More reliable agents may have an incentive to exclude riskier agents from netting arrangements, or to force them to undergo screening before they can take part (Kahn and Roberds 2003).

3.1. Hybrid systems as suppliers of information

An appealing feature of hybrid arrangements is that they can provide supplemental information on the behavior of individuals over their past trades. Kocherlakota (1998) and related papers have demonstrated that an accumulated pile of money may sometimes be a sufficient statistic for the relevant trading history, but in general it will be desirable to supplement this statistic with additional records. An important function of hybrid (inside-money) payment systems is to provide such information.

Suppose, for example, that trades are only directly observable by the two parties to the trade. Rather than exchange a good for an oral promise, it becomes useful to exchange a good for documentation—evidence that the trade took place. IOU's and receipts are in fact important parts of payments arrangements: in their own ways they are also stores of value—in this case, evidentiary value. (IOU's in particular serve as hostages; redeeming them forestalls lawsuits). Characteristically, IOU's and receipts are one-sided evidence, in the spirit of Lacker and Weinberg (1989): you can easily hide a receipt, but it's difficult to counterfeit one. Kahn and Roberds (forthcoming) explore a costly enforcement environment in which optimal mechanisms include incentives for collecting evidence and presenting to a court in the case of counterparty malfeasance. In this model, inside money in the form of personally issued transferable debt further extends the parameter space in which efficient transactions can be achieved (see also Mills, 2004). In other words, transferable debt that is tied to an agent's identity (e.g., a check) can provide greater evidence of the course of transactions than would the circulation of an anonymous coin or government-issued paper.

It is worthwhile to point out that the need for evidence has nothing to do with spatial isolation in any meaningful sense—in fact it is likely to be compatible with Walrasian mechanisms. The nearest thing we have to Walrasian mechanisms in the real world are the (remaining) open-outcry pits on the Chicago futures exchanges. Nonetheless, in the midst of all that public competition, consummation of an individual trade is accomplished by a pair of traders locking eyes and nodding—an act unobserved by anyone else in the room. Thus, even in this centralized situation, identifying evidence in the form of trade slips is indispensable.

The role of hybrid arrangements in establishing identity is explored in recent papers by Kahn and Roberds (2005) and Martin, Orlando, and Skeie (2006). Devices such as credit cards become a way of reducing the cost of establishing identities. To the extent that such identification is imperfect, fraud in the form of identity theft arises as an equilibrium phenomenon. Payment systems only survive if they keep fraud to a manageable level, but it would be prohibitively expensive to eliminate it entirely.

4. Payments and monetary policy

The widespread use of RTGS systems for payments between banks implies a close connection between payments and monetary policy. Overnight "policy" interest rates (e.g., the fed funds rate in the U.S.) are, after all, simply the prices over which a central bank is willing to provide funds over an RTGS system today (e.g., Fedwire) in terms of a reverse transfer the following day. As noted in the Introduction, the use of RTGS requires that traditional, "overnight" monetary policy be complemented some mechanism for the provision of intraday liquidity. Such a mechanism effectively constitutes an intraday monetary policy—a set of conditions under which a central bank is willing to provide funds during its daily operations. Such intraday credit is usually provided on generous terms, relative to overnight credit. In the U.S., banks were historically given access to free, uncollateralized daylight overdrafts, and even now they are charged a very low (36 basis point) fee, and only for overdrafts that exceed a cap.[18] Other central banks require collateral for access to intraday credit, but conditions for such collateral is often liberal in terms of asset eligibility and haircuts. There is usually no charge for such intraday credit, apart from the implicit cost of posting the necessary collateral.

A number of papers, beginning with Freeman (1996), have sought to explore the dichotomy between overnight and intraday monetary policy.[19] A recurring theme in this literature is that intraday central bank credit serves a fundamentally different function from overnight credit. Papers in this literature argue that while the demand for overnight central bank policy is dictated by the private agents' need to smooth consumption over time, the demand for central bank policy arises from the agents' inability to control the timing of incoming and outgoing payment demands during the day. To the extent that the requirement to use RTGS systems constrains the ability of private agents to contract away these timing mismatches, the implication is that central banks should make intraday credit available on relatively easy terms (Green 2003).

[18] A substantially higher fee is charged if the intraday overdraft is not repaid by the end of the day.
[19] Subsequent papers include Green (1997); Freeman (1999); Zhou (2000); Kahn and Roberds (2001b); Temzelides and Williamson (2001); Williamson (2003); Martin (2004); Fujiki (2006); Lester et al. (2006); Mills (forthcoming).

This literature thus provides some support for existing central-bank intraday monetary policies, but this support is subject to some important caveats. In these models, when private agents are allowed access to more sophisticated arrangements (effectively, some form of net settlement) their need for intraday central bank credit often either diminishes or vanishes altogether. And, as Lacker (2006) points out, in today's world where the "overnight" period of monetary policy spans just a few hours,[20] the distinction between overnight and intraday monetary policy has become less and less relevant. Liquidity now provided via specialized intraday credit facilities could also be provided through other mechanisms—in the U.S. case, Lacker proposes the use of intraday repos combined with payment of interest on excess reserves (mimicking systems already in place in countries such as Australia and New Zealand).

A related literature[21] has provided a more detailed analysis of strategic behavior by banks under various configurations for large-value payment systems. Although this literature confirms the central logic behind RTGS—that injection of central bank liquidity through an RTGS system can stem contagious failures (see Freixas and Parigi 1998; Freixas, Parigi, and Rochet 2000)—this literature has also highlighted some of the drawbacks of RTGS. A common theme is that RTGS systems are inherently more susceptible to coordination problems and dominated outcomes (e.g., delay of payments in an individually rational, but collectively futile attempt to economize on liquidity) than are systems that make use of net settlement.

Such analyses point to a more fundamental, but still relatively unexplored question: why should banks pay each other in central bank balances, i.e., why RTGS? The beginnings of an answer may be found in papers such as Holmström and Tirole (1998) and Kocherlakota (2001), which argue that the coercive powers of government allow it to create collateral (government debt) in states where adverse shocks might undermine forms of collateral available to the private sector. Building on these arguments, Kahn and Roberds (2003) construct a model in which a central-bank sponsored payment system economizes on collateral needed for payment during times of duress, relative to the col-

[20] Fedwire closes at 6:30 p.m. New York time and Asian wholesale systems (e.g., BOJ-NET) open just a few hours later, enabling a virtually continuous flow of funds around the globe.
[21] Including Kobayakawa (1997); Angelini (1998, 2000); Bech and Garratt (2003); Kahn, McAndrews and Roberds (2003); Freeman and Hernandez-Verme (2004).

lateral that would be required by a private-sector system. But this robustness may come at price. Rochet and Tirole (1996) note that central bank involvement in payment systems may introduce a too-big-to-fail problem, diluting banks' incentives to monitor counterparty exposures that may arise in a payments context. A similar argument is made by Fujiki, Green, and Yamazaki (forthcoming), who construct a mechanism design problem in which payment disruptions are sometimes necessary to induce agents to private information about other agents' creditworthiness. In their setup, attempts to mandate an absence of payment disruptions—to remove "systemic risk," as through the requirement to pay via RTGS—may result in a deadweight loss, since it can cause such valuable information to be withheld. On the other hand, the effect of such shocks will also depend on banks' exposures and on the degree of interconnectedness between banks. Kahn and Santos (2006) argue that participation in publicly maintained wholesale payment systems necessarily imposes a high degree of interconnectedness between banks, perhaps in excess of what would be efficient under private contracting; maintenance of a public system may be defensible, however, given that much of the utility of a payment system derives from its universality.

5. The industrial organization of retail payments

Papers in the monetary literature often assume that payment arrangements are administered by a benevolent social planner, or by a club of payment system participants. Such exercises offer useful benchmarks, particularly for wholesale payments where government-sponsored or cooperative arrangements are the rule. At the retail level, however, payment services, apart from those provided by currency, are usually provided by profit-seeking firms. As purveyors of "information goods" (Varian 1998), payment service providers are subject to large economies of scale and various types of externalities. These factors have led to significant concentration in the retail payments industry, which in turn has inspired a burgeoning theoretical literature on the industrial organization of payments.[22]

[22] As this literature has been extensively surveyed elsewhere (e.g., Chakravorti 2003; Hunt 2003; Rochet and Tirole 2004; Evans and Schmalensee 2005; Rochet and Tirole 2006a), this section provides only a brief outline of some of the relevant topics.

At the center of this literature is a debate about the pricing of payments using credit and debit cards. While there seems to be widespread agreement that electronic forms of payment offer the potential for greater efficiency, card-based payments in particular have in many cases remained more expensive (for merchants) than paper alternatives.[23] The price of card payments is invisible to most purchasers, because contractual agreements (*no-surcharge rules*) between the card providers and merchants prohibit merchants from charging extra for card payments. Instead the cost of card payments is reflected in *merchant discounts*, fees paid by merchants to the card companies. When the cards are provided through an association such as MasterCard or Visa, an *interchange fee*, paid by the merchant's bank to the purchaser's bank, comprises a significant portion of the merchant discount. Interchange fees are set cooperatively by the card associations. Card providers have also imposed *honor-all-card rules*, requiring merchants to accept all types of cards issued by the provider, if they accept any.

Do the level of prices for card payments and their peculiar structure reflect the exercise of market power by the card providers, or do they simply reflect the nature of the service provided? Rochet and Tirole (2004) argue that the prevalence of the merchant discounts reflects, most fundamentally, a failure of the Coase Theorem to hold in payment situations, leading to what is termed a "two-sided" markets for payment services. If for example, as in Rochet and Tirole (2002) the typical retail transaction is between a household and a merchant with some degree of monopoly power, charging consumers for card payments may inefficiently discourage the use of credit, compounding the welfare losses due to monopoly. Charging the other side of the market, i.e., the merchants, for card use (and imposing the no-surcharge rule to make such pricing meaningful) eliminates this problem but may inefficiently subsidize card payments. A similar analysis can be applied to honor-all-cards rule (Rochet and Tirole 2006b). Beginning with Baxter (1983), the literature has also defended interchange fees as necessary incentives to guarantee participation by all parties (buyer, seller, and their associated payment service providers) to an account-based transaction (for an updated treatment see e.g., Wright 2003).

While this literature has provided numerous theoretical insights, some notable challenges remain. The first is a closer integration of the I/O literature with the monetary

[23] See e.g., Humphrey et al. (2003); Garcia-Swartz et al. (2006a,b).

literature surveyed above. The second is a need for more consensus building in economists' assessment of current pricing structures in the industry (Evans and Schmalensee 2005; Rochet and Tirole 2006a), a process that has been hindered by the lack of available data.

6. Empirical work in payments

Payment systems are in the business of producing and storing large amounts of data, potentially providing some of the richest datasets in economics. But since the use of this "primary" type of data poses serious confidentiality issues, to date relatively few researchers have enjoyed access to it. A major challenge in this area is to find some means by which researchers may analyze data collected by payment systems, without compromising privacy.

In the comparatively small number of cases where researchers have gained access to payments data, some meaningful studies have been produced. Most of these have been performed by central bank economists with privileged access to data on wholesale systems. In a pioneering effort, Humphrey (1986) used data on the CHIPS system to determine that under its then-prevalent settlement rules, the likelihood of a chain of settlement failures over CHIPS was markedly larger than had previously been supposed. McAndrews and Rajan (2000) examine Fedwire data to show that the pattern of payments made over that system—peaking in the late afternoon as participants await payments from other participants—accords well with the predictions of models in the theoretical literature. Also employing Fedwire data, Furfine (2000) demonstrates a strong statistical link between interbank payments volume and short-term fluctuations in the overnight federal funds rate; Kamhi (2006) provides a similar analysis for Canada's LVTS system. A number of recent studies use data on various wholesale systems to investigate the potential for efficiency gains through the use of queuing algorithms, selective netting, and related techniques (e.g., Johnson, McAndrews, and Soramäki 2004; Galos and Soramäki 2005).

Data recorded by wholesale payment systems can also provide an invaluable record of banks' behavior during times of duress. For example, Furfine (2006) employs Fedwire data to analyze the market for unsecured interbank lending during the Long-Term Capital Management episode, finding little evidence of disruption in this market.

McAndrews and Potter (2002) analyze activity over Fedwire during and shortly after September 11, 2001, providing evidence that the extraordinary liquidity demands experienced during this time resulted primarily from banks' inability to coordinate payment flows.

In the area of retail payment systems, most studies have focused on the issue of consumers' choice of payment method, with a particular emphasis on the shift from paper to electronic payment methods. Lacking access to more precise data, most studies of retail payments have tried to infer consumer payments behavior through household surveys. The most widely employed dataset is the (U.S.) Survey of Consumer Finances (SCF). Studies employing various vintages of the SCF (Kennickell and Kwast 1997; Stavins 2001; Klee 2006a, Mester 2006) have established that demographic factors such as age, income, and education strongly influence consumers' payment choices, and have documented the shift towards electronic means of payment in recent years (this trend is confirmed very extensive payments industry survey reported in Gerdes et al. 2005). Carow and Staten (1999) obtain broadly similar results using data from a survey by Purdue University. Zinman (2006) uses SCF data to infer that debit card use is more common among consumers who are likely to be credit-constrained. Another approach in the literature has been to infer consumer choice from aggregate data on payment systems and data from industry sources. Among the papers in this literature are Humphrey et al. (2000), Humphrey (2004), Bolt et al. (2005), Garcia-Swartz et al. (2006a,b).

While these analyses have been informative, their lack of transaction-specific data has limited researchers' abilities to model the microeconomic behavior of consumer. This shortcoming has been partially addressed in some recent studies that make use of surveys more specifically targeted at consumers' perceptions of various modes of payment. Hayashi and Klee (2003) use data from a survey by American Bankers Association to link consumers' use of electronic means of payment with their use of other information technologies. Loix et al. (2005) find similar results using data from Belgian survey. Jonker (2005) analyzes data obtained in a survey in the Netherlands, indicating that consumers there continue to prefer cash for many transactions, despite the perceived convenience of paying by debit card. Recent papers by Borzekowski and Kiser (2006) and Borzekowski et al. (2006) combine the two prevalent approaches in the empirical literature. They are able to estimate demand functions for various methods of payment, using data from the Michigan Survey (demographic data plus consumers' attitudes toward different types of payment) with data on the "average" characteristics of certain payment types (electronic versus paper, time of use, bank fees, etc.).

Only a few empirical studies of retail payments have been able to use actual payments data, some notable examples being Klee (2004, 2006b), Fusaro (2006), and Rysman (forthcoming). Using data provided by a grocery retailer, Klee finds that a major determinant of consumers' payment choice is simply transaction size, with cash being highly favored for small-value transactions involving just a few items. Analysis of the same dataset indicates a marked transaction-time advantage for debit cards over checks, helping to explain the recent popularity of the former. Fusaro uses a sample of bank accounts to examine behavioral explanations for consumers' preference for debit over credit card transactions. Rysman uses data collected by Visa to determine that while consumers may hold multiple payment cards, in practice they tend to concentrate card payments on a single card network.

In short, empirical studies have provided some instructive snapshots of payments behavior at the retail and wholesale levels. But there are limits to what information can be gleaned from surveys and aggregate statistics—imagine what financial economics would be like if the only available data were surveys of traders and year-end closing prices. Obtaining a more complete understanding of how payment systems actually work, and why people choose to use them, will have to await the availability of more extensive, and more accessible data. And, better empirical work will lead to better theory.

7. Conclusion

Payment is more than a mechanical act. It is, in a sense, the quintessential economic activity, the "glue" that binds together the gains from trade. As such, an act of payment also represents a decision. The choice of whether, when, and how to pay depends on a variety of characteristics of the agents involved in the trade: demographics, value, variability, and liquidity of assets, differential information, risk aversion. The choice also depends on the environment: legal structure enforcing contracts, importance of reputation and ease of damaging it, market interest rates and prices. And in the background, the ability to establish particular terms—indeed the feasibility of trade itself—rests on the availability and cost of a payment system.

In times when payment infrastructure is unchanging, the background becomes opaque. The frequency, the very ordinariness of payment renders it difficult to analyze. We are *not* living in such times. The technological changes of the Internet, the dramatic declines in the costs of computation, and, not least, the institutional changes wrought by international competition and convergence, have led to important changes in the payment systems, both retail and wholesale—and thus made it obvious that payments and system design matter. In such periods renewed interest also arises in the institutional histories and workings of earlier payment regimes. On the horizon are linkups with the law and economics literature, reappraising our understanding of contracts and payments law (for example, the U.C.C.) in terms of the new insights. Also on the horizon are models of macroeconomic fluctuation with a better micro-foundation than the cash-in-advance paradigm provides.

While the empirical work is still in its infancy, the recent study of payments has also led to interesting theoretical insights and potential policy prescriptions. Our examination of this theoretical work leads to the following messages:

1. Despite our habit of modeling of money in geographically-dispersed frameworks, payments still matter even if the world is not decentralized. As long as a time mismatch in trade arises and enforcement of credit arrangements is imperfect, payment systems design will affect an economy.

2. Informational limitations are a key factor in limiting enforcement, and thus are a key factor in the success of a payment system. Verification of identity is central to accounts systems, just as counterfeit protection is central to store-of-value systems.

3. Informational limitations are not set in stone. At a cost, they can be and are overcome. Although anonymity places severe limitations on payment arrangements, anonymity can be pierced at a cost. If the benefit is high enough the institutions for gathering or recording the information develop.

Our models have made a start—but the models have just scratched the surface. Even if others find plumbing boring, plumbers (and plumbing inspectors) won’t be out of a job any time soon.

References

Aiyagari, S.R. and Wallace, N., (1997). Government transaction policy, the medium of exchange, and welfare, J. Econ. Theory 74, 1-18.

Aiyagari, S.R. and Williamson, S.D., (2000). Money and dynamic credit arrangements with private information, J. Econ. Theory 91, 248-279.

Andolfatto, D. and Nosal, E., (2003). A theory of money and banking. Working Paper, Federal Reserve Bank of Cleveland.

Angelini, P., (1998). An analysis of competitive externalities in gross settlement systems, J. Banking Finance 22, 1-18.

Angelini, P., (2000). Are banks risk averse? Intraday timing of operations in the interbank market, J. Money Credit Banking 32, 54-73.

Araujo, L., (2004). Social norms and money, J. Monet. Econ. 51, 241-256.

Bank of England, (2002). Access to Settlement Accounts. Accessed online at www.bankofengland.co.uk/financialstability/paymentsettlementsystems/pdf/boesettleaccs021128.pdf.

Baxter, W.F., (1983). Bank exchange of transactional paper: legal and economic perspectives, J. Law Econ. 26: 541-588.

Bech, M.L. and Garratt, R., (2003). The intraday liquidity management game. J. Econ. Theory 109, 198-219.

Berentsen, A., Camera, G., and Waller, C., (forthcoming). Money, credit and banking. J. Econ. Theory.

Bliss, R., (2003). Bankruptcy law and large complex financial organizations: a primer. Fed. Reserve Bank Chicago Econ. Perspect. 27(1), 48–58.

Bolt, W., Humphrey, D., and Uittenbogaard, R., (2005). The effect of transaction pricing on the adoption of electronic payments: a cross-country comparison. Working paper, Federal Reserve Bank of Philadelphia.

Borzekowski, R., and Kiser, E.K., (2006). The choice at the checkout: quantifying demand across payment instruments. Finance and Economics Discussion Paper, Board of Governors of the Federal Reserve System.

Borzekowski, R., Kiser, E.K., and Ahmed, S., (2006). Consumers' use of debit cards: patterns, preferences and price response. Finance and Economics Discussion Paper, Board of Governors of the Federal Reserve System.

Bullard, J. D. and Smith, B.D., (2003). Intermediaries and payments instruments, J. Econ. Theory 109, 172-197.

Carow, K.A., and Staten, M.E., (1999). Debit, credit, and cash: survey evidence on gasoline purchases, J. Econ. Bus. 51, 409-422.

Cavalcanti, R., and Wallace, N., (1999). A model of private banknote issue, Rev Econ. Dynam. 2, 104–36.

Chakravorti, S., (2003). Theory of credit card networks: a survey of the literature, Rev. Network Econ. 2, 50-68.

Committee on Payment and Settlement Systems, (1990). Report of the Committee on Interbank Netting Schemes of the central banks of the Group of Ten Countries, Bank for International Settlements, Basel.

Committee on Payment and Settlement Systems, (1996). Settlement Risk in Foreign Exchange Transactions, Bank for International Settlements, Basel.

Committee on Payment and Settlement Systems, (2006). Statistics on Payment and Settlement Systems in Selected Countries—Figures for 2004, Bank for International Settlements, Basel.

Corbae, D., and Ritter, J., (2004). Decentralized credit and monetary exchange without public record keeping, Econ. Theory 24, 933-951.

Corbae, D., Temzelides, T., and Wright, R. (2003). Directed matching and monetary exchange, Econometrica 71, 731-756.

Diamond, D. and Dybvig, P., (1983). Bank runs, deposit insurance, and liquidity, J. Polit. Econ. 91, 401-419.

Diamond, P., (1990). Pairwise credit in search equilibrium, Quart. J. Econ. 105, 285-319.

European Central Bank and Federal Reserve Bank of Chicago, Issues Related to Central Counterparty Clearing, Joint Conference, Frankfurt am Main, 03-04 April 2006. Proceedings available at http://www.ecb.int/events/conferences/html/ccp.en.html

Evans, D.S., and Schmalensee, R., (1999). Paying with plastic: the digital revolution in buying and borrowing, Cambridge, MA: The M.I.T. Press.

Evans, D.S., and Schmalensee, R., (2005). The economics of interchange fees and their regulation: an overview. Paper delivered at Interchange Fees in Credit and Debit Card Industries: What role for Public Authorities?, Federal Reserve Bank of Kansas City, May.

Fratianni, M., and Spinelli, F., (2006). Did Genoa and Venice kick a financial revolution in the quattrocento? Working Paper, Oesterreichische Nationalbank.

Freeman, S., (1996). The payments system, liquidity, and rediscounting, Amer. Econ. Rev. 86, 1126-1138.

Freeman, S., (1999). Rediscounting under aggregate risk, J. Monet. Econ. 43, 197-216.

Freeman, S., and Hernandez-Verme, P., (2004). Default and fragility in the payments system. Paper presented at the Economics of Payments, Federal Reserve Bank of Atlanta, March.

Freixas, X., and Parigi, B.M., (1998). Contagion and efficiency in gross and net interbank payment systems, J. Fin. Intermediation 7, 3-31.

Freixas, X., Parigi, B.M., and Rochet, J.C., (2000). Systemic risk, interbank relations, and liquidity provision by the central bank. J. Money Credit Banking 32, 611-638.

Fujiki, H., (2006) Institutions of foreign exchange settlement in a two-country model, J. Money Credit Banking, 38, 697-719.

Fujiki, H., Green, E.J. and Yamazaki, A. (forthcoming). Sharing the risk of a settlement failure, J. Econ. Theory.

Furfine, C.H., (2000). Interbank payments and the daily federal funds rate, J. Monet. Econ. 46, 535-553.

Furfine, C.H., (2006). The costs and benefits of moral suasion: evidence from the rescue of Long-Term Capital Management, J. Bus. 79, 593-622.

Fusaro, M.A., (2006). Debit vs. credit: a rational model of self-control with evidence from checking accounts. Mimeo, East Carolina University.

Galos, P. and Soramäki, K., (2005). Systemic risk in alternative payment system designs. Working Paper, European Central Bank.

Garcia-Swartz, D. D., Hahn, R.W., and Layne-Farrar, A., (2006a). The move toward a cashless society: a closer look at payment instrument economics, Rev. Network Econ. 5, 175-198.

Garcia-Swartz, D. D., Hahn, R.W., and Layne-Farrar, A., (2006b). The move toward a cashless society: calculating the costs and the benefits, Rev. Network Econ. 5, 199-228.

Gerdes, G., Walton, J.K., Liu, M.X., and Parke, D.W., (2005). Trends in the use of payment instruments in the United States, Fed. Res. Bull. 91, 180-201.

Green, E.J., (1997). Money and debt in the structure of payments. Bank Japan Monet. Econ. Stud. 15, 63-87.

Green, E.J., (1999). Panel: thoughts on the future of payments and central banking: we need to think straight about electronic payments, J. Money Credit Banking 31, 668-670.

Green, E.J., (2004). Challenges for research in payments. Invited lecture at the Economics of Payments, Federal Reserve Bank of Atlanta, March.

Green, E.J., and Todd, R.M., (2001). Some thoughts on the Fed's role in the payments system. Fed. Reserve Bank Minneapolis Quart. Rev. 25(1): 12-27.

He, P., Huang, L., and Wright, R., (2005). Money and banking in search equilibrium. International Economic Review, 46(2): 637-670.

Holmström, B., and Tirole, J., (1998). Public and private supply of liquidity, J. Polit. Econ. 106, 1-40.

Humphrey, D.B., (1986). Payment finality and the risk of a settlement failure, in: Saunders, A., White, L.J. (Eds.), Technology and the regulation of financial markets: Securities, Futures, and Banking, Lexington, MA, Heath, 97-120.

Humphrey, D.B., (2004). Replacement of cash by cards in U.S. consumer payments, J. Econ. Bus. 56, 211-225.

Humphrey, D.B., Pulley, L.B., and Vesala, J.M., (2000). The check's in the mail: why the United States lags in the adoption of cost-saving electronic payments, J. Fin. Services Res. 17, 17-39.

Humphrey, D.B., Willesson, M., Lindblom, T., and Bergendahl, G., (2003). What does it cost to make a payment? Rev. Network Econ. 2, 159-174.

Hunt, R.M., (2003). An introduction to the economics of payment card networks, Rev. Network Econ. 2, 80-96.

Intraday Liquidity Management Task Force, (2000). Intraday liquidity management in the evolving payment system: a study of the impact of the Euro, CLS bank, and CHIPS finality, Payments Risk Committee, New York.

Jin, Y. and Temzelides, T., (2004). On the local interaction of money and credit, Rev. Econ. Dynam. 7, 143-156.

Johnson, K., McAndrews, J.J., and Soramäki, K., (2004). Economizing on liquidity with deferred settlement mechanisms, Fed. Reserve Bank New York Econ. Pol. Rev. 10(3), 51-72.

Jonker, N., (2005). Payment instruments as perceived by consumers-a public survey. Working paper, De Nederlandsche Bank.

Judson, R. and Porter, R., (2003) Estimating the worldwide volume of U.S. counterfeit currency: data and extrapolation., Finance and Economics Discussion Paper, Board of Governors of the Federal Reserve System.

Kahn, C.M., (2006). Payment cards—economy wide benefits: the pattern of developments in payments provision. Working paper, The University of Illinois.

Kahn, C.M., McAndrews, J.J., and Roberds, W., (2003). Settlement risk under net and gross settlement, J. Money, Credit, Banking 47, 299–319.

Kahn, C.M. and Roberds, W., (2001a). The CLS bank: a solution to the problem of international payments settlement? Carnegie-Rochester Conf. Ser. Public Pol. 191-226.

Kahn, C.M. and Roberds, W., (2001b). Real-time gross settlement and the costs of immediacy, J. Monet. Econ. 47, 299-319.

Kahn, C.M. and Roberds, W., (2003). Payments Settlement under limited enforcement: public versus private systems. Working paper, Federal Reserve Bank of Atlanta.

Kahn, C.M. and Roberds, W., (2005). Credit and identity theft. Working Paper, Federal Reserve Bank of Atlanta.

Kahn, C.M. and Roberds, W., (forthcoming). Transferability, finality, and debt settlement, J. Monet. Econ.

Kahn, C.M. and Santos, J.A.C., (2006). Endogenous financial fragility and prudential regulation. Mimeo, Federal Reserve Bank of New York.

Kamhi, N., (2006). LVTS, the overnight market, and monetary policy. Working Paper, Bank of Canada.

Kashyap, A., Rajan, R. and Stein, J., (2002). Banks as liquidity providers: an explanation of the co-existence of lending and deposit-taking. J. Finance 57, 33-73.

Kennickell, A.B., and Kwast, M., (1997). Who uses electronic banking? Finance and Economics Discussion Paper, Board of Governors of the Federal Reserve System.

Kiyotaki, N. and Moore, J., (1997). Credit chains. Mimeo, London School of Economics

Kiyotaki, N. and Moore, J., (2000). Inside money and liquidity. Mimeo, London School of Economics.

Kiyotaki, N. and Wright, R., (1989). On money as a medium of exchange, J. Polit. Econ. 97, 927-954.

Klee, E., (2004). How people pay? Evidence from grocery store data. Mimeo, Board of Governors of the Federal Reserve System.

Klee, E., (2006a). Families' use of payment instruments during a decade of change in the U.S. payment system, Finance and Economics Discussion Paper, Board of Governors of the Federal Reserve System.

Klee, E., (2006b). Paper or plastic? The effect of time on check and debit card use at grocery stores. Finance and Economics Discussion Paper, Board of Governors of the Federal Reserve System.

Kocherlakota, N.R., (1998). Money is memory, J. Econ. Theory 81, 232-251.

Kocherlakota, N.R., (2001). Risky collateral and deposit insurance, Advances in Macroeconomics 1(1), article 2.

Kocherlakota, N.R., and Wallace, N., (1998). Incomplete record-keeping and optimal payment arrangements, J. Econ. Theory, 81, 272-289.

Koeppl, T.V. and Monnet, C., (2006). Central counterparties. Paper presented at Joint Conference of the ECB and the FRBC on the role of Central Counterparties, European Central Bank, April.

Kohn, M., (1999). Early deposit banking. Working paper, Dartmouth University.

Kuttner, K.N. and McAndrews, J.J., (2001). Personal on-line payments, Fed. Reserve Bank New York Econ. Pol. Rev. 7(3), 35-50.

Lacker, J.M., (2006). Central bank credit in the theory of money and payments. Invited lecture at the Economics of Payments, Federal Reserve Bank of New York, March.

Lacker, J.M. and Weinberg, J.A., (1989). Optimal contracts under costly state falsification, J. Polit. Econ. 97, 1345-1363.

Lagos, R., and Wright, W., (2005). A unified framework for monetary theory and policy analysis, J. Polit. Econ. 113, 463-484.

Lester, B., Millard, S., and Willison, M., (2006). Optimal settlement rules for payment systems. Paper presented at the Economics of Payments, Federal Reserve Bank of New York, March.

Li, Y., (2001). A search model of money and private debt with applications to monetary policy, Int. Econ. Rev. 42, 925-946.

Loix, E., Pepermans, R., and Van Hove, L., (2005). Who's afraid of the cashless society? Belgian survey evidence. Paper presented at Consumer Behavior and Payment Choice, Federal Reserve Bank of Boston, October 2005.

Lowe, P., (2005). Payments system reform: the Australian experience. Paper delivered at Interchange Fees in Credit and Debit Card Industries: What role for Public Authorities?, Federal Reserve Bank of Kansas City, May.

Martin, A., (2004). Optimal pricing of intraday liquidity. J. Monet. Econ. 51, 401-424.

Martin, A., Orlando, M., and Skeie, D., (2006). Payment networks in a search model of money. Mimeo, Federal Reserve Bank of New York.

McAndrews, J.J. and Rajan, S., (2000). The timing and funding of Fedwire funds transfers, Fed. Reserve Bank New York Econ. Pol. Rev. 6(2): 17-32.

McAndrews, J.J., and Trundle, J., (2001). New payment system design: causes and consequences. Bank Engl. Fin. Stability Rev. 11, 127-136.

McAndrews, J.J. and Potter, S.M., (2002). Liquidity effects of the events of September 11, 2001, Fed. Reserve Bank New York Econ. Pol. Rev. 8(2), 59-79.

Mester, L.J., (2006). Changes in the use of electronic means of payment: 1995-2004. Fed. Reserve Bank Philadelphia Bus. Rev. 0(2), 26-30.

Milne, A.K.L. (forthcoming), Introduction to Clearing and Settlement of Financial Markets: Europe and Beyond, conference held at Cass Business School, London, UK, 16-18 June, 2005, Journal of Banking and Finance. Proceedings available at http://www.cass.city.ac.uk/conferences/clearingandsettlement/index.html

Mills, D.C., (2004). Mechanism design and the role of enforcement in Freeman's model of payments, Rev. Econ. Dynam. 7, 219-236.

Mills, D.C., (forthcoming). Alternative central bank policies for liquidity provision in a model of payments, J. Monet. Econ.

Nosal, E. and Rocheteau, G., (2006). The economics of payments. Policy discussion paper, Federal Reserve Bank of Cleveland.

Ostroy, J.M., and Starr, R.M., (1974). Money and the decentralization of exchange, Econometrica 42, 1093-1113.

Quinn, S., and Roberds, W., (2006). An economic explanation of the early Bank of Amsterdam, debasement, bills of exchange and the emergence of the first central bank. Working Paper, Federal Reserve Bank of Atlanta.

Rochet, J.-C. and Tirole, J., (1996). Interbank lending and systemic risk, J. Money, Credit, Banking, 28, 733-762.

Rochet, J.-C. and Tirole, J., (2002). Some economics of payment card associations, Rand J. Econ. 33, 549-570.

Rochet, J.-C. and Tirole, J., (2004). Two-sided markets: an overview. IDEI working paper, University of Toulouse.

Rochet, J.-C., and Tirole, J., (2006a). Externalities and regulation in card payment systems, Rev. Network Econ. 5, 1-14.

Rochet, J.-C. and Tirole, J., (2006b). Tying in two-sided markets and the honor all cards rule. Mimeo, IDEI, University of Toulouse.

Rysman, M., (forthcoming). An empirical model of payment card usage. Forthcoming in J. Industrial Econ.

Samuelson, P.A., (1958). An exact consumption-loan model of interest with or without the contrivance of money, J. Polit. Econ. 66, 467-482.

Shi, S., (1996). Credit and money in a search model with divisible commodities, Rev. Econ. Stud. 63, 627-652.

Speight, G., Norman, B., and Shaw, R., (2006). The evolution of payment arrangements and central banks. Mimeo, Bank of England.

Stavins, J., (2001). Effect of consumer characteristics on payment instruments. Fed. Reserve Bank Boston New England Econ. Rev. 0(3): 19-31.

Starr, R. M., The Price of Money in a Pure Exchange Monetary Economy with Taxation, Econometrica, 1974, 42 (1), 45-54.

Taub, B., (1994). Currency and credit are equivalent mechanisms, Int. Econ. Rev., 35, 921-956.

Temzelides, T., and Williamson, S.D., (2001). Payments system design in deterministic and private information environments, J. Econ. Theory 99, 297-326.

Temzelides, T., and Yu, J., (2004). Lack-of-recall and centralized monetary trade, Int. Econ. Rev. 45, 1221-1227.

Townsend. R.M., (1980). Models of money with spatially separated agents, in: Karaken, J.H., Wallace, N. (Eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis Press., Minneapolis 265-303.

Townsend, R.M., (1987). Economic organization with limited communication, Amer. Econ. Rev. 77, 954-971.

Varian, H.R., (1998). Markets for information goods. Mimeo.

Williamson, O. E., The Economic Institutions of Capitalism, New York: The Free Press, 1985, p. 177

Williamson, S.D., (1999). Private money, J. Money, Credit, Banking 31, 469-491.

Williamson, S.D., (2003). Payments systems and monetary policy, J. Monet. Econ. 50, 475-495.

Willison, M., (2005). Real-time gross settlement and hybrid payment systems: a comparison. Working paper, Bank of England.

Wright, J., (2003). Optimal card payment systems, Europ. Econ. Rev. 47, 587-612.

Zinman, J., (2006). Debit or credit? Mimeo, Dartmouth College.

Zhou, R., (2000). Understanding intraday credit in large-value payment, Fed. Reserve Bank Chicago Econ. Perspect. 24 (3), 29-44.

Transferability, Finality, and Debt Settlement*

Charles M. Kahn
Department of Finance
University of Illinois
c-kahn@uiuc.edu

William Roberds
Research Department
Federal Reserve Bank of Atlanta
william.roberds@atl.frb.org

May 24, 2006

Abstract

Payment, fundamental to exchange in a decentralized economy, often takes the form of transfers of inside money, i.e., specialized forms of debt. Associated with each type of inside money is a set of rules that governs both the legitimacy of such transfers as means of extinguishing other debts, and the allocation of the ensuing risks.

In this paper we develop a model of debt as inside money. In a simple mechanism design framework we show that transferable debt that can be used to settle other debt obligations with finality can be a welfare improving arrangement in the presence of limited enforcement powers. Transferable debt has two advantages over simple chains of credit: it allows for removal of less-than-perfectly reliable agents from the chain in a timely fashion, and it allows agents to direct payments to the proper party without direct communication with other members of the credit chain.

*The authors are grateful to Dan Bernhardt, Roberto Chang and Bart Taub for helpful comments, as well as to participants in seminars at the Bank of England, the European Central Bank, the Federal Reserve Banks of Atlanta, Cleveland, Kansas City, New York, and Richmond, the University of Iowa, the University of Illinois at Chicago, the London School of Economics, Pennsylvania State University, Purdue University, Stanford University, the Swiss National Bank, Texas Christian University and Tokyo University. Remaining errors are the authors' responsibility. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

Since at least Wicksell (1935), economists have focused on money as a means of overcoming the double-coincidence problem. Before fiat money, goods were commonly traded for gold and gold for goods. But there have long been means of overcoming this problem other than the exchange of gold. Debt, in particular, ties up fewer resources than does the use of precious metal. Nonetheless, not all debt serves as money: only debt which is passed from hand to hand feels like money—mere chains of debt obligations do not. Two characteristics are crucial to the money-ness of debt. The first is *transferability*—the enforceability of the debt should not diminish when the debt is passed to a third party. The second is *finality*—debt can only be considered a good substitute for commodity money when it can be used to discharge other debts.

The transfer of debt has served as a fundamental building block of both historical and contemporary financial systems. The special role of certain types of debt in providing final payment, e.g., bank debt, is firmly entrenched into the laws and business practices of developed economies. In the U.S., for example, a cashier's check (a check drawn by a bank on itself) can serve much the same function as legal tender in terms of its capacity to discharge an obligation.[1]

In this paper we develop a simple story of transferable debt. In our account, transferable debt, when used to extinguish other debts with finality, efficiently effects trades that could not be accomplished by other means. The story is explicitly dynamic: The agents find it desirable to extend credit at one time and then to extinguish that credit through a payment—that is, by transfer of someone else's debt. Thus, unlike previous models of inside money, ours accounts for *settlement of debt,* not simply for exchange of inside money for goods. In our account, transferable debt has two noteworthy advantages: First and fundamentally, it allows for finality—it makes it possible for less-than-reliable agents to be removed from a credit chain in a timely fashion. In addition, it allows agents in a bilateral transaction to receive information concerning the actions of third parties: at some point in the past, the agent with whom I am currently dealing has provided value to other agents. More to the point, if the agent to whom I originally gave the

[1] Despite its legal ubiquity, the concept of a final debt transfer might seem odd from the standpoint of contract theory. Having recourse to debtors in certain eventualities would generally seem to be advantageous from the point of view of risk sharing and useful in generating optimal incentives. Cutting off this flexibility by making a debt transfer final at least requires some justification.

transferable note does *not* present it to me, it must have been used to pay somebody else. These two roles of debt transfer makes it a common and indispensable catalyst of decentralized exchange.

The rest of this section provides a brief introductory account of the model and places the institutions of transferable debt and finality within a historical context. Section 2 provides a mechanism design framework for considering decentralized trading with limited enforcement, which is then formally applied to specialized examples in the subsequent sections. The penultimate section of the paper relates our model to others in the literature, and the final section offers concluding remarks.[2]

In order to demonstrate the significance of transferable debt, it is necessary to compare it with a non-transferable alternative. We introduce a non-transferable instrument for providing evidence—an "IOU." We show how transferable debt provides efficiency gains relative to non-transferable IOUs.[3]

1.1 Introduction to the model

In our model, trade occurs through a sequence of bilateral meetings.[4] Preferences, endowments, and trading opportunities are such that within a period there is never a double coincidence of wants, ruling out barter exchanges. Nor is fiat money available to facilitate exchange. Instead, trade must be carried out on the basis of promises to pay.

We suppose there are three agents, A, B, and C. A can provide a good ("flour"). B can produce a finished good ("bread") using A's input. The finished good is always desired by C, *provided he can acquire it early.* At a later date, C is endowed with a good ("gold") that is desired by both A and B. In other words, C wants to consume bread before he has the assets to pay for it.

The specific sequence of meetings is depicted in figure 1. A and B meet in period 1, when A may supply flour. B and C meet in period 2, when B may supply bread. Following the arrival of C's endowment of gold in period 4, C meets again with B (period 4) and C meets directly with A (period 6).

[2] An earlier version, available from the authors, uses a similar framework to examine a historically prevalent set of rules for assigning risks and effecting payments.

[3] An extended version of the paper also examines a second type of non-transferable evidence, a stylized version of a "receipt."

[4] In our examples, the order of meetings is fixed, but this is because our analysis focuses on the most problematic sequence (see the discussion in section 5 below). The underlying framework is general, and at the cost of additional complexity, the sequence used in our examples could be embedded in a less restricted specification.

We allow two additional meetings between A and B (periods 3 and 5); these meetings will become significant when we compare payments methods. Meetings other than these are not feasible, and in particular there is never a meeting of all three agents.

We assume that enforcement is in the hands of a court. The court cannot observe the trades themselves, but it can carry out limited rewards and punishments, based on reports made by the parties subsequent to the trades. If either party to a trade makes a report, then the court examines the available documentation and may decide to penalize the nonperforming counterparty. The court has a technology available for transferring part of a penalty to other parties. The court's enforcement technology has two limitations: there is a limit to the size of a feasible punishment, and the transfers occur in Okun's "leaky bucket"—part of the receipts are not transferred, but wasted.

Finally, it is costly for agents to access the court. The cost of making a report to the court is stochastic; the draw is privately revealed to the individual just before he decides whether to make a report.

Two-player case In order to understand the issues, first let us focus on the even simpler subcase depicted within the dotted rectangle in figure 1. Two players, B and C, are attempting to consummate a deal: C wants B's bread in the earlier period (period 2) and B wants C's gold in the later period (period 4). If access to the court were costless, B could deliver bread to C with the understanding that C would reciprocate by delivering gold to B later. If C failed to reciprocate B would protest to the court, and the court would punish C. As long as B receives no reward for his announcement he has nothing to gain by lying to the court. Hence, under the threat of a sufficiently large penalty, C would voluntarily transfer his gold to B.

Such an arrangement breaks down, however, when making reports to the court is costly. Then B will no longer make such announcements, unless the court is willing to transfer resources from C to compensate B for the costs of making reports. If the court offers a low amount of compensation to B in such cases, then B will come forward only when his cost of making announcements is low. If his cost is high, B will make no report to the court. As a result, C may be tempted not to deliver gold to B. On the other hand, if the court offers a high amount of compensation to B, the frequency of announcements may make the arrangement prohibitively expensive.

In this case, trade can be sustained by the introduction of physical evidence (figure 2). Suppose C provides an IOU to B in period 2, acknowledging delivery of the bread and, in effect, promising delivery of gold in period 4, with the IOU to be returned by B to C when the gold is delivered. If C does not perform, B can make an announcement of nonperformance to the court. If the complaint is backed by C's IOU, the court can punish C, and reward B. If no IOU is forthcoming from B, then the court can punish B for making a frivolous complaint. In other words, the holder of the IOU is willing to redeem it for something of value; the writer of the IOU wants to redeem it rather than suffer a penalty at the hands of the court.[5] And in equilibrium, no reporting costs are incurred.

Three-player case In the three-player case, a system of IOUs—a "credit chain" in the sense of Kiyotaki and Moore (1997) —can likewise induce the agents to honor their commitments (figure 3): A produces and delivers his good in period 1 to B in return for B's IOU and B delivers bread in period 2 to C in return for C's IOU. C redeems his IOU with a delivery of gold to B in period 4, and B extinguishes his IOU by transferring gold to A in period 5. If an agent makes a complaint to the court, appropriate punishments can be meted out based on an examination of unredeemed IOUs.

For a credit chain to be effective, the punishments available must be severe enough to ensure that B is not tempted to "take the money and run." In other words, a weak point in the enforcement arrangement is ensuring that when B receives gold from C, he is willing to pass gold along to A.

If C knew that such indebtedness existed, trade could be sustained by having C deliver gold directly to A in period 6. But without some form of evidence, this final delivery will also require possibly expensive reporting to sustain it. It can be sustained more cheaply by the use of transferable debt (figure 4). As before, B pays A with an IOU in period 1, and C pays B with an IOU in period 2. Suppose we permit B to extinguish his debt by transferring C's IOU to A in period 3. Then C pays A directly, redeeming his IOU. This arrangement can be sustained by allowing A to bring to the court any IOU he still has after period 6 trade.[6] In this case, relatively mild threats of punishment can enforce payment, and in equilibrium, A never actually has to make a report. In other words, the use of transferable debt allows the elimination of B, the weak link in the credit chain, in a timely fashion. This case is analyzed in detail in the main part of the paper, section 3.

[5]This case is examined in detail in the extended version of the paper.

[6]In our example, the terms specified on the IOU are inessential; more generally, if the penalty imposed by the court were on a sliding scale according to the terms on the IOU, it could be used in various states to sustain a variety of trades.

Extension One source of artificiality in the above example is that every agent in the model knows the exact pattern of trade to be followed, and, therefore, who owes what to whom at any moment. When such information is not generally available, transferable debt takes on a second role, providing information about trades that have taken place elsewhere in the economy.[7] We consider an extension of the three-player case in which transferable debt allows for the elimination of weak links in a credit chain while simultaneously playing the information role.

Suppose that B sometimes can make bread without the use of A's input. Thus he accepts raw material from A only part of the time. This possibility leads to a new difficulty: C knows whether B supplied him bread, and he knows the ex ante probability that B desired A's input; but he does not know whether A actually contributed to its production. Suppose that, in the event that B wants flour from A in period 1, the efficient flow of goods is as in the three-player case (figure 1), and in the event that B does not want flour the efficient flow is as in the two-player case (figure 2).

In this extended model, enforcing a credit chain becomes even more difficult. C is uncertain about whether A has delivered flour to B, i.e., whether B owes a debt to A. If C has to rely, for example, on information from B about whether payment should go to A or to B, then B will be tempted to lie about his own indebtedness, unless the court can again apply heavy penalties to B.

In this case, transferable debt plays two roles. If A delivers flour to B in period 1, then B issues an IOU to A, payable in period 3 by transfer of a third party's (C's) debt to A. C's debt is then ultimately redeemed by the transfer of gold from A to C, upon presentation of C's IOU in period 6. In the event where A does not deliver a good to B, then C's debt stays with B and is redeemed by B in period 4.

The court can enforce such an arrangement as follows: If B issues a debt to A and fails to discharge his debt in the manner described above, A reports this to the court and shows B's debt as evidence. The court announces that B is in default, indicating that any payments by C should be to the court and not to B. The court would also apply penalties to B, sufficient to induce performance, but these could be smaller than before, since by defaulting on his period 3 payment, B removes any chance that he could receive a payment from C in period 5.

In effect, A's loan to B is a short-term loan; it falls due before gold is available to pay it. Payment with transferable debt thus allows for debt owed

[7] Compare with Kocherlakota (1998) and Townsend (1987).

to a certain agent (i.e., by C to B) to "cancel out" another debt owed by the same agent (by B to A). This cancellation plays two roles: it eliminates B from the credit chain, and it informs C as to who is owed the gold. So long as such cancellation is understood by all agents to be a feature of the trading environment, it allows for trouble-free enforcement of creditors' claims, even in cases where other enforcement options are limited.[8] This extension is analyzed in detail in section 4.

1.2 Institutional and historical context

The foregoing example, while stylized, captures the basic idea of all inside-money payment systems: the settling of a debt between two parties by tender of a third party's debt. Such transfers are risky by nature, and associated with every inside-money payment system is a set of rules that govern the allocation of risks that may arise in the course of exchange.

In the legal and historical literatures, the rules are often described by the terms *transferability* and *finality*. Roughly speaking, a debt is "transferable" ("assignable") if a third party who receives the debt retains the same creditor's rights against the debtor as original debt holder. Under contemporary U.S. law, for example, a check or similar instrument may be freely transferred to a third party via endorsement, whereas a credit or debit card payment can only be cleared through certain prespecified channels.[9]

A debt transfer is "final" ("the debt is discharged") when the transfer extinguishes an obligation between two parties. Otherwise put, if discharge has not occurred then one or both parties have recourse, i.e., the right to compel the other side to undertake additional actions in fulfillment of the contract. In practice, finality may hold in some circumstances but not others. A modern check payment, for example, typically does not extinguish an obligation but only suspends the obligation pending settlement of the check. The the higher degree of finality, the more money-like the character of a debt transfer.

Modern notions of transferability and finality are the result of a very long evolution. Medieval debt contracts were generally not transferable (Kohn 1999), but as trade expanded, debts of individuals began to be used as a

[8] This feature of transferable debt corresponds to the idea of net settlement (cf. netting the arrangements analyzed in Kahn, McAndrews, and Roberts 2001). These correspondences are examined in greater detail in section 5.

[9] Strictly speaking, a check is not debt since it represents an order to pay rather than a promise to pay. However, a check or similar instrument may effectively become a debt obligation if it is "accepted," as in a certified check (a check accepted by a bank).

means of payment. Circulating debt became widespread in the Low Countries in 16th and 17th centuries, with the establishment of the legal concept of *negotiability*.[10] Bills of exchange and similar debt "instruments" became "negotiable," or generally acceptable in exchange, because they were freely transferable and their transfer was subject to certain widely accepted finality rules.[11] It is worth emphasizing that bills of exchange and related types of instruments were in use for centuries (in non-circulating form) before the development of negotiability. However, once established, the use of negotiable instruments quickly spread to other countries. Rogers (1995) argues that the adoption of negotiable instruments in 17th and 18th century England resulted in a gradual reorganization of trade, essentially from sequences of spot transactions to something more like the credit chains that prevail today.[12]

The concept of negotiability survives to the present day in U.S. law, where it forms the basis for much of the law governing check payments. As the debt of individuals and non-bank firms no longer circulates as money, negotiability per se is less important than in earlier times (Winn 1998, Mann 1999). The use of bills of exchange and similar instruments persists in less developed economies, however. Ickes (1998) describes the widespread use of "veksels," essentially bills of exchange, by Russian industrial firms during the 1990s.[13] More modern forms of payment, such as credit cards and wire transfers, incorporate different rules for transferability and finality. Nonetheless all of these arrangements share the fundamental feature, in the course of normal trade, of pulling the middle party out of the credit chain.

The formal model below gives some precise economic meaning to the concepts of transferability and finality. Specifically, the model illustrates how transferable debt can implement optimal allocations in cases where a credit chain cannot.

---

[10] See van der Wee (1997) on the origins of negotiable instruments.

[11] Like checks, negotiable bills of exchange were "order instruments" that became debt only after they had been accepted by the party instructed to pay (see Rogers 1995).

[12] A parallel development took place in Edo-era Japan. There, according to Tamaki (1995), a number of merchant bankers (*ryogae*) issued notes payable either at a fixed term, or on demand. Like their European counterparts, these notes circulated as means of payment. We are grateful to Masato Shizume for making us aware of the existence of these arrangements.

[13] Unlike early negotiable instruments, the veksels could not be freely transferred but instead were only transferable to specific parties, usually the next participant in a given supply chain.

2 The Mechanism Design Framework

Although we will deal only with an example where there are six periods and three agents, it should be apparent that the framework described in this section readily extends to arbitrary numbers of agents and periods and arbitrary trading patterns.

Each period consists of a *trading stage* followed by an *announcement stage.* To describe the economy, we must specify not only agents' preferences, endowments, and technologies for transforming goods, but also the technologies for making trades, announcements, and transfers (rewards or punishments).

Endowments, Production Technology and Preferences For simplicity, all goods are indivisible. A is endowed with a unit of flour in period 1. C is endowed with a unit of gold in period 4. B can produce one unit of bread from one unit of flour. The three agents attach the following utilities to goods:

agent	flour	bread	gold
A	v	0	u
B	0	v	w
C	0	u	v

where

$$u \geq w > v > 0 \qquad (1)$$

Recall that bread is only valuable to agent C if it is received before the end of period 2; if it is received later it is worthless to him.

Trading Technology The *trading technology* specifies which agents are matched at each period for trading and the set of feasible offers that an agent can make to his trading partner. (Offering nothing is always feasible.) The series of bilateral meetings is specified in figure 1. In the case at hand, it is feasible for A to offer flour to B in period 1, it is feasible for B to offer one unit of bread to C in period 2, provided he has received the required amount of flour in period 1, and so forth. There is an additional meeting between A and B in period 3, at which neither has goods to exchange. This meeting will only become significant when we introduce transferable debt.

At each trading stage, each agent in a trading pair simultaneously makes a *trade proposal,* which consists of a feasible offer plus a demand from the other individual (not necessarily feasible, since the agent may not know what offers are feasible for the partner). If the proposals match then the trade takes place, otherwise no trade takes place. Only the trading pair observe their own activity. In other words, the trading period models private-information spot transactions.

In each period of our model, at most one side will have goods to trade. The other side may, however, exchange goods for "evidence." Evidence comes in the form of either non-transferable "IOUs" or transferable debt. Both forms represent a promise to repay at a future date. Each is costlessly produced by any agent, and each is impossible for non-issuers to copy or counterfeit. By definition, transferable debt can be passed to third parties; non-transferable IOUs can only be returned to the issuer.

Announcement Technology The *announcement technology* specifies the costs to the agent (in utils) of making reports to the enforcement authority. A report may include evidence. Each player independently decides whether to send a report to the center, and if so what message to send. The center can only observe reports, not the trades themselves. The cost of making a report does not vary with the message an agent decides to send; however, if he decides not to make a report, the agent pays no announcement cost. For convenience, reports are public announcements.

Agent A can make a report in period 6; agent B can make a report in period 4. In each case the agent learns the cost of reporting immediately before he decides whether to make a report. The costs of reporting are independent draws from the distribution G_i for $i = A, B$. We confine attention to the case where reports by C are always prohibitively expensive.[14]

In the extension we will add reports in period 3 (see below).

Transfer Technology The *transfer technology* specifies feasible rewards and fines. For simplicity we will assume that these are all assessed in utils after the final period of play. The transfer technology is parameterized by two numbers, F and α, where $F > 0$ denotes the maximum feasible fine and $\alpha \in (0, 1)$ indexes the efficiency of the transfer mechanism. Rewards must be paid out of the other players' fines, where a fraction $(1 - \alpha)$ of any fine is wasted.

[14] Again, the extended version of the paper considers more general cases.

Mechanism A *mechanism* is a specification of feasible rewards and fines for all agents in the economy as a function of the history of reports. A mechanism generates a game, played by the agents in the economy. In this game, strategies consist of sequences of trade proposals and reports as functions of the history of play, measurable in the usual way with respect to the individual's information set at any round of play. A mechanism *implements* an outcome if it is a subgame perfect Nash equilibrium of the game induced by the mechanism.[15] We will be evaluating mechanisms according to the desirability of the trades that they induce, the cost of the announcements made by the players, and the wastage incurred in running the transfer technology. The ultimate standard is the "efficient mechanism":

Definition 1 *An* efficient mechanism *is one which has a subgame perfect Nash equilibrium in which, along the equilibrium path*

1) The trades are Pareto efficient and participation is individually rational.

2) No announcements are made

3) The center makes no transfers

In other words, in an efficient mechanism, no enforcement costs are incurred.

3 Results

This section demonstrates our primary result: using transferable debt to settle other debts achieves efficiency in circumstances where nontransferable IOUs alone fail.

Clearly, in the absence of enforcement problems the set of individually rational, Pareto efficient trades is as follows: A gives a unit of flour to B; B uses it to make a unit of bread, and gives the bread to C and C gives gold to A. As an alternative to delivering the gold directly to A in period 6, C can deliver it indirectly, passing it to B for transfer to A (periods 4 and 5).

[15]Note therefore, we are dealing with weak-implementation in subgame perfect Nash equilibria. In mechanisms relying on comparing agents' reports, strong implementation is generally not possible. Instead, threats generated by other Nash equilibria of subgames induce desired behavior along the equilibrium path.

Extensions to the case of correlated equilibrium, allowing for a publicly observed signal before messages are sent, would probably increase the generality of the results, at the cost of considerable additional overhead. Introducing cheap talk in the trading stage, while unnecessary in our example, would also be useful in more complex environments.

As a preliminary result, we point out that an efficient mechanism must employ some sort of evidence.

Theorem 2 *If reports by C are prohibitively costly, then there is no efficient mechanism that does not use evidence in its reports.*

Proof. (Outline). An efficient mechanism requires that C give some gold to either A or B. This will not happen unless the mechanism provides a positive probability of punishment to C for misbehavior, which in turn requires that A or B report misbehavior to the center with positive probability. But reports are expensive. Therefore efficiency also requires that no reports are made if C behaves properly.

First, suppose that the mechanism has C give A some gold in period 6. Only A can report misbehavior in this case. We cannot reward A for reporting misbehavior, relative to reporting good behavior, because that will leave an incentive for non-truthful reporting. But reports of bad behavior must be given with positive probability. Thus reports of good behavior must also receive positive rewards, meaning that the reporting mechanism will be costly with positive probability even if C behaves properly.

Thus the mechanism cannot pay anything to A in period 6. Similarly, the mechanism cannot have B pay A in period 4, otherwise it would require A to report on good behavior by B. But this means A can receive no gold, violating the individual rationality requirement for A. ■

3.1 Non-transferable IOUs.

In this subsection, we restrict evidence to non-transferable IOUs. These are costlessly produced by the issuer in exchange for goods and can be redeemed in subsequent trades between the issuer and the initial recipient. IOUs cannot be passed to third parties.

Theorem 3 *There exists an efficient mechanism with non-transferable IOUs if and only if the following conditions hold:*

$$G_A(2\alpha F) \geq w/F \tag{2}$$

$$u \geq \int^{G_A^{-1}(w/F)} G_A(t)\, dt \tag{3}$$

Corollary 4 *The efficient outcome can be enforced by the following mechanism: if A presents B's IOU to the center, he receives a reward. The reward is set at a level that would induce him to report a fraction w/F of the time.*

If A presents, then B and C each receive the maximal punishment $-F$. If B presents C's IOU to the center, all *agents receive maximal punishment (i.e. B is not permitted to provide evidence). In all other cases the court provides neither rewards nor punishments.*

Proof. See Appendix. ■

The following trade pattern achieves the efficient outcome: B provides A an IOU in return for A's endowment. C provides B an IOU in return for B's endowment. C redeems his IOU from B with gold, and B uses the gold to redeem his IOU from A. In other words, a credit chain of IOUs implements the efficient outcome. If everybody delivers, no IOU is left at the end.

In effect, the conditions of the theorem are requirements on the parameter F: It must be feasible for the court to impose sufficiently large punishments and rewards. If B is subject to the maximal penalty, then having A report a fraction w/F of the time A does not receive payment, is just enough to induce B to pay A. To make a report with that frequency requires a reward of $G_A^{-1}(w/F)$. Condition (2) requires that it be feasible to reward A enough to induce this frequency of reporting, and condition (3) requires that given such a reward, A not be tempted to fraudulently claim misbehavior by B. As the maximum feasible punishment F increases, frequency of punishment can decrease, and thus the reward necessary to induce A to announce becomes lower, and thus less likely to incite fraudulent behavior.

If A does not receive gold from B, then both B and C are punished. Collective punishment is adequate for the job at hand. We simply want to thwart misbehavior; when it happens, it is of no consequence which of the two possible miscreants is guilty. Obviously, in more complex environments, a more refined mechanism would be necessary.[16]

[16] Collective punishment does however have a limitation. While the mechanism implements with a subgame perfect Nash equilibrium, the equilibrium does not survive some reasonable refinements. In particular, even if C did not receive bread from B, as the mechanism stands he would prefer to pass the gold to B for transmission to A, rather than face the possibility of punishment due to A's announcement. This means that B is in a powerful position in period 2. If he were to refuse to supply the bread to C, it would be reasonable for C to take this as a signal that B will make an offer to receive the gold in period 5 without returning an IOU. C would find it better to comply with this extortion than to suffer the expected penalties from A's announcement. Since this deviation requires implicit coordination by B and C, it does not violate subgame perfection, but it does violate some refinements of perfection.

The extended version of this paper also provides necessary and sufficient conditions for implementing with non transferable IOUs under a stronger notion of equilibrium.

3.2 Transferable Debt

So far we have only allowed non-transferable IOUs. When debt is extinguished, the IOU is returned to the debtor; in the meanwhile it has remained in the hands of the creditor. In contrast *transferable debt* can be passed from the creditor to a third party. In this section we demonstrate the usefulness of transferability, by showing that it relaxes the requirements for mechanisms to deliver the efficient outcome. As before, transferable debt is costless for the issuer to manufacture and impossible for other parties to counterfeit. A party other than the initial debtor can only deliver that debtor's transferable debt if he has previously received it.

Allowing for debt to be transferable increases the options available in designing a mechanism. Thus any mechanism which was feasible with non-transferable IOUs remains feasible. The following theorem demonstrates that use of transferable debt leads to efficiency in situations where it could not otherwise be sustained.

Theorem 5 *The efficient outcome can be enforced by a transferable debt if and only if the following conditions hold:*

$$G_A(2\alpha F) \geq v/F \tag{4}$$

$$u \geq \int^{G_A^{-1}(v/F)} G_A(t)\, dt \tag{5}$$

Corollary 6 *The following mechanism will enforce a transferable debt arrangement: if A presents B's IOU or C's transferable debt to the center in period 6, then B and C receive maximal punishment, and A receives a reward of a magnitude designed to induce him to announce a fraction v/F of the time. If B presents evidence to the center in period 4, A, B, and C all receive maximal punishment. In all other cases the court provides neither rewards nor punishments.*

Proof. See Appendix. ∎

Thus, with transferable debt, trade takes the following form: In period 1, A trades flour for B's IOU. In period 2 B trades bread for C's transferable debt. In period 3, A returns B's IOU in exchange for C's transferable debt. In period 4 and 5 no trade takes place, and in period 6 C redeems his transferable debt with gold.[17]

[17] Under conditions analogous to those presented in the appendix in the proof of the previous theorem, the mechanism can also be made immune to the extortion by player B described in footnote 16.

Comparing Theorem 5 with Theorem 3 reveals the advantage of transferable debt. Since $v < w$, the set of enforcement systems which can support the efficient outcome has increased. It is only necessary to threaten punishment a fraction v/F of the time—that is, efficiency can be achieved with less dramatic fines and rewards.

The result of this section points to one important advantage of transferable debt with finality: A less-than-reliable individual can settle a debt and be removed from a credit chain, allowing payment to bypass him and making compliance easier.

4 An Extension with Uncertainty

In the presence of uncertainty transferable debt not only permits the timely elimination of an individual from a credit chain, it also provides evidence to the debtor as to who should receive ultimate payment. In this section we provide an example incorporating uncertainty in order to demonstrate this point. The example combines the two-player and three-player cases.

This example is important to our argument because it demonstrates the robustness of the difference between credit chains and transferable debt. In the previous section, because there was no underlying uncertainty, the correct set of trades in equilibrium was automatically known by all individuals: In that section, the problem is that the gold gets into the wrong hands at some point in the process. Perhaps transferable debt would not be needed if only the arrangement could have been set up initially to route the gold correctly. In general, structures in which all agents meet at the beginning of time severely limit the usefulness of transferable debt.[18]

Thus to assure the robustness of the results of the previous section, we need to consider models in which initial meetings are infeasible. One natural case of this sort occurs when the raw goods supplier is unknown at

[18]In the certainty case, it can be shown, for example, if the meeting in period 1 were *trilateral*, then a new type of credit chain would become effective. In period 1, let C hand an IOU to A, let A hand flour to B and let B hand an IOU to C. This new credit chain, with C in the middle, is enforced entirely by announcements by A. If A presents C's IOU to the court in period 6, C and B are punished. If B fails to deliver to C, then C sends a piece of evidence to A in period 6, which A may present to the court. If he does, then A receives a reward, B receives a punishment and C receives neither reward nor punishment. (The only difference between this "early" credit chain and transferable debt arises because the reward that A receives is limited by the fact that C does not contribute to the total; as we add more resources from which rewards can be paid—for example, additional agents in the economy—then the difference disappears.)

the beginning date, as would occur, for example, if the identity of the raw goods supplier were randomly determined later. To keep things simple, we will have two possible suppliers of raw materials: A and B himself.

Specifically, we modify the model as follows: with probability $\frac{1}{2}$, meetings and preferences are as described previously. With probability $\frac{1}{2}$, player B does not need A's flour in order to make a unit of bread. In this case, player A does not meet player B in period 1 but all other meetings are unaffected. B and C's preferences are as before. A can make reports in periods 3 or 6 and B can make reports in period 4. The costs for A in periods 3 and 6 are independent draws.

We first examine exchange using non-transferable debt, extending theorem 3 to the uncertainty case.

Theorem 7 *In the uncertainty case, there exists an efficient mechanism with non-transferable IOUs only if there exists an efficient mechanism with non-transferable IOUs in the certainty case. If $w \geq 2v$ then there conditions for existence of an efficient mechanism are identical in the certainty and uncertainty case.*

Proof. See Appendix. ■

Thus, in general, it is more difficult to achieve efficient allocations with credit chains in the uncertainty case than in the certainty case. Nonetheless, in some conditions, the requirements for efficient outcomes can still be relaxed by introducing transferable debt:

Theorem 8 *Suppose $2v < w < F$. Then following conditions are sufficient for a transferable debt arrangement to be efficient.*

$$G_A(2\alpha F) \geq w/(w+F) \tag{6}$$

$$u \geq \int^{G_A^{-1}(w/(w+F))} G_A(t)\, dt \tag{7}$$

$$G_B(2\alpha F) \geq v/F \tag{8}$$

$$w \geq \int^{G_A^{-1}(v/F)} G_A(t)\, dt \tag{9}$$

Corollary 9 *The following mechanism will enforce a transferable debt arrangement: if A presents B's IOU to the center in period 4 or C's transferable debt to the center in period 6, then B and C receive maximal punishment, and A receives a reward of a magnitude designed to induce him to* *announce a fraction $w/(w+F)$ of the time. If B presents evidence to the center in period 4, A and C receive maximal punishment and B's reward is sufficient to induce him to report v/F of the time.*

Proof. See Appendix. ■

If A meets B in period 1 he offers to trade flour for an IOU. In period 3 if he has an IOU, he offers to trade it for transferable debt. If he has B's IOU at the end of the period he announces it to the center if the draw of costs is sufficiently low. He makes no offers in period 5. If he has C's transferable debt in period 6, he offers to trade it for gold. If he still has C's transferable debt at the end of period 6 he announces it to the center if the draw of costs is sufficiently low. In the mechanism if B is declared bankrupt, subsequent announcements by him are ignored. (Intuitively, once B is declared bankrupt, C makes payment directly through the court to A, rather than paying B).[19]

Trade in this mechanism works as follows: In period 1, if A appears, he trades flour for B's IOU. In period 2 B trades bread for C's transferable debt. In period 3, if A holds B's IOU, he trades it for C's transferable debt. In period 4, if B has not been declared bankrupt and if B possesses C's transferable debt, then B trades it for gold. In period 5 no trade takes place. In period 6, if A possesses C's transferable debt, C redeems it for gold.

Since $w/(w+F) < w/F$, Theorem 8 relaxes the conditions on G_A relative to Theorem 7.[20] On the other hand new conditions on G_B have been introduced, which were not there in the credit chain. These conditions are stricter precisely because by allowing collective punishment we have biased the account in favor of credit chains, in which A's announcements are used to provide indirect punishment to C. Were we to restrict collective punishment we would make the advantage of transferable debt even greater.

The example we use further restricts the advantage of transferable debt by confining our examination to mechanisms in which A is given only one chance to use each form of evidence: he is punished if he presents B's IOU

[19] Since bankruptcy never occurs on the equilibrium path, we make the payment from C to A maximal for convenience. If we extended the model in such a way that agents occasionally went bankrupt, it would be important to consider the minimal payments necessary from A to make the mechanism effective.

[20] Note that the parametric restriction $w < F$ of the theorem is implied by (2) and hence is a necessary condition for the credit chain to be an efficient mechanism. Hence the result that transferable debt relaxes the restrictions on efficiency holds trivially for $F < w$.

to the court late (that is, in period 6). We do this to emphasize that the advantage of the transferable debt comes not from the increase in the number of reporting opportunities, but from the *timeliness* of the report on B's misbehavior that transferable debt offers.

Note that the implementation of this arrangement requires that A have the ability to contact the center at an earlier stage, and that the center have the ability to make announcements. These additional facilities provided no additional benefit or power to the mechanism without uncertainty. The communication only is of use because it allows the center to check in a timely way on whether B has transferred value to A in the form of transferable debt.[21] Without transferable debt earlier announcements would have no bite. Moreover, note that these announcements are only used off the equilibrium path.

Conditions (6-9) are sufficient, not necessary. By exploiting the fact that A has two opportunities to make a report, and thus two chances at low costs of reporting on B's malfeasance, we could expand the set of efficient implementation slightly.

5 Relationship to the literature

In Arrow-Debreu economies, no one makes payments. Instead all agents keep running tabs with a reputable and powerful central authority, secure in the knowledge that budgets will be balanced in the fullness of time. Payments only become necessary with diminution of the central authority and a consequent limitation on agents' reliability: under certain circumstances, an agent is "good" for debt only up to a limit (less than the value of his future income and possibly zero). If there is ample liquidity, i.e., if there are sufficient durable assets and such assets are attachable as collateral, payments can be made in these assets. Techniques for economizing on liquidity become important when such assets are scarce, or where the legal structure's ability to enforce takings of collateral deteriorates. This is the basic rationale for the circulation of debt claims.

Various analyses of circulating debt can be found in the recent literature on inside money (e.g., Freeman 1996, 1999, Cavalcanti and Wallace 1999, Kiyotaki and Moore 2000, Bullard and Smith 2003, Mills 2004). Each of these approaches contains at its core a cycle of trade that begins with the

[21] In effect, a default by B (failure to transfer value to A) renders B unable to enforce debt against another agent (C). This is a standard consequence of default in models with limited commitment (see for example, Azariadis and Lambertini 2003).

issue of debt and ends with its redemption. A common conclusion is that, as long as enforcement is possible at the end of the cycle, that same enforcement can be used to give value to promises traded at earlier stages and thus to allow mere promises to circulate as assets do. Often such enforcement is possible, or economical, only for a certain class of agents. The creditworthiness of these "strong credits" gives rise to the circulation of their debts, which in equilibrium may relax liquidity constraints for all agents. In other papers (Williamson 1992, Williamson 1999, Temzelides and Williamson 2001), fairly broad classes of agents may issue circulating claims, but if there is private information concerning the creditworthiness of the issuers, the potential efficiencies conferred by circulating debt may be undermined by adverse selection.

Despite apparent differences in structure, our approach borrows much from the literature described above. As in many of the above models, trade is restricted to a succession of bilateral encounters, enforcement is limited, and only the debt of the strongest credit circulates. What is different in our approach, however, is the enforceability of agents' debt is not associated with their inherent creditworthiness (all debtors are subject to the same enforcement technology), but instead with the mixture of assets held. A decision to re-order holdings—to pay a debt—changes the reliability of the agent. Transferable debt becomes complementary with illiquid debt: even an inherently unreliable "middle" agent can temporarily borrow with the assurance that transferable debt will be used to pay off later. The circulation of debt is as much as a cause of creditworthiness as a consequence of it.

The advantages of circulation in our structure can be illustrated using a Wicksell triangle example from Kiyotaki and Moore (2000). Their example, depicted in figure 5, while apparently similar to our economy, nonetheless differs in a fundamental way. In our model we want A to transfer some good to B who then transfers some good to C who then transfers some good to A. In the Kiyotaki and Moore example, the first two transfers are reversed in time: the desired arrangement is for B to transfer some good to C and then for A to transfer some good to B and then for C to transfer some good to A. Because the transfers are in this reverse order, a single piece of inside paper money can pass from hand to hand to accomplish all the trades in the Kiyotaki-Moore example: C passes the paper to B who passes the paper to A who passes it back to C. Such timing is natural in an endowment economy, where the typical individual receives the payment and then uses it to buy goods.[22]

[22]This timing also underlies cash-in-advance economies.

In a production economy, a producer receives raw materials and uses them to make finished product. In other words, there is a need for working capital. If inside claims must be used to accomplish the work, the number of claims needed will be much larger in our economy than in the Kiyotaki-Moore example. Thus in our economy, unlike theirs, there is a role for settling debt with other debts.[23]

Crucial to our story of finality is the idea of setoff. Even with limited enforcement, creditors' priority can be maintained as long as debts are allowed to cancel each other (which is effectively what occurs when B transfers C's debt to A). The role of final debt transfer in this regard is closely related to that of net settlement of payment obligations. What is different here is that the practice of net settlement is typically associated with centralized clearing arrangements, whereas final payment by debt transfer can and does occur in decentralized settings.

Our emphasis on the informational and evidentiary role of transferable debt is reminiscent of Kocherlakota (1998) on money as a substitute for "memory." But the differences from that work are as important as the similarities. Kocherlakota's main result compares fiat money with centralized record keeping. It concludes that fiat money is in general inferior to centralized record keeping, although in some circumstances it can perform the same role. Our emphasis is on the comparison of two different sorts of debt: transferable and non-transferable. Strictly speaking, Kocherlakota's result has nothing to say about such a comparison; it only says that either form of debt is inferior to centralized record-keeping.

Thus the paper which is closest in spirit to our investigation of the information and evidentiary roles of transferable debt is Townsend (1987). That paper considers the role of "tokens" in overcoming trading frictions posed by private information and spatial separation. In Townsend's setup, "patient" agents may lack incentives to provide consumption goods in an early period, because there is a risk that they may be relocated, where such relocation eliminates all records of their having provided goods. Since an agent's patience is private information, all agents claim to be patient once relocated.

[23]In a more advanced model, Kiyotaki and Moore (2000) have roles for both liquid and illiquid debt. Compressed into the context of the example, they consider the case where C is simultaneously both more reliable than other agents and better able to enforce debts against them. Goods can then flow from A to B to C (as in our model) if C both issues "liquid" debt (to A) and holds "illiquid" debt (from B). Note that this is in effect a type credit chain (from B to C to A) different from the one considered above (from C to B to A), and a natural one to consider given C's advantages. (See also footnote 18.) Unlike our setup, however, it does not lead to a role for the settlement of one debt with another.

These frictions may be overcome if patient movers carry with them a token as proof of having provided a consumption good at their original location. Once at a new location, patient agents can present tokens to the planner and be rewarded with higher levels of late consumption.

Townsend's tokens thus play an evidentiary role similar to that of the IOUs considered above. But again, in Townsend's model transferability is not necessary for the tokens to be useful. Townsend's tokens can be presented to the planner at zero cost and, for the simplest cases, are able to completely overcome the trading frictions he considers. There is no incentive problem for the planner in honestly evaluating the tokens. Absent private information, tokens would not be needed, and even with private information, tokens can be reminted for each new round of use, without any social loss. In our environment, by contrast, enforcement is always costly even in the presence of evidence. Hence the existence of evidence may not in and of itself deliver efficiency. Non transferable IOUs can still be used to pass information from one individual to the next in a chain, but they do not provide the necessary incentives. Transferable debt transfers the evidence itself, not just a report of the evidence.

The other literature with which this account has important links is the literature on trade credit. The advantages and disadvantages of trade credit as an alternative to bank lending have been examined by a variety of authors. Frank and Maksimovic (1998), for example, emphasize the advantage of the supplier of raw materials as a lender to the downstream firm, because of comparative advantage at using repossessed collateral. In the context of emerging markets of Eastern Europe, Hege and Ambrus-Lakatos (2000) emphasize the threats inherent in continuing relationships among chains of upstream and downstream firms as the reason that trade credit can provide an imperfect substitute for bank lending in times of illiquidity.

However, Ickes' (1998) description of veksels in Russia emphasizes several aspects of the arrangement that seem to support our view of these instruments as important for their transferability. In particular, payment in the form of veksels was less subject to opportunistic diversion than payment in relatively anonymous bank funds. By diminishing the scope for successive defaults, Ickes argues, the use of veksels allowed firms to maintain the integrity of their supply chains in spite of difficulties posed by the prevalent legal environment.

6 Conclusion

Inside money is first and foremost, debt, but it is debt with a difference. Associated with every type of inside money is a system of rules that govern the circumstances of monetary transfer. The most critical rules are those that determine when a transfer may occur and when a transfer discharges another debt.

In a model with centralized, but limited enforcement, we have shown that it can be valuable to allow the use of transferable debt to discharge other debt. Not only does transferable debt provide useful information about the behavior of individuals at a distance from the current transaction, it also allows less-than-perfectly-reliable agents to exit a credit chain in a timely fashion. We would argue that these two features are the central logic behind the use of transferable debt as payment.

Laws governing the use of negotiable instruments as payment have existed for centuries in essentially unaltered form (at least within the Anglo-Saxon legal tradition; see Winn 1999). These are becoming less relevant, however, as paper-based payments instruments are being supplanted by electronic, centralized forms of funds transfer. Electronic systems can offer their participants a higher degree of assurance than paper-based systems, both in terms of payment finality and in terms of protection against fraud. There are downsides to such systems, however, including lack of flexibility (e.g., one must have a merchant account to receive a credit card payment) and a loss of privacy (since there is a centralized record of all transactions). A comparative analysis of traditional and centralized systems for the transfer of inside money, as well as a comparison of the finality aspects of inside and outside money, should provide fertile ground for future research.

7 Appendix: Proofs

A mechanism specifies payments $P_i(j,k)$ by the center to each player i, where $j \in \{A,B\}$ indicates who made an announcement and $k \in \{B,C\}$ indicates whose debt is provided as evidence. (In a credit chain mechanism, A never holds C's debt.) In an efficient mechanism, no payments are made if no one makes an announcement. Without loss of generality, we assume that an agent is penalized for making an announcement without providing evidence. All mechanisms are subject to the following feasibility restrictions:

$$
\begin{aligned}
P_i(j,k) &\geq -F \text{ for all } i,j,k && (10)\\
\sum_i \max\{P_i(j,k), \alpha P_i(j,k)\} &\leq 0 \text{ for all } i,j,k && (11)
\end{aligned}
$$

Each mechanism is also subject to incentive restrictions described within each proof.

7.1 Theorem 3 and Corollary 4

Proof. When A holds B's IOU at the end of period 4, he will choose to report if $t_A < P_A(A,B)$. When B holds C's IOU at the end of period 6, he will choose to report if $t_B < s$, where

$$s = P_B(B,C) - G_A(P_A(A,B))P_B(A,B) \tag{12}$$

Restrictions (13-17) are the necessary and sufficient incentive conditions for efficient behavior by the players in each period:

$$w + G_A(P_A(A,B))P_B(A,B) \leq 0 \tag{13}$$

$$\int^{P_A(A,B)} G_A(t)\,dt \leq u \tag{14}$$

$$v + G_B(s)P_C(B,C) + (1 - G_B(s))G_A(P_A(A,B))P_C(A,B) \leq 0 \tag{15}$$

$$G_A(P_A(A,B))P_B(A,B) + \int^{s} G_B(t)\,dt \leq 0 \tag{16}$$

$$v + G_A(P_A(A,B))P_B(A,B) \leq 0 \tag{17}$$

(13) says in period 5 B must prefer to give up gold if he holds it. (14) says in period 5 A must prefer to accept the gold rather than to hold onto B's IOU. (To see this, note that the left side, by integration by parts, equals $\int^{P_A(A,B)} P_A(A,B) - t\ dG(t)$, the expected reward, net of costs, from providing the court with the IOU). (15) says C must prefer to give up the gold

in period 4 (rather than run the risk of either A or B making a subsequent report). (16) says B must prefer to accept the gold in period 4 than to hold onto C's IOU. (17) says B must prefer to give C the bread in period 2; however, this condition can be dropped, as it is implied by conditions (1) and (13). C's willingness to enter the trade in period 2, and A's willingness to enter the trade in period 1 are also guaranteed by (1). B receives 0 in the efficient arrangement, therefore he is indifferent between participating and not participating.

Our concern is to find necessary and sufficient conditions for a feasible set of values $P_i(j,k)$,to exist. Without loss of generality, we can set $P_i(j,k) = -F$ for $i \neq j$ (the only case where this is not immediate is $P_B(A,B)$, because this variable has an indirect effect through s. The derivatives of the left sides of (16) and (15) with respect to $P_B(A,B)$ are, respectively

$$-G_A(P_A(A,B))G_B(s) + G_A(P_A(A,B)) > 0 \tag{18}$$

$$(-F + G_A(P_A(A,B))F)\frac{d}{dP_B(A,B)}G_B(s) > 0 \tag{19}$$

In each case the constraint is relaxed by reducing $P_B(A,B)$.)

Using (12) to eliminate $P_B(B,C)$, we can rewrite the problem as follows:

Find $P_A(A,B)$ and s such that

$$P_A(A,B) \geq -F; s - G_A(P_A(A,B))F \geq -F \tag{20}$$

$$P_A(A,B) \leq 2\alpha F;\ s - G_A(P_A(A,B))F \leq 2\alpha F \tag{21}$$

$$w \leq G_A(P_A(A,B))F \tag{22}$$

$$\int^{P_A(A,B)} G_A(t)\, dt \leq u \tag{23}$$

$$v \leq G_B(s)F + (1 - G_B(s))G_A(P_A(A,B))F \tag{24}$$

$$\int^{s} G_B(t)\, dt \leq w + G_A(P_A(A,B))F \tag{25}$$

Note that (1) and (22) imply (24). Given the remaining conditions, we can, without loss of generality set s at its minimum value according to (20). The only restrictions that are not automatically satisfied are

$$\begin{aligned} P_A(A,B) &\leq 2\alpha F \\ w &\leq G_A(P_A(A,B))F \\ \int^{P_A(A,B)} G_A(t)\, dt &\leq u \end{aligned} \tag{26}$$

There exists $P_A(A,B)$ satisfying these conditions (which we can without loss of generality take as $G_A^{-1}(w/F)$) if and only if (2-3) are satisfied. ■

7.2 Theorem 5 and Corollary 6

Proof. A will choose to report if $t_A < P_A(A,.)$. B will choose to report, when he holds transferable debt, if $t_B < s$, where s is defined in (12).

Restrictions (27-32) are the necessary and sufficient incentive conditions for efficient behavior by the players in each period:

$$v + G_A(P_A\,(A,C))P_C(A,C) \leq 0 \tag{27}$$

$$\int^{P_A(A,C)} G_A(t)\;dt \leq u \tag{28}$$

$$0 \geq \int^{s} G_B(t)\;dt + G_A(P_A(A,B))P_B(A,B) \tag{29}$$

$$u \geq P_A(B,C)G_B(s) + (1 - G_B(s)) \int^{P_A(A,B)} G_A(t)\;dt \tag{30}$$

$$v + G_A(P_A(A,B))P_B(A,B) \leq 0 \tag{31}$$

$$u \geq v + G_A(P_A(A,B))P_C(A,B) \tag{32}$$

(27) says in period 6 C must prefer to give up gold if he holds it. (28) says in period 6 A must prefer to accept the gold rather than to hold onto C's transferable debt. (29) says B must prefer to pay A in period 3 with C's transferable debt. (To see this rewrite the condition as follows:

$$0 \geq \int^{s} (P_B(B,C) - t)\;dG_B(t) + (1 - G_B(s)(G_A(P_A(A,B))P_B(A,B)).$$

The right side is the payoff from retaining C's debt: he may get the reward; but if he decides not to announce, he runs the risk of A's retaliation. The left side is the continuation payoff from behaving correctly—zero since all payments owed to C will have already been received before he makes the announcement.) (30) says A must prefer to return B's IOU for C's transferable debt, rather than run the risk of B's complaint on hopes of receiving the reward from announcing. (31) says B must prefer to give C the bread in period 2. (32) says C must prefer making a trade in period 2 to autarky. Finally, entering a trade in period 1 is individually rational for A because of (1), and B is indifferent between autarky and entering the trade.

We can assume $P_i(j,k) = -F$ for $i \neq j$. (For $P_B(A,B)$ condition (29) is identical to condition (16); the derivative of the right side of (30) is

$$\left(P_A(B,T) - \int^{P_A(A,B)} G_A(t)\,dt\right) \frac{d}{dP_B(A,B)} G_B(s) > 0 \qquad (33)$$

As before, all incentive constraints are relaxed by reducing $P_B(A,B)$.) Note that $s \geq 0$ without loss of generality. Use (12) to eliminate $P_B(B,C)$, rewriting the problem as follows:

Find $P_A(A,B), P_A(A,C), s$ satisfying

$$\begin{gathered}
2\alpha F \geq P_A(A,j) \geq G_A^{-1}(v/F) \text{ for } j \in \{B,C\} \\
2\alpha F + G_A(P_A(A,B))F \geq s \geq 0 \\
\int^{P_A(A,C)} G_A(t)\,dt \leq u \\
0 \geq \int^{s} G_B(t)\,dt - g_{AB}F \\
u \geq -FG_B(s) + (1 - G_B(s)) \int^{P_A(A,B)} G_A(t)\,dt
\end{gathered} \qquad (34)$$

Now $P_A(A,C)$ exists satisfying the lot if and only if (4) and (5) are also satisfied. Given these two conditions hold, we can set $s = 0$ and $P_A(A,B) = P_A(A,C) = G_A^{-1}(v/F)$ to satisfy (34) with no additional restrictions. ∎

7.3 Theorem 7

Proof. Without loss of generality, we can ignore announcements by A in period 3: they would only be useful in order to report misbehavior by B, but B's only misbehavior at that point would be non-acceptance of the initial supply of flour. Thus such an announcement would have to be made by A without evidence. But if such an announcement were profitable to A, it would also be profitable for A to (falsely) make announcements when A had no useful good to supply in the first place.

A will choose to report, when he holds an IOU, if $t_A < P_A(A,B)$. B's decision about reporting will depend on whether A has evidence he can use in period 6. If A has no such evidence, B will choose to report, when he holds an IOU, if $t_B < P_B(B,C)$. If A has such evidence, then B will choose to report, when he holds an IOU, if $t_B < s$, where s is given in (12).

Period 5 constraints are identical to the non-random version of the model, (13-14). If A holds B's IOU, then the period 4 constraint on B is identical to the non-random version, (16); otherwise the constraint is

$$\int^{P_B(B,C)} G_B(t) \leq w \tag{35}$$

When C is contemplating giving up gold in period 4, he does not know whether B will need to pass the gold on to A. Thus his constraint will compare the benefit v of deviating, with the expected penalty of either A or B making a subsequent report:

$$\begin{aligned} v + \frac{1}{2}\left(G_B(s)P_C(B,C) + (1-G_B(s))G_A(P_A(A,B))P_C(A,B)\right) \\ + \frac{1}{2}G_B(P_B(B,C))P_C(B,C) \leq 0 \end{aligned} \tag{36}$$

In period 2, in the case where B has acquired flour from A, the requirement that B prefer to give the bread to C is identical to the requirement in the non-random case, condition (17). If B was endowed with bread, then willingness to give it up is simply individual rationality, guaranteed by (1). C's willingness to enter the trade in period 2, and A's and B's willingness to enter the trade in period 1 are also guaranteed by (1).

In summary, the conditions are the same as those in the non-random case, with the exception of the substitution of condition (36) for the corresponding condition (15), and the addition of condition (35). As before, we can rewrite the problem:

Find $P_A(A,B)$ and s subject to conditions (20-23) and

$$\frac{2v}{F} \leq G_B(s) + (1-G_B(s))G_A(P_A(A,B)) + G_B(s - G_A(P_A(A,B)F) \tag{37}$$

$$\int^{s-G_A(P_A(A,B)F} G_B(t) \leq w \tag{38}$$

$$\int^{s} G_B(t)\, dt \leq G_A(P_A(A,B))F \tag{39}$$

These restrictions are more stringent than those in the certainty case. However, if we assume $v < w/2$, then (37) follows from (22). Under this assumption, we can, without loss of generality, set $s = -F + G_A(P_A(A,B))F$. Then the necessary and sufficient conditions are identical to those in the certainty case. ■

7.4 Theorem 8 and Corollary 9

Proof. We consider a mechanism in which A is not permitted to report evidence B in period 6. A will choose to report B in 4 or C in 6 if cost in that period $t_A < P_A(A,.)$. B will choose to report, if $t_B < P_B(B,C)$.

Necessary and sufficient conditions for efficiency are (27-28), (31-32) and

$$v + G_B(P_B(B,C))P_C(B,C) \leq 0 \tag{40}$$

$$\int^{P_B(B,C)} G_A(t)\, dt \leq w \tag{41}$$

$$0 \geq G_A(P_A(A,B))P_B(A,B) + (1 - G_A(P_A(A,B))w \tag{42}$$

$$u \geq \int^{P_A(A,B)} G_A(t)\, dt \tag{43}$$

Conditions for periods 1, 2, and 6 are identical to the certainty case. (40-41) are period 4 conditions; the second guarantees that B prefers to give up C's transferable debt for gold when B has not used A's input; it sufficies to guarantee B will still prefer to give up the debt to C even if he had held it illegitimately (by not paying it to A at the proper time). (42) says B prefers to pay A in period 3 with C's transferable debt (This condition anticipates the fact that if B holds the debt and is not reported by A, he will transfer the debt to C in period 4). (43) says A prefers to return B's IOU in return for C's transferable debt, rather than receive a reward for announcing it. As before, $P_i(j,k) = -F$ for $i \neq j$. Thus the problem reduces to

$$P_i(i,j) \leq 2\alpha F \text{ for } (i,j) \in \{(A,B),(A,C),(B,C)\} \tag{44}$$

$$v \leq G_A(P_A(A,C))F \tag{45}$$

$$\int^{P_A(A,C)} G_A(t)\, dt \leq u \tag{46}$$

$$v \leq G_B(P_B(B,C))F \tag{47}$$

$$\int^{P_B(B,C)} G_A(t)\, dt \leq w \tag{48}$$

$$G_A(P_A(A,B))(F + w) \geq w \tag{49}$$

$$u \geq \int^{P_A(A,B)} G_A(t)\, dt \tag{50}$$

$$v \leq G_A(P_A(A,B))F \tag{51}$$

The conditions of the theorem are necessary and sufficient for (47-50). Given $2v < w < F$ the remaining conditions are less stringent. ■

References

[1] Azariadis, C. and L. Lambertini, 2003, Endogenous debt constraints in lifecycle economies, Review of Economic Studies 70, 461-487.

[2] Bullard, J. D. and B. D. Smith, 2003, Intermediaries and payments instruments, Journal of Economic Theory 109(2), 172-197.

[3] Cavalcanti, R. and N. Wallace, 1999, A model of private banknote issue, Review of Economic Dynamics 2, 104-136.

[4] Frank, M. and V. Maksimovic, 1998, Trade credit, collateral, and adverse selection, working paper, University of British Columbia.

[5] Freeman, S., 1996, The payments system, liquidity, and rediscounting, American Economic Review 86: 1126-38.

[6] Freeman, S., 1999, Rediscounting under aggregate risk, Journal of Monetary Economics 43: 197-216.

[7] Hege, U. and L. Ambrus-Lakatos, 2000, Trade credit chains and liquidity supply, working paper, Tilburg University.

[8] Ickes, B. W., 1998, Notes on veksels, unpublished manuscript, Pennsylvania State University.

[9] Kahn, C. M., J. McAndrews, and W. Roberds, 2001, Settlement risk under gross and net settlement, Journal of Money, Credit, and Banking 47(2): 299-319.

[10] Kiyotaki, N. and J. Moore, 1997, Credit chains, unpublished manuscript, London School of Economics.

[11] Kiyotaki, N. and J. Moore, 2000, Inside money and liquidity, unpublished manuscript, London School of Economics.

[12] Kocherlakota, Narayana R, 1998, Money is memory, Journal of Economic Theory 81(2): 232-51.

[13] Kohn, M., 1999, Bills of exchange and the money market to 1600, working paper 99-04, Department of Economics, Dartmouth College.

[14] Mann, R. J., 1999, Payment Systems and Other Financial Transactions. (New York, Aspen Law and Business).

[15] Mills, D. C., 2004, Mechanism design and the role of enforcement in Freeman's model of payments, Review of Economic Dynamics 7(1): 219-236.

[16] Rogers, J. S., 1995, The Early History of Bills and Notes: A Study of the Origins of Anglo-American Commercial Law (Cambridge: Cambridge University Press).

[17] Tamaki, N., 1995, Japanese Banking: A History 1859-1959 (Cambridge: Cambridge University Press).

[18] Temzelides, T. and S. D. Williamson, 2001, Private money, settlement, and discounts, Carnegie-Rochester Conference Series on Public Policy 54: 85-108.

[19] Townsend, R.M., 1987, Economic organization with limited communication, American Economic Review 77: 954-971.

[20] van der Wee, H., 1997, The influence of banking on the rise of capitalism in northwest Europe, fourteenth to nineteenth century, in: A. Teichova, G. Kurgan-van Hentenryk, and D. Ziegler eds., Banking, Trade, and Industry (Cambridge University Press: New York) 173-190.

[21] Wicksell, K., 1935, Lectures on Political Economy, Volume II: Money (New York: Macmillan).

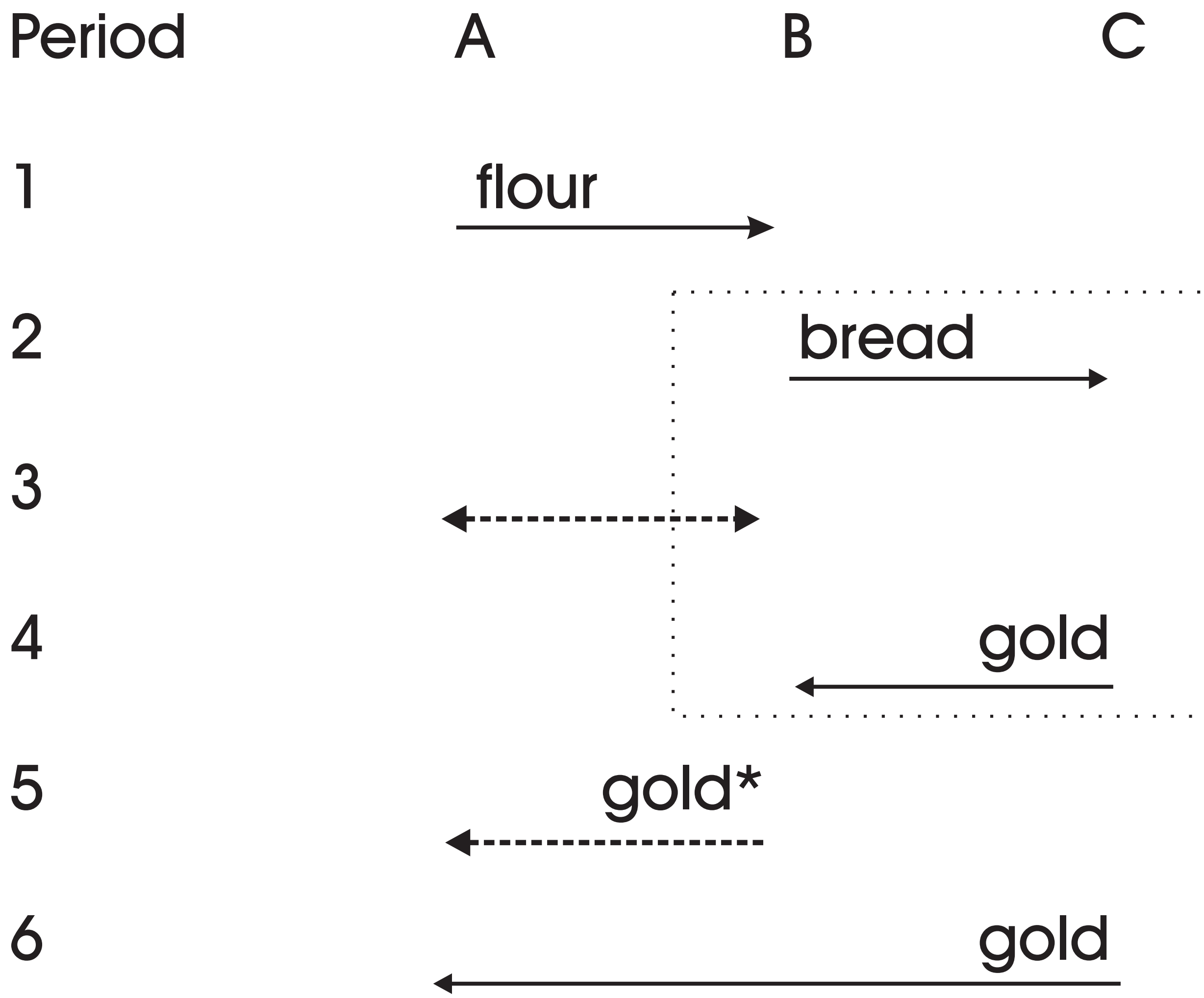
[22] Williamson, O.E., 1983, "Credible Commitments: Using Hostages to Support Exchange," American Economic Review, 73: 519-540.

[23] Williamson, S. D., 1992, Laissez faire banking and circulating media of exchange, Journal of Financial Intermediation 2: 134-167.

[24] Williamson, S. D., 1999, Private money, Journal of Money, Credit and Banking 31: 469-491.

[25] Winn, J. K., 1998, Couriers without luggage: negotiable instruments and digital signatures, South Carolina Law Review 29: 739-785.

[26] Winn, J. K., 1999, Clash of the titans: regulating the competition between established and emerging electronic payment systems, Berkeley Technology Law Journal 14: 675.



* only feasible if B receives gold from C

Figure 1: Meetings and Flow of Goods

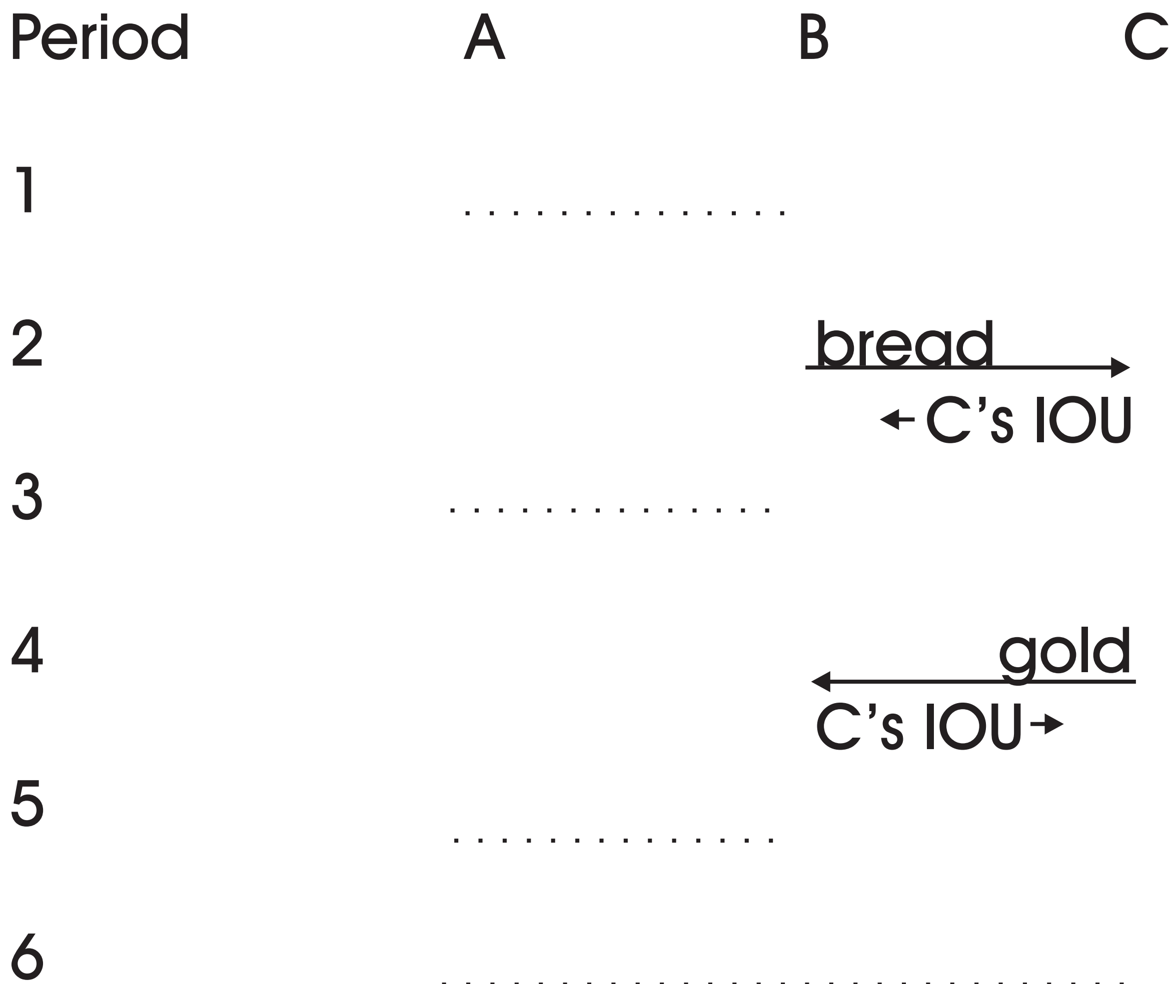


Figure 2: Trading pattern,
Two-player case

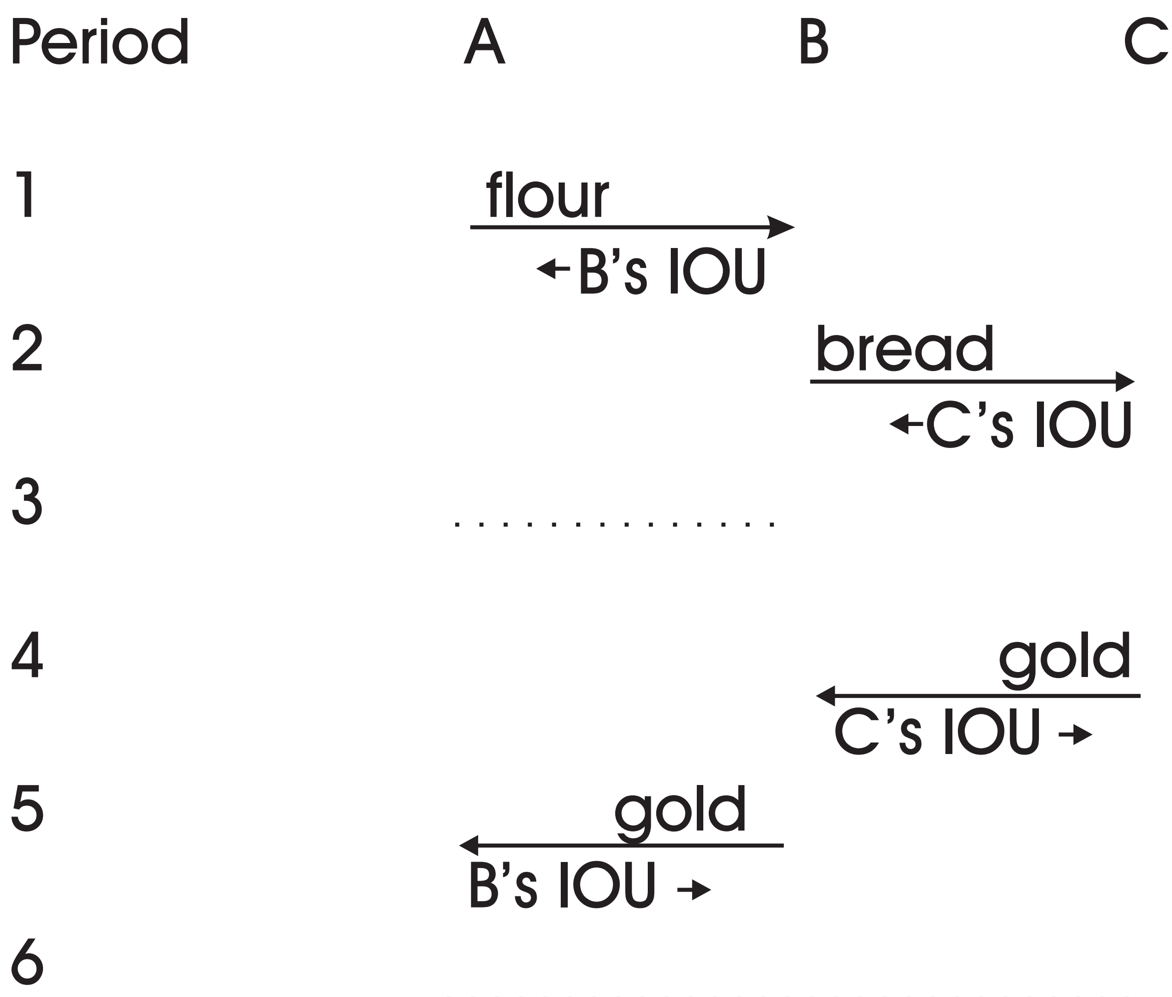


Figure 3: Trading pattern with credit chain
(Three-player case)

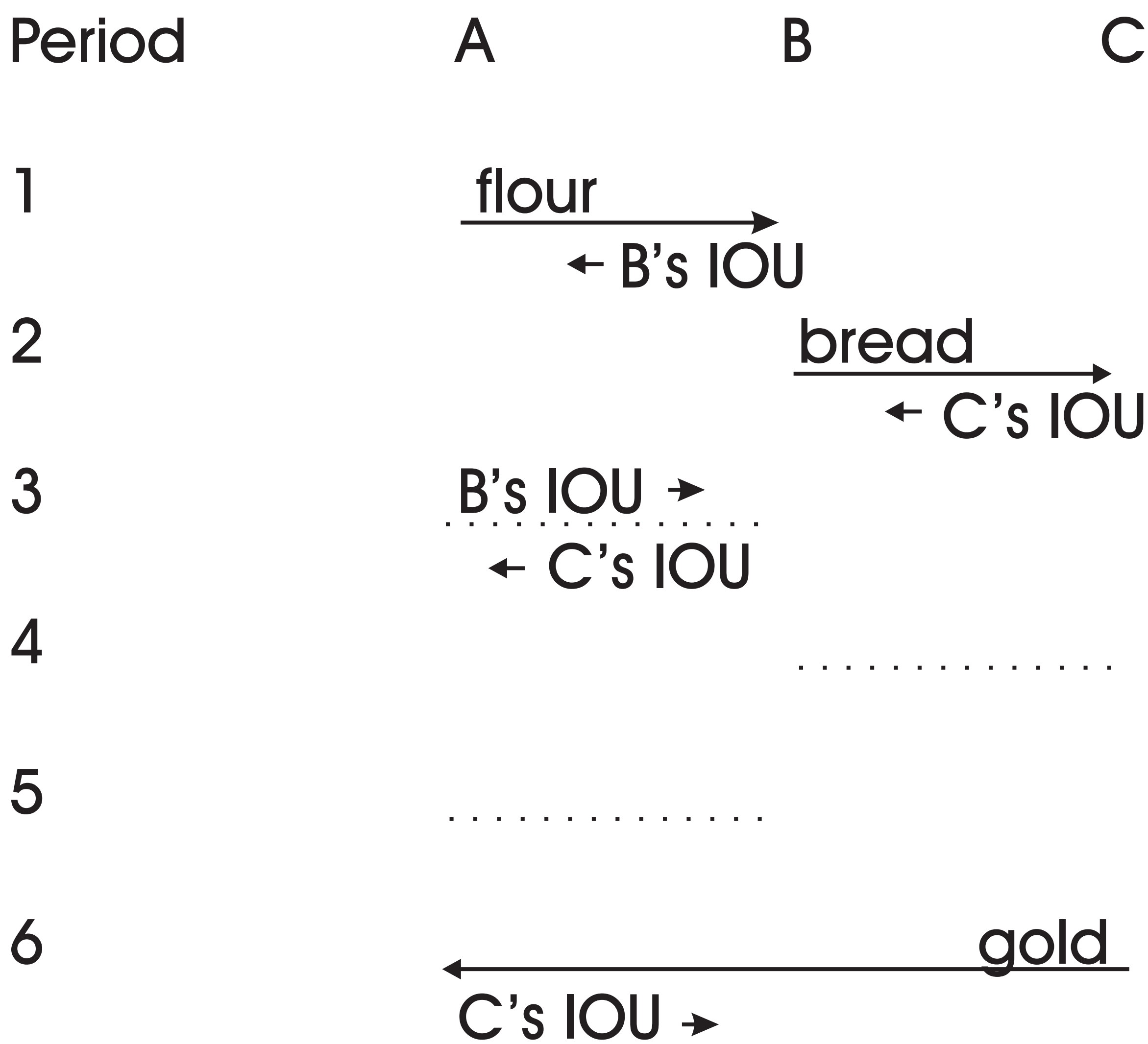
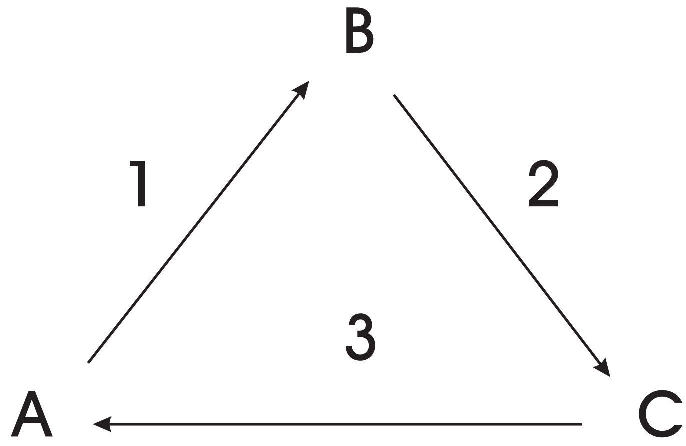


Figure 4:Trading pattern with transferable debt (Three-player case; Also mixed case when A supplies)

This Paper



Kiyotaki and Moore (2000)

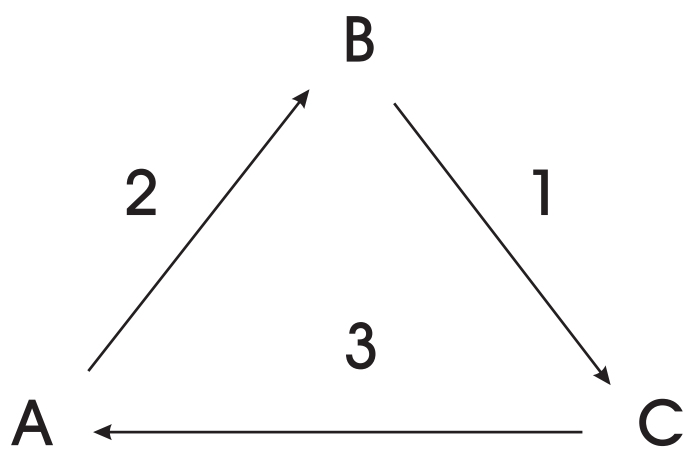


Figure 5: Flow of Goods in Intertemporal Wicksell Triangles