

AN ECONOMIC ANALYSIS OF LIQUIDITY-SAVING MECHANISMS

- Liquidity-saving mechanisms (LSMs) are queuing arrangements for payments that operate alongside traditional real-time gross settlement (RTGS) systems.
- LSMs allow banks to condition the release of queued payments on the receipt of offsetting or partially offsetting payments; as a result, banks are less inclined to delay the sending of payments.
- An analysis of LSMs finds that these mechanisms typically perform better than pure RTGS systems when it comes to settling payments early.
- RTGS systems can sometimes be preferable to LSMs, such as when many banks that send payments early in RTGS choose to queue their payments when an LSM is available.

1. INTRODUCTION

Large-value payments systems, used by banks to settle financial and commercial transactions, play a key role in the financial system. The importance of these payments systems can be illustrated by the large amounts they settle. Every year in the United States, the systems process value equal to approximately 100 times GDP.

Innovations in the design of large-value payments systems have led to many improvements in their operations. For example, over the last twenty years, many countries have adopted real-time gross settlement (RTGS) systems for their large-value payments. In an RTGS system, each payment is settled individually, on a gross basis, at the time the payment is sent. RTGS systems offer many advantages—for instance, they limit the risk exposure of payments system participants and allow for rapid final settlement of payments during the day. However, RTGS systems require large amounts of central bank balances to function smoothly.

More recent innovations have occurred in the design and implementation of various liquidity-saving mechanisms (LSMs) that are used in conjunction with RTGS systems.¹ An LSM gives participants in the payments system an additional option not offered by RTGS alone: A payment can be put into a queue and then released from the queue if some prespecified event occurs. Such mechanisms can reduce the amount of central bank balances necessary to operate the system smoothly as well as quicken the settlement of payments.

Antoine Martin is a research officer and James McAndrews a senior vice president at the Federal Reserve Bank of New York.
<antoine.martin@ny.frb.org>
<james.mcandrews@ny.frb.org>

The authors thank Enghin Atalay for excellent research assistance. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

As we describe in detail below, an LSM allows banks to send payments conditional on the receipt of payments, and it can accommodate some netting of payments.

Over the past decade, researchers have been able to simulate the performance of various LSMs. In most of these simulations, the researcher makes assumptions about the behavior of the parties in the system and measures various consequences of the assumed behavior. This approach has great potential to yield useful answers to a number of questions.

This article outlines a different approach to the study of LSMs in a payments system. It examines a theoretical model of the behavior of parties, which for simplicity we refer to as *banks*. Each bank has particular motivations and constraints; as a result, its behavior can be determined as an equilibrium

Every year in the United States, [large-value payments] systems process value equal to approximately 100 times GDP.

outcome in response to the incentives it faces. The theoretical approach has the advantage of allowing banks' reactions to alternative payments system designs to be determined within a theoretical model, rather than be assumed by the researcher. This approach also allows outcomes to be compared consistently across a number of designs.

Innovations in LSMs are numerous and, in some cases, quite complex. A 2005 report by the Bank for International Settlements, "New Developments in Large-Value Payment Systems," is an authoritative source on the many alternative systems introduced until that time. The research presented here generalizes the various types of LSMs by placing their essential characteristics into one of two categories: *balance reactive* or *receipt reactive*.

A simplified description of an LSM's operation is as follows. A bank wishes to make a payment and has a choice of when to submit it to the payments system. Upon submitting the payment, the bank has a second choice to make: It can either submit the payment to a central queue (the LSM part of the payments system, which will be called *the LSM channel*), or attempt to settle the payment at the time of submission (the RTGS part of the payments system, which will be called

the RTGS channel). If the bank submits the payment to the LSM channel, that payment will settle only when certain conditions have been met. If the bank attempts to settle via the RTGS channel and it has sufficient funds available, the payment will settle immediately.

One condition that might trigger the settlement of a payment (and is common to both types of LSMs) occurs when the request in the LSM channel is made in the presence of an offsetting payment in the queue of the bank to which the payment is to be made. If the two payments offset, then both can be released by the LSM.

A condition of another sort is determined by the type of LSM used. In a balance-reactive LSM, a bank has the choice of reserving some level of its account balances for the settlement of payments submitted via the RTGS channel. If the bank's balances exceed the predetermined reserve level, then that bank's payments that were previously placed in the LSM can be released. In a receipt-reactive LSM, a queued payment can be released against recent receipts of the bank (rather than against its accumulated balances) and at any level of balances for the bank.

The research reported here reveals two sources of potential value to implementing an LSM. First, an LSM gives a bank a new tool not available in an RTGS system: It gives the bank the option of making a payment *conditional* on the receipt of another payment. In RTGS, banks can find themselves in the positions of those two overly polite gentlemen in the old vaudeville routine, who repeatedly say to each other: "After you, Alphonse." "No, after you, Gaston."² That is, so long as central bank balances are costly, each bank would prefer to have its counterparty make a payment first.

In an RTGS system, there is no way to condition the settlement of a payment upon the future receipt of a payment; however, this course of action is possible with an LSM. The LSM essentially allows the two banks to solve the precedence problem that bedevils them in RTGS.³ The ability to condition payments on the receipt of offsetting payments provides banks with some insurance against the risk of having to borrow funds from the central bank (see, for example, Mills and Nesmith [2008]). In turn, this means that banks are more willing to submit payments to the LSM earlier than they would to the RTGS, which usually, but not always, has beneficial effects.

¹ In this article, *liquidity-saving mechanism* refers to a mechanism intended to economize on the use of central bank reserves. These reserves can typically be obtained intraday from the central bank either against collateral or for a small fee. A liquidity-saving mechanism can allow payments to be settled with fewer central bank reserves. As we discuss, under some circumstances, an LSM can indeed economize on the use of central bank balances as well as lead banks to submit payments earlier to the payments system. In general, this outcome can be defined as making the payments system more "liquid."

² The American version of this routine is based on a comic strip by Frederick Burr Opper, "Alphonse and Gaston," which was popular in the early 1900s and pokes fun at exaggerated politeness.

³ Note that an internal queue, rather than an LSM, could allow a bank to send a payment conditional on receiving another payment. However, investment in an internal queue has benefits to other banks that cannot be appropriated by the investing bank. In addition, an internal queue does not solve the precedence problem; the bank with the internal queue arranges for its payments to follow the receipt from other banks. As a result, RTGS systems that have internal queues can still suffer from excessive delay in payments.

The second source of potential value to implementing an LSM comes from the offsetting of payments within the queue, which reduces the need for central bank balances. The benefit arising from payments offsetting in the queue is well understood. An important contribution of our research is to show that even when no such offsetting occurs, the first potential benefit described above means that an LSM can improve welfare.

Several authors have examined theoretical behavior in RTGS systems. Angelini (1998, 2000) considers the behavior of banks in an RTGS system in which they face delayed costs for payments as well as costly borrowing of funds. He shows that the natural payment timing equilibria of RTGS systems (in the absence of LSMs) involve excessive delay of payments, as banks do not properly internalize the benefits to other banks from the receipt of funds. Bech and Garratt (2003) carefully specify a game-theoretic environment in which they find that RTGS systems can be characterized by multiple equilibria, some of which can involve excessive delay. Roberds (1999) compares gross and net payments systems with systems offering an LSM. Examining the incentives banks have to engage in more

The research presented here generalizes the various types of [liquidity-saving mechanisms] by placing their essential characteristics into one of two categories: balance reactive or receipt reactive.

risk-taking behavior in the different systems, Roberds finds that, under certain circumstances, the risk profiles of LSMs and net systems are identical. Mills and Nesmith (2008) also study the impact of incentives on banks' payment patterns. McAndrews and Trundle (2001) and the Bank for International Settlements (2005) provide extensive descriptive material on LSMs.

Willison (2005) examines the behavior of banks in an LSM, and this work is most similar to ours. He models agents as having an ordering of payment priority, which is similar in spirit to our assumption that some banks' payments are time sensitive. Willison models the extension of credit from the central bank as an ex ante amount to be borrowed by banks, while in our study the credit is tapped ex post, depending on a bank's per-period balance. Our model extends Willison's analysis in two dimensions that prove to be important: We consider a wider array of LSMs and, crucially, we allow for liquidity shocks, which we define as shocks to the level of a bank's balances on account.

Our study proceeds as follows. In the next section, we describe the environment in place when banks decide on their payment submission strategies. In subsequent sections, we consider various scenarios, including banks' behavior in an RTGS system, a balance-reactive LSM, and a receipt-reactive LSM.

2. THE ENVIRONMENT

This section describes the economic environment in which banks operate. It specifies the economic agents, which we refer to as *banks*; the banks' objectives; and other factors that influence banks' decisions. We study a simple payments system. While some features of our model are unrealistic because of their simplicity, we believe that the model captures essential economic frictions that affect banks' behavior in the payments system. The simplicity of the model allows us to obtain explicit results and to provide transparent intuition for these results.⁴

In our model, the essential features of an LSM are as follows. Payment orders are put in a queue, and the release of a payment order from the queue will occur when some conditions have been satisfied. Here, the conditions are that a bank has sufficient funds and that an offsetting payment has been received. While actual LSMs can be considerably more complicated, they all share those basic, essential features. By capturing the essential features shared by LSMs, our model can describe the basic economic mechanisms associated with these systems in a tractable way.

To incorporate additional features of queues, we would require a more complicated model that may be difficult to solve explicitly. Those additional features could include: 1) limits on the amount of central bank balances that can be committed to a particular bank or set of banks, 2) a time before or after which a payment should be sent, and 3) different payment priorities that can change the ordering of payments submitted to the queue away from first in, first out. Also, queues may or may not be transparent. If a queue is transparent, banks can see pending payments in their favor.⁵ In addition to the tractability issue, it is not clear that adding these features would modify our key insights into the potential benefits of the LSM's essential feature: to allow conditionality in the settlement of some payments.

⁴ Our analysis summarizes our more technical paper (Martin and McAndrews 2008).

⁵ In our model, queues are opaque—that is, banks cannot see particular payments in other banks' queues. However, banks are forward looking and, in equilibrium, they will expect banks of different types to have (or have not) submitted some payments to the LSM, and these expectations will be fulfilled.

Each day, the payments system operates during two periods: the morning and the afternoon. A large number of banks are involved, and each bank must send payments to other banks and will receive payments from other banks. For simplicity, all payments are of the same size.

We assume that banks are risk neutral. Our definition of risk neutrality is that a bank would be willing to pay up to 50 cents for the opportunity to participate in a lottery that promises one dollar with probability 1/2 and nothing with probability 1/2. We believe that our results would extend to more complicated economies in which there are many periods and payments of different sizes. Banks have rational expectations about the probability of receiving a morning payment. In other words, banks are able to calculate the correct probability of receiving a payment in the morning.

Banks must choose whether to send their payments to another bank in the morning or in the afternoon. Three factors influence this decision. First, banks face a cost if they must borrow funds at the central bank. Second, some banks must

If a time-sensitive payment is not settled until the afternoon, the sending bank incurs a delay cost. This cost could arise because delay creates a bad reputation for the bank in the eyes of its customers or counterparties.

make time-sensitive payments. Third, banks may receive a positive liquidity shock, a negative liquidity shock, or no shock. Each factor is explained in more detail below.

Banks face a cost if they must borrow funds from the central bank. All banks start the day with zero reserves at the central bank and, because of the symmetry assumed by the model, end the day with zero reserves as well. Banks are not allowed to hold negative reserves at the end of the morning period, but they can borrow reserves from the central bank at a cost.⁶ For example, a bank that sends a payment but does not receive a payment in the morning period would have to borrow reserves. The cost of borrowing can represent either an explicit fee imposed by the central bank, as in the United States, or the implicit cost of collateral, as in many countries where central bank balances are

⁶ We assume that excess reserves have zero return. In our model, banks cannot trade with one another the funds needed to make payments. This assumption is a common one in the literature and reflects the empirical observation that markets for intraday funds are either nonexistent or operate only in extraordinary circumstances.

available for free but only against collateral. Evidence provided by the Bank for International Settlements (2005) suggests that the cost of central bank balances can influence banks' payment behavior.

Some banks must make time-sensitive payments. If a time-sensitive payment is not settled until the afternoon, the sending bank incurs a delay cost. This cost could arise because delay creates a bad reputation for the bank in the eyes of its customers or counterparties. The behavior of banks suggests that they perceive some payments as being more time sensitive than others. For example, payments made to close a real estate transaction may be more time sensitive than payments for previous deliveries of supplies to a manufacturer.⁷ However, delaying non-time-critical payments until the afternoon is costless.

Banks may receive a positive liquidity shock, a negative liquidity shock, or no liquidity shock. The liquidity shock is modeled by assuming that each bank must send a second payment to (and will receive a payment from) a settlement institution. For the purpose of this article, U.S. settlement institutions are represented jointly by CLS Bank, CHIPS, and DTC.⁸ The settlement institution is intended to capture other payment or settlement systems whose pay-ins and payouts are made using the central bank's payments system. This rearrangement of balances, which we refer to as a *shock*, introduces another source of heterogeneity between banks and leads to a more diverse set of strategic interactions.

In the model, banks find out when they must send a payment to the settlement institutions and when they will receive an offsetting payment from the settlement institutions *before* they decide whether to send their payment to other banks. We assume that payments to settlement systems cannot be delayed. We say that a bank that receives a payment from the settlement institutions in the morning and does not have to make an offsetting payment until the afternoon has received a positive liquidity shock, as its account balances are boosted throughout the day prior to its late offsetting payment. Conversely, a bank that receives a payment from the settlement institutions in the afternoon but must make a payment in the

⁷ Armantier, Arnold, and McAndrews (2008) find that Fedwire payments tend to settle earlier in the day on days with higher values of customer payments. This behavior could reflect higher costs of delay for customer payments relative to interbank payments.

⁸ CLS Bank is a payment-versus-payment settlement system that settles foreign exchange transactions in fifteen currencies. CLS Bank is operated by CLS Bank International, a bank-owned Edge Act corporation incorporated in the United States. CHIPS (the Clearing House Interbank Payments System) is a private, large-value U.S. dollar payments system owned and operated by the Clearing House Payments Company. DTC (the Depository Trust Company) is a securities settlement system that settles the majority of U.S. corporate securities and commercial paper. DTC is a wholly owned subsidiary of Depository Trust and Clearing Corporation.

morning receives a negative shock, as its balances are depleted early. When the payment from the settlement institutions is received in the same period in which the bank's outgoing payment to the settlement institutions must be sent, the bank experiences no liquidity shock, as the offsetting payments have a neutral effect on its balances.

We assume that payments sent among banks are larger than payments sent to and received from settlement institutions. This implies that while banks may face large liquidity shocks, these shocks are relatively small compared with the banks'

Each bank can have one of six possible profiles: It may or may not have to make a time-sensitive payment, and it may experience a positive, negative, or no liquidity shock.

expected payment activity. In addition, payments made through settlement institutions occur either at the very beginning or very end of the business day. Thus, a bank that begins its business day with a negative liquidity shock, then receives a payment from another bank in the morning, and then delays sending a payment to another bank will have a positive balance at the end of the morning period. In contrast, a bank that receives a positive liquidity shock at the start of the business day, then proceeds to send a payment to another bank in the morning, and then does not receive a payment from another bank during the morning period will end the morning with a negative balance. The latter bank must borrow from the central bank, incurring a borrowing cost in the process.

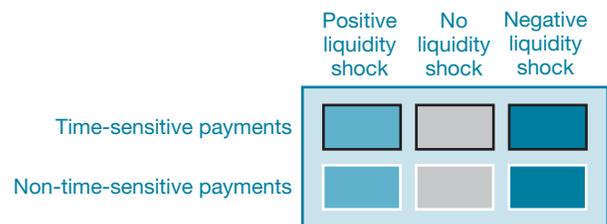
With all those factors explained, we can now trace out the intuition of the model. By itself, the cost of borrowing gives banks an incentive to send payments at the same time as other banks (either in the morning or in the afternoon). For example, absent liquidity shocks and delay costs, banks would be willing to send their payments in the morning (or afternoon) if they knew that all other banks would send theirs in the morning (or afternoon). That way, they would never have to borrow from the central bank. However, the cost of delay provides an incentive for banks to make payments early. For example, if the cost of delay is sufficiently high, a bank will prefer to send a time-critical payment early even if all other banks delay their payments. This will happen if the cost of delay is greater than the cost of borrowing. A cost of borrowing and a high cost of delay will therefore result in banks preferring to send payments

all at the same time in the morning. Introducing the liquidity shock, together with the borrowing cost, provides an incentive for banks to delay payments. For example, a bank that experiences a negative liquidity shock and that must make a non-time-critical payment prefers to send it in the afternoon even if all other banks pay early. Indeed, in that case, the bank incurs no cost if it delays its payment, but it receives the benefit of not having to borrow from the central bank.

In summary, each bank can have one of six possible profiles: It may or may not have to make a time-sensitive payment, and may experience a positive, negative, or no liquidity shock. Exhibit 1 displays the six profiles in the form of a color key that will describe the equilibria of our economy. We assign the color light blue to banks that receive a positive liquidity shock, gray to banks that receive no liquidity shock, and dark blue to banks that receive a negative liquidity shock. A black border indicates banks that must make time-sensitive payments, while a white border indicates banks that make non-time-sensitive payments.

The remainder of this article examines different settlement systems for the payments system described above. We consider real-time gross settlement alone and an RTGS system supplemented by two types of liquidity-saving mechanisms. In each case, we describe the timing of payments predicted by the model in equilibrium. In equilibrium, every bank chooses a submission time for its direct payments that minimizes its delay and borrowing costs given the payment submission strategies of all other banks in the payments system. We also compare the desirability of different settlement systems according to our model.

EXHIBIT 1
Banks Using Large-Value Payments Systems



Note: The key applies to Charts 1-10. It describes six types of banks: those sending time-sensitive payments that experience a positive liquidity shock, no liquidity shock, or a negative liquidity shock; and those sending non-time-sensitive payments that experience a positive liquidity shock, no liquidity shock, or a negative liquidity shock.

3. REAL-TIME GROSS SETTLEMENT

This section studies our model economy in terms of a real-time gross settlement system. With RTGS, payments are settled on an individual basis at the time they are sent. We assume that banks have sufficient borrowing capacity at the central bank to make payments, even if they do not receive a prior payment and/or have experienced a negative liquidity shock. Banks have the choice of sending payments in the morning or delaying them until the afternoon. They compare the expected cost of each option and choose the least expensive timing strategy. As shown in Exhibit 2, there is a pattern to how events unfold during the day. First, banks learn if they must send a time-critical payment and if they have a positive, negative, or no liquidity shock. Next, early payments to settlement institutions are made. Afterward, banks decide whether to send their direct payments in the morning period or to delay sending them until the afternoon period. At the end of the morning period, banks incur a delay cost if they have delayed a time-sensitive payment or a borrowing cost if they must borrow from the central bank. In the afternoon, delayed payments to other banks are made, as are late payments to settlement institutions.

Liquidity shocks force some banks to start the day with negative balances. If those banks do not have time-sensitive payments to make, then they can be counted on to delay their direct payments, since otherwise even the receipt of an offsetting payment would mean that they must borrow. So banks with negative liquidity shocks and non-time-sensitive payments will surely delay their payments until the afternoon period. Following domino-type logic, banks—even those that enjoyed a positive liquidity shock—will delay all other non-time-sensitive payments until the afternoon, as they cannot

count on offsetting payments during the morning period. In sum, in a world with liquidity shocks, we expect that all non-time-sensitive payments will always be delayed.

We present the equilibria associated with RTGS in Charts 1-4. The dark gray bars indicate payments sent to settlement institutions. These payments are made either early in the morning or late in the afternoon. Payments to other banks are indicated in colors that correspond to a specific type of sender as described in Exhibit 1. Payments sent by banks that have experienced a negative liquidity shock are indicated in dark blue, those sent by banks that have received no liquidity shock are in gray, and those sent by banks that have experienced a positive liquidity shock are in light blue. A dark border corresponds to time-sensitive payments while a white border corresponds to non-time-sensitive payments. Given this color scheme, the behavior of each of the six types of banks can be seen within the bar, indicating whether a bank of a certain type sent its payments in the morning or the afternoon.

Charts 1-4 show how the pattern of payments changes as the cost of delay decreases. In Chart 1, the cost of delay is high, so all time-sensitive payments are sent early. If the cost of delay decreases a little (Chart 2), banks that have experienced a negative liquidity shock will choose to delay their time-sensitive payments. These banks have the highest expected cost of borrowing from the central bank, so if the cost of delay is not too high, banks will aim to reduce their need to borrow from the central bank by delaying payments. If the cost of delay decreases still further (Chart 3), only banks that have experienced a positive liquidity shock will send their time-critical payments early. Finally, if the cost of delay is sufficiently low (Chart 4), all banks will delay payments regardless of their liquidity profiles.

EXHIBIT 2

Real-Time Gross Settlement Timeline

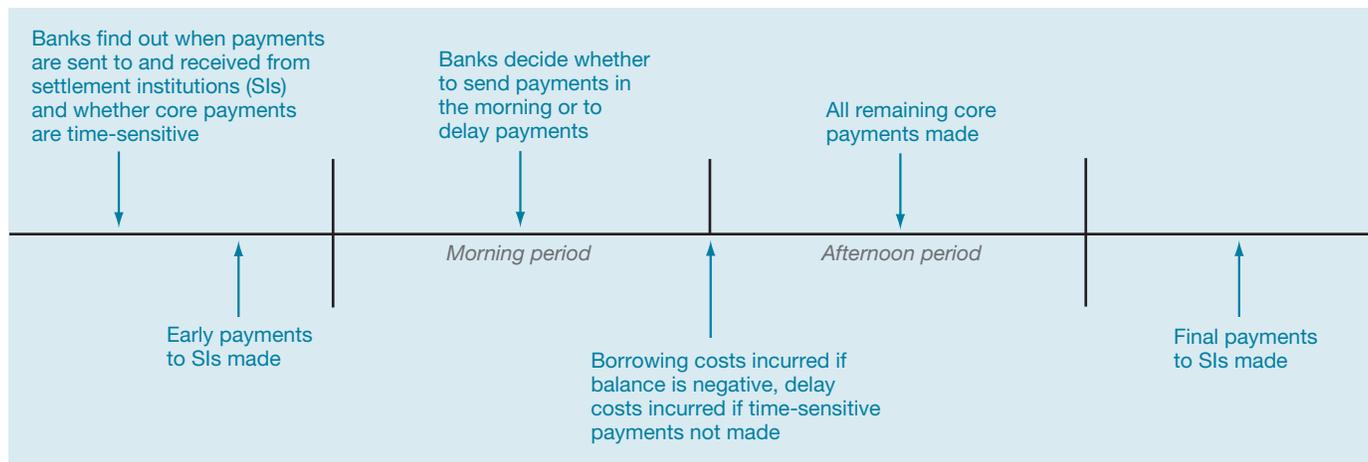
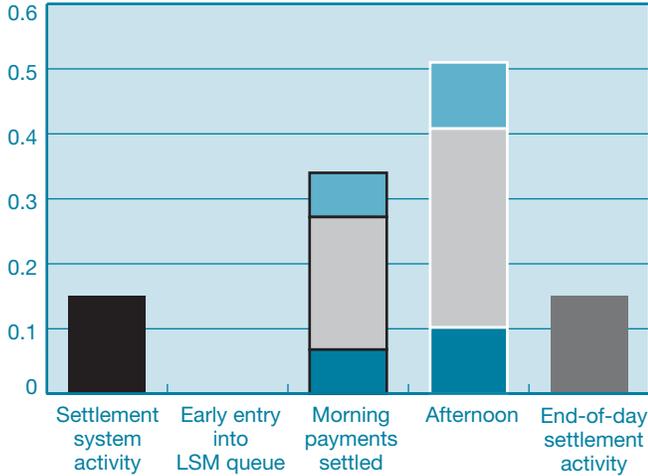


CHART 1

Real-Time Gross Settlement with High Costs of Delay

Value of payments (normalized to total 1)



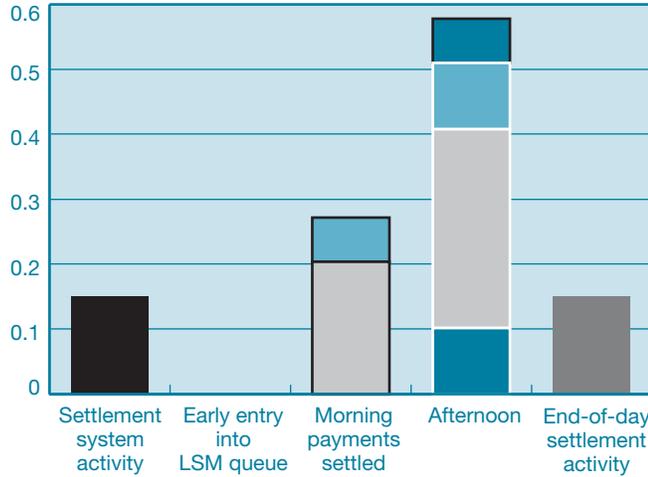
Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

CHART 2

Real-Time Gross Settlement with Medium Costs of Delay

Value of payments (normalized to total 1)



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

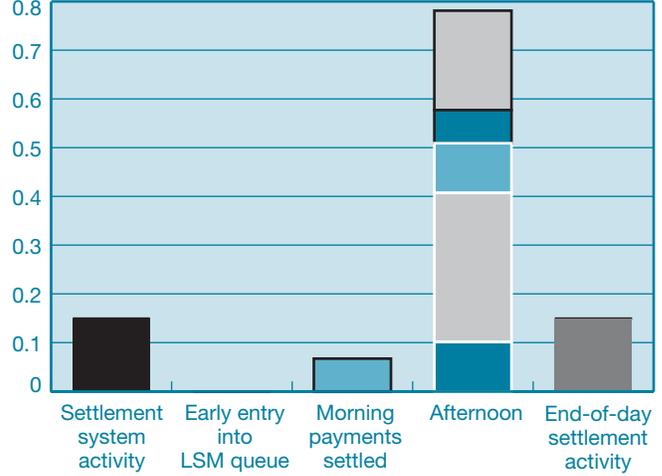
In general, multiple equilibria can exist. Each equilibrium described in Charts 1-4 may arise for the same set of parameters. There can be multiple equilibria because a bank's incentives to pay early depend on the behavior of other banks.⁹

⁹ By multiple equilibria, we mean that, for the same set of parameters of the model, different sets of beliefs are consistent with an equilibrium.

CHART 3

Real-Time Gross Settlement with Lower Costs of Delay

Value of payments (normalized to total 1)



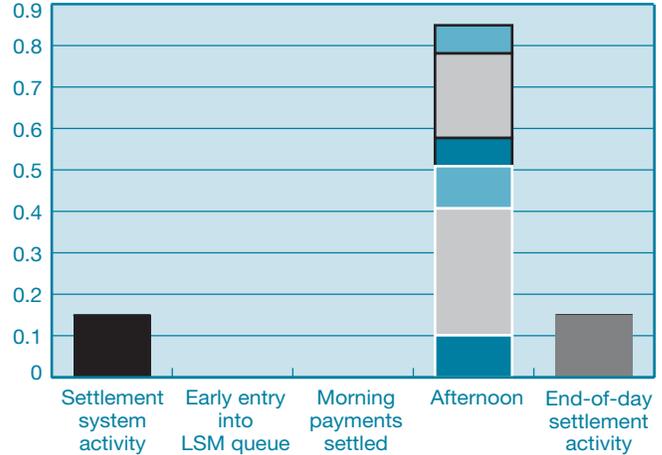
Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

CHART 4

Real-Time Gross Settlement with Lowest Costs of Delay

Value of payments (normalized to total 1)



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

If all time-sensitive payments are sent early, then the probability of receiving a payment in the morning is high, implying that the probability of having to borrow from the central bank is low. If the expected cost of borrowing from the central bank is low, banks have a strong incentive to make their time-sensitive payments early to avoid the cost of delay.

Conversely, if all time-sensitive payments are delayed, then the probability of receiving a payment in the morning is low (in this case, zero). It follows that banks will have to borrow from the central bank if they make payments early. With a high expected cost of borrowing, agents have only weak incentives to send payments early, even if this would allow them to avoid the delay cost.

We can compare the welfare associated with different equilibria. Our measure of welfare is the expected utility of the average of the six possible types of banks. To measure this average payoff, we consider what would happen if all banks

Banks have the choice of sending payments in the morning or delaying them until the afternoon. They compare the expected cost of each option and choose the least expensive timing strategy.

could meet before the beginning of the day. Before determining their respective liquidity shocks and whether they have to make a time-sensitive payment, all banks are identical and thus have the same preferences for payment patterns.

When several equilibria coexist, welfare is highest in the equilibrium in which the greatest number of payments is sent in the early period.¹⁰ Prior to knowing its particular situation—that is, whether it has a positive or negative balance in the morning and whether it has a time-sensitive payment—the average bank would prefer to make its payments at the same time as all other banks. This simultaneous payment pattern reduces banks' expected borrowing needs. Given the level of coordination, the average bank would prefer to make simultaneous payments in the morning, thereby reducing delay costs. For all equilibria, our measure of welfare decreases with the cost of delay and with the cost of borrowing.

We can illustrate the role liquidity shocks play in our analysis by looking at how equilibrium outcomes change when these shocks are suppressed. This would correspond to the case where all banks make their pay-ins to the settlement institutions and receive their payouts from the settlement institutions early. Absent liquidity shocks, all banks start the morning period with zero balances and the only difference among them is whether or not their payments are time sensitive. Three types of equilibria are possible: 1) all payments

are sent early, 2) only time-sensitive payments are sent early, or 3) all payments are delayed.

Consider the equilibrium in which all payments are sent in the morning period. This is an equilibrium, regardless of the magnitude of borrowing and delay costs, because if all other banks are sending their payments in the morning, an individual bank is at least as well off sending its payments in the morning as delaying them until the afternoon. If the bank has a time-sensitive payment, the bank is strictly better off sending it in the morning; if the bank does not, it is equally well off sending its payments in the morning to coincide with its receipts. This type of equilibrium does not exist when some banks have experienced a negative liquidity shock.

The costs of delay and borrowing matter for the other two equilibria. If the cost of delaying time-sensitive payments is low, if the borrowing cost is high, and if banks without a time-sensitive payment plan to pay in the afternoon, then banks that must make a time-sensitive payment would prefer to delay it. Indeed, banks want to avoid high borrowing costs, even if they must incur relatively low delay costs. Likewise, if the cost of delay is high relative to the cost of borrowing, then banks with time-sensitive payments would prefer to make them early, even if banks without time-sensitive payments choose to delay making theirs.

4. BALANCE-REACTIVE LIQUIDITY-SAVING MECHANISM

This section studies our model economy under a particular design of a liquidity-saving mechanism. An LSM can be thought of as a queue, into which banks enter their payments as an alternative to sending or delaying payments outright. A queued payment is released in the morning if the bank that queued the payment receives an offsetting payment or if a group of queued payments offset multilaterally. Otherwise, the payment is released from the queue in the afternoon.

The probability of a payment being released from the queue depends on the underlying pattern of payments. Exhibits 3 and 4 illustrate two cases. In the exhibits, the X_s denote banks and the arrows denote a payment that must be made from one bank to another. In Exhibit 3, all payments form a unique cycle that links all banks. Here, payments in the queue cannot offset multilaterally unless all payments are queued. In Exhibit 4, payments form cycles of length 2 that link banks in pairs. This pattern of payments maximizes the probability that payments in the queue offset bilaterally. The case of a unique cycle, illustrated in Exhibit 3, is particularly interesting because it allows us to disentangle two different roles played by LSMs. One is that the LSM allows agents to condition the release of

¹⁰ Note that the pattern of payments such that all payments, including time-sensitive ones, are made early may not provide maximum welfare. This does not contradict the result stated above, as this pattern of payments is not an equilibrium.

EXHIBIT 3
A Unique Cycle

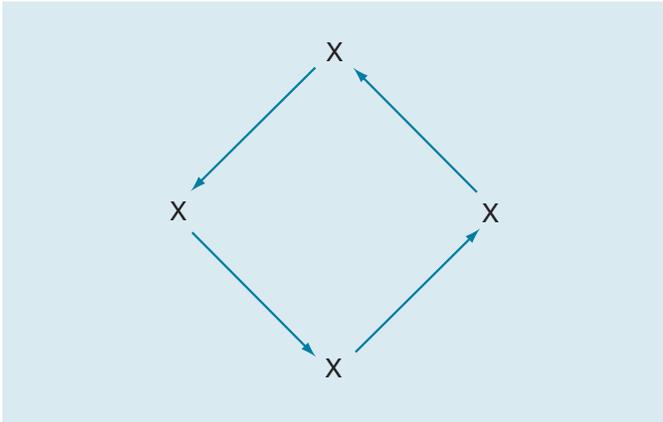
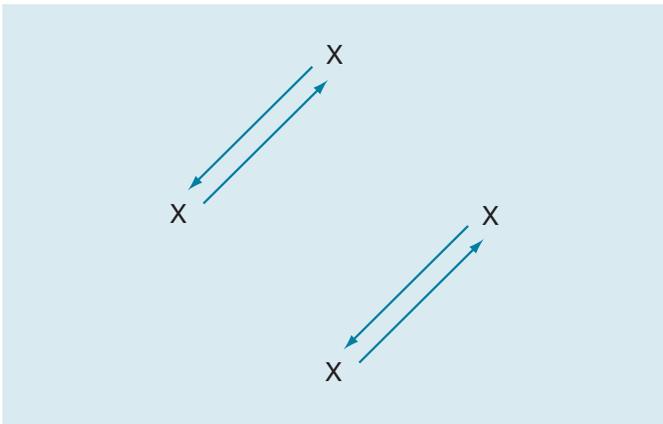


EXHIBIT 4
Cycles of Length 2

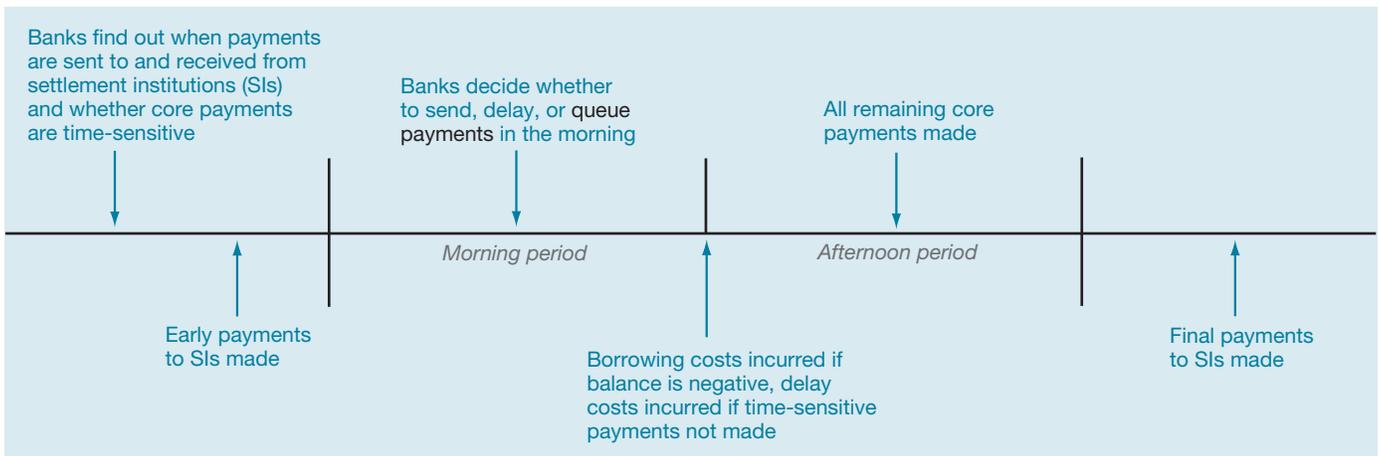


a payment on the receipt of an offsetting payment. A different, though complementary, role is that an LSM can permit bilateral or multilateral netting. In the case of a unique cycle, however, no netting occurs unless all payments are in the queue. Here we can study, in isolation, the role of the LSM in allowing banks to make payments conditional on the receipt of offsetting payments.

Exhibit 5 shows the timing of payments in a system in which banks are able to submit payments to the LSM. The difference between this and pure RTGS is that now banks have the option of queuing their payments. In other words, a bank can condition its sending of a queued payment on the receipt of an offsetting payment, in addition to sending or delaying the payments outright. The decision to queue a payment is made after the bank's liquidity shock is known, allowing the bank to take into account the shock's effect on the level of the balance before deciding whether to queue a payment. Thus, we call this a *balance-reactive* LSM. Later, we will consider an LSM in which the release of a payment from the queue does not depend on a liquidity shock.

Charts 5-8 illustrate the equilibria arising with a balance-reactive LSM. The visual coding in these charts is the same as in Charts 1-4. That is, payments from banks that experience a positive liquidity shock are indicated in light blue, those from banks that experience no liquidity shock are in gray, and those from banks that experience a negative liquidity shock are in dark blue. A black border indicates time-sensitive payments while a white border indicates payments that are not time sensitive. In addition, dark shading indicates payments entered in the queue, which we emphasize by adding a queue-submission stage. Payments released from the queue have a light shading, while payments sent outright or delayed have no

EXHIBIT 5
Balance-Reactive Liquidity-Saving Mechanism Timeline



shading. Some payments from the queue will be released in the afternoon unless all payments are queued.

In Chart 5, all payments are queued. This happens if both the cost of delay and the liquidity shocks are not too high. In this case, all payments are released in the morning and no delay cost is incurred. Banks with a negative liquidity shock, however, must borrow from the central bank.

If the cost of delay is higher and the liquidity shocks remain moderate, the equilibrium depicted in Chart 6 can occur. In this equilibrium, banks that experience positive liquidity shocks choose to make time-sensitive payments outright during the early period. Banks that experience negative liquidity shocks delay non-time-sensitive payments. All other payments—time-sensitive payments from banks that experience negative or zero liquidity shocks and non-time-sensitive payments from banks that experience positive or zero liquidity shocks—are queued. Some payments are released from the queue in the morning and some are released in the afternoon.

The intuition is as follows. Because the cost of delay is sufficiently high, banks that have experienced a positive liquidity shock prefer to insure themselves against the risk of having to suffer the cost of delay. For this reason, they send payments outright. Banks that receive a negative liquidity shock prefer to delay non-time-sensitive payments. By delaying their own outgoing payments, these banks avoid the need to

borrow from the central bank if they receive a payment from another bank in the morning. Note that for parameters in which the equilibria in Charts 5 and 6 coexist, the latter equilibrium is not robust.

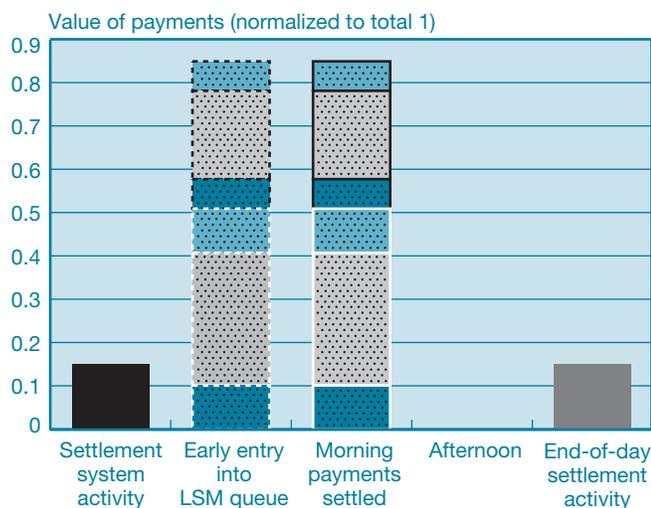
If the cost of delay is higher still, the equilibrium presented in Chart 7 can occur. In this equilibrium, all banks that must make time-sensitive payments choose to send them outright. Banks that experience a negative liquidity shock delay non-time-sensitive payments, while those that experience a positive or no liquidity shock queue non-time-sensitive payments.

If the cost of delay is neither very high nor very low and the liquidity shocks are large, then the equilibrium in Chart 8 will occur. Because the liquidity shock is so large, banks that receive a negative liquidity shock at the beginning of the day delay both time-sensitive and non-time-sensitive payments. Only banks that receive a positive liquidity shock at the beginning of the day choose to send time-sensitive payments early; all other payments are queued.

The same types of equilibria exist in a unique cycle and in cycles of length 2. However, the parameters under which these equilibria exist in each case can be different. With cycles of length 2, there can be multiple equilibria, as is true with pure RTGS. In contrast, with a unique cycle, only one equilibrium exists for any configuration of parameters.

To understand this difference, note that an LSM allows banks to condition the release of a payment on other banks'

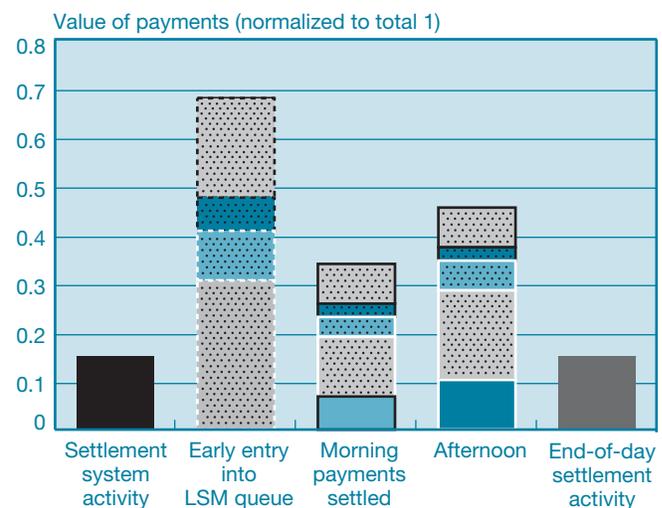
CHART 5
Balance-Reactive Liquidity-Saving Mechanism (LSM) with Low Costs of Delay



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

CHART 6
Balance-Reactive Liquidity-Saving Mechanism (LSM) with Medium Costs of Delay



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

actions, which reduces strategic interaction between banks. In RTGS, by contrast, the level of strategic interaction is high: Whether or not a bank chooses to send a payment during the early period depends in large part on that bank's beliefs about the strategy other banks have adopted for the timing of their payments. With the LSM, the bank need not concern itself with such complicated reasoning because it can simply submit a payment to settle automatically if an offsetting payment is received, decoupling the bank's decision to submit the payment to the LSM from the bank's beliefs about the plans of other banks.

Whenever multilateral netting is expected to occur in the queue, a new set of strategic interactions emerges. A bank's incentives to submit payments to the queue depend on the bank's belief about what other banks will do. If many other banks submit their payments to the queue, then more netting will occur and the incentive to queue will be high. In contrast, if only a few other banks submit their payments to the queue, less netting will occur and the incentive to queue will be low. As a result, the possibility of multilateral netting of queued payments reintroduces strategic interaction into banks' submission behavior, leading to multiple equilibria. In contrast, since there is no netting in a unique cycle, multiple equilibria do not occur.

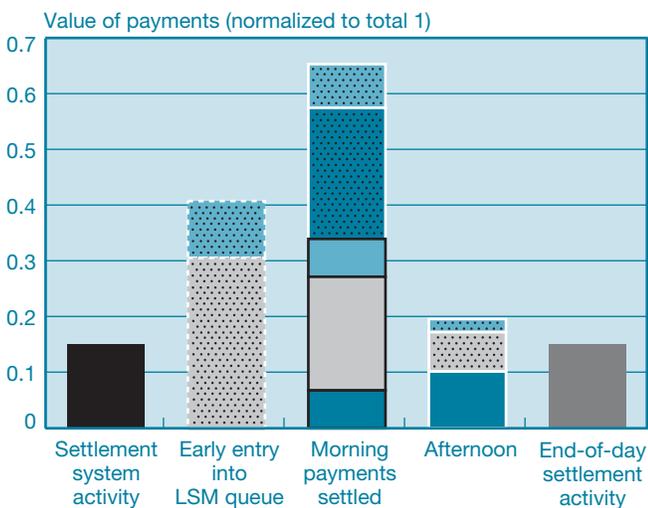
We can compare the welfare of banks if settlement occurs according to RTGS or to a balance-reactive LSM. Welfare is

higher with the balance-reactive LSM if the liquidity shocks are small, if the cost of delay is either sufficiently high or sufficiently low, and if the probability of a liquidity shock occurring is large compared with the fraction of time-sensitive payments. In one specific case, a balance-reactive LSM can lead to a loss of some beneficial coordination and result in lower welfare.

When liquidity shocks are small, banks incur a large borrowing cost only if they send a payment early and do not receive an offsetting payment. An LSM increases welfare because it allows banks to insure themselves against this risk because the bank's payment is released from the queue only when an offsetting payment is received. Also, because banks can queue their payments rather than delay them outright, more payments are released in the morning and fewer banks incur a delay cost.

If the cost of delay is sufficiently high, banks send all time-sensitive payments early, regardless of the settlement mechanism. However, banks that do not receive a negative liquidity shock queue their payments when an LSM is available, rather than delay them as they would in RTGS. Hence, more payments are released early and the expected cost of borrowing is reduced for banks that send payments outright. If the cost of delay is sufficiently low, all banks will find it beneficial to queue payments when an LSM is available so that all payments are released early.

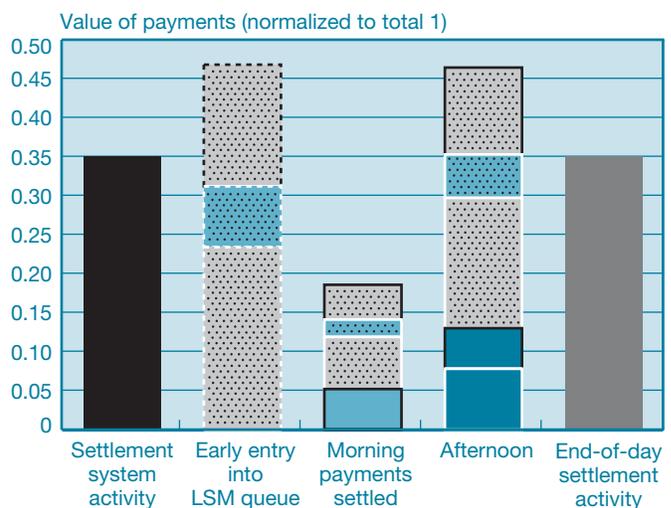
CHART 7
Balance-Reactive Liquidity-Saving Mechanism (LSM) with High Costs of Delay



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

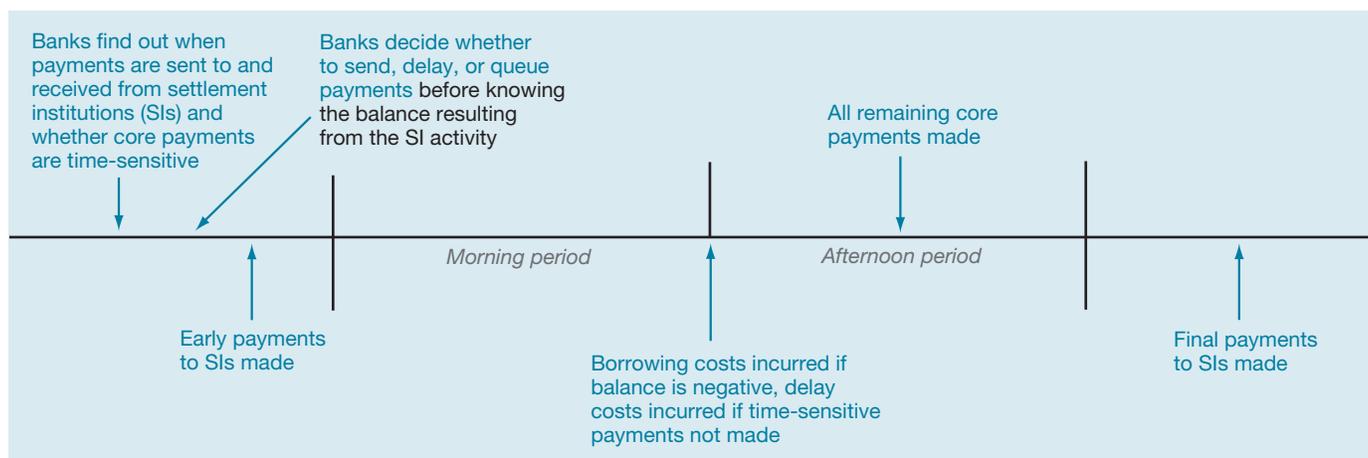
CHART 8
Balance-Reactive Liquidity-Saving Mechanism (LSM) with Large Liquidity Shock



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

Receipt-Reactive Liquidity-Saving Mechanism Timeline



If the fraction of time-sensitive payments is relatively small, then the benefit of queuing non-time-sensitive payments under a balance-reactive LSM—rather than delaying them, which would occur in RTGS—is large. Here too a balance-reactive LSM provides higher welfare than would RTGS.

If none of these conditions is satisfied, an instance may arise in which RTGS provides higher welfare than does a balance-reactive LSM. There exists a set of parameters for which the equilibrium described in Chart 2 is the unique RTGS equilibrium. In this equilibrium, banks that do not receive a

We can illustrate the role of liquidity shocks by considering what happens when they are suppressed. Absent liquidity shocks, all payments are released early.

liquidity shock send a time-critical payment early. For this set of parameters, the best LSM equilibrium is described in Chart 8. In this equilibrium, banks that do not receive a liquidity shock prefer to queue a time-critical payment, rather than send it early, because it provides them with some insurance against the risk of having to borrow from the central bank. However, this action lowers the number of payments that are released in the morning and reduces welfare. This example shows that RTGS can create beneficial coordination among banks to send some

payments early. In this case, the presence of a queue unravels that coordination.

Again, we can illustrate the role of liquidity shocks by considering what happens when they are suppressed. Absent liquidity shocks, all payments are released early. This could happen if all the payments are queued, if they are all sent in the morning outright, or if some are queued and the others are sent in the morning outright.

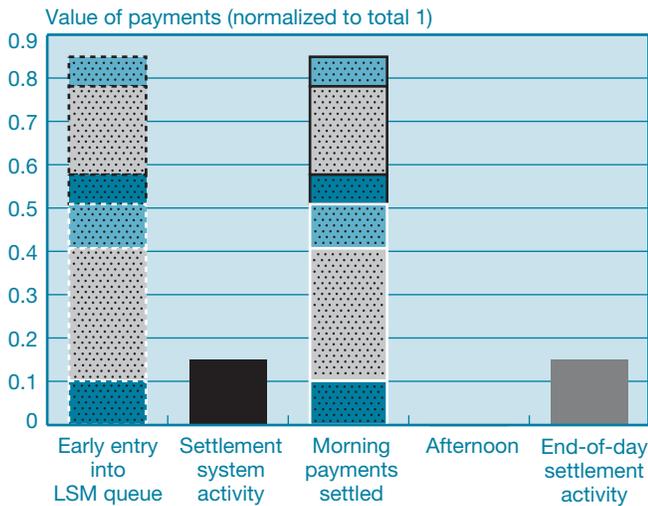
5. RECEIPT-REACTIVE LIQUIDITY-SAVING MECHANISM

We now consider a different LSM design. Previously, we assumed that banks could make their decision to queue a payment conditional on their liquidity shock or, equivalently, on their balance. Here we assume that banks do not know their liquidity shock when they decide either to queue payments or to pay in the morning period. We also assume that the decision to queue is irrevocable. Since banks can condition their behavior only on the receipt of other payments, but not on their balance, we call this a *receipt-reactive* LSM.¹¹

This case illustrates another possible feature of an LSM that can affect bank behavior. In the receipt-reactive LSM, banks are given a tool that enables them to commit to making a payment at a particular time. This ability is valuable to banks. The new timing of events is shown in Exhibit 6.

¹¹ Receipt-reactive LSMs were first discussed and studied in Johnson, McAndrews, and Soramäki (2004).

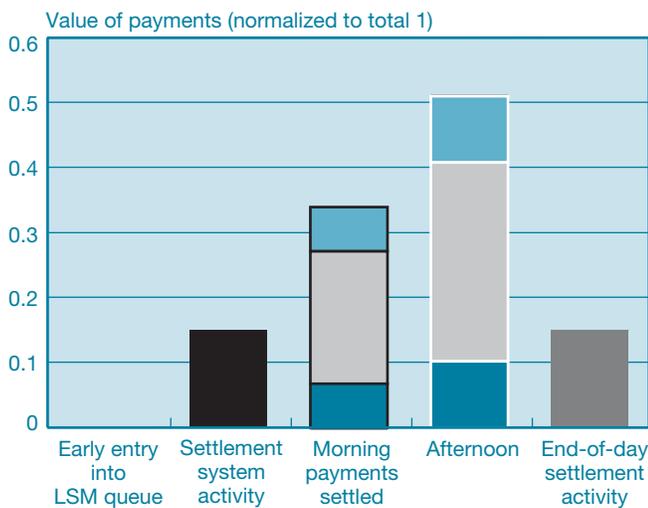
CHART 9
Receipt-Reactive Liquidity-Saving Mechanism (LSM)
with Low Costs of Delay



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

CHART 10
Receipt-Reactive Liquidity-Saving Mechanism (LSM)
with High Costs of Delay



Source: Authors' calculations.

Note: The six types of banks/colors are described in Exhibit 1.

Charts 9 and 10 illustrate the two types of equilibria that can occur with a receipt-reactive LSM. In the equilibrium described in Chart 9, all payments are queued and hence released in the morning. Note that early payments to settlement institutions, which represent the liquidity shock, occur after payments have been sent to the queue. In the equilibrium described in Chart 10, all time-sensitive payments are made early and non-time-sensitive payments are delayed.

We can now compare welfare among RTGS, a balance-reactive LSM, and a receipt-reactive LSM. In our model, a receipt-reactive LSM always provides welfare at least as high as RTGS. This outcome stands in contrast to a balance-reactive LSM, which provides lower welfare than RTGS under the circumstances outlined above. A receipt-reactive LSM may or may not provide higher welfare than a balance-reactive LSM does. If the cost of delay is sufficiently large, for example, a balance-reactive LSM will provide higher welfare. A balance-reactive LSM will also provide higher welfare when the cost of delay and the probability of a liquidity shock are small.

6. CONCLUSION

This article studies a model in which banks settle daily payments while seeking to minimize the costs associated with payment delays and intraday borrowing. The novel feature of our model is that banks are subject to two types of shocks. First, banks are randomly assigned to have either time-critical payments, whose late-period settlement imposes a cost on the bank, or non-time-critical payments. Second, banks are subject to liquidity shocks at the start of the day because of the nature of settlement institutions' operations. Together, these two types of shocks yield a rich array of strategic situations. The important parameters in our model are the cost of delay, the cost of borrowing intraday funds from the central bank, the relative size of the payments made to the settlement system with respect to bank-to-bank payments, and the proportion of time-critical payments.

To model the working of a balance-reactive liquidity-saving mechanism, we study two extreme cases. In the first, there is no possibility of netting any strict subset of payments. In the second, payments offset each other bilaterally. These two models provide different motives for using the LSM. In the first case, banks do not assign payments to the LSM queue in the hope of offsetting them within the queue, but rather to have them settle only *conditional* on receiving another payment. In the second, banks can also anticipate that, some of the time, their queued LSM payments will be offset and will settle inside the queue.

In most cases, the presence of a balance-reactive LSM increases welfare compared with a real-time gross settlement system alone, but, perhaps surprisingly, welfare may also be reduced. A balance-reactive LSM provides higher welfare if the cost of delay is high enough or low enough, and if the size of the outside settlement system and the proportion of time-critical payments are relatively low. When this is not true, RTGS can achieve higher welfare. The intuition is that RTGS creates some beneficial coordination of payments that can be undone by the presence of a queue. In our example, some banks that send payments early under the unique RTGS equilibrium choose to put the payments in the queue when they are available. The resulting reduction in the number of payments settled early leads to lower welfare.

With a receipt-reactive LSM, the level of welfare achieved is always at least as high as the level achieved in RTGS alone. Here the intuition is simpler. As banks cannot condition the sending of payments on balances, they either submit all their payments to the queue or they simply make all the time-critical payments in the early period.

In comparing balance-reactive and receipt-reactive LSMs, we find that when delay costs are high and the payments to

settlement systems are not too large, the balance-reactive LSM yields a better outcome than its receipt-reactive counterpart. As a result, while our results point to LSMs being at least weakly preferred to RTGS for all parameter configurations, the practical choice can present more of a dilemma to the operator of the large-value payments system. The dilemma is that our results show that the LSM design matters. If the wrong LSM is implemented, it can yield either lower welfare than RTGS or lower welfare than a competing LSM design. The challenge for a payments system operator is to know the sizes of the four parameters of interest. Here we have considered basic design elements in choosing the LSMs to model; more complex designs would introduce other behavioral considerations that are beyond the scope of this article.

Future research in this area would thus benefit from focusing on the empirical magnitudes of the parameters of interest. The cost of delaying payments and the proportion of payments that are time critical are especially important to measure and difficult to observe. Research employing alternative distributions of these parameters will be important, as will be the extension of the current model to include several periods.

REFERENCES

- Angelini, P.* 1998. "An Analysis of Competitive Externalities in Gross Settlement Systems." *JOURNAL OF BANKING AND FINANCE* 22, no. 1 (January): 1-18.
- . 2000. "Are Banks Risk Averse? Intraday Timing of Operations in the Interbank Market." *JOURNAL OF MONEY, CREDIT, AND BANKING* 32, no. 1 (February): 54-73.
- Armantier, O., J. Arnold, and J. McAndrews.* 2008. "Changes in the Timing Distribution of Fedwire Funds Transfers." Federal Reserve Bank of New York *ECONOMIC POLICY REVIEW* 14, no. 2 (September): 83-112.
- Bank for International Settlements.* 2005. "New Developments in Large-Value Payment Systems." Committee on Payment and Settlement Systems, Publication no. 67. May. Basel, Switzerland.
- Bech, M. L., and R. Garratt.* 2003. "The Intraday Liquidity Management Game." *JOURNAL OF ECONOMIC THEORY* 109, no. 2 (April): 198-219.
- Johnson, K., J. J. McAndrews, and K. Soramäki.* 2004. "Economizing on Liquidity with Deferred Settlement Mechanisms." Federal Reserve Bank of New York *ECONOMIC POLICY REVIEW* 10, no. 3 (December): 51-72.
- Martin, A., and J. McAndrews.* 2008. "Liquidity-Saving Mechanisms." *JOURNAL OF MONETARY ECONOMICS* 55, no. 3 (April): 554-67.
- McAndrews, J., and J. Trundle.* 2001. "New Payment System Designs: Causes and Consequences." Bank of England *FINANCIAL STABILITY REVIEW* 11, December: 127-36.
- Mills, D. C., and T. D. Nesmith.* 2008. "Risk and Concentration in Payment and Securities Settlement Systems." *JOURNAL OF MONETARY ECONOMICS* 55, no. 3 (April): 542-53.
- Roberds, W.* 1999. "The Incentive Effects of Settlement Systems: A Comparison of Gross Settlement, Net Settlement, and Gross Settlement with Queuing." Bank of Japan *INSTITUTE FOR MONETARY AND ECONOMIC STUDIES DISCUSSION PAPER SERIES*, no. 99-E-25, September.
- Willison, M.* 2005. "Real-Time Gross Settlement and Hybrid Payments Systems: A Comparison." Bank of England Working Paper no. 252.

The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.